On the solution of mathematical programming problems with equilibrium constraints

Roberto Andreani *   José Mario Martínez †

June 18, 2001 ; Updated August 16, 2008

Abstract

Mathematical programming problems with equilibrium constraints (MPEC) are nonlinear programming problems where the constraints have a form that is analogous to first-order optimality conditions of constrained optimization. We prove that, under reasonable sufficient conditions, stationary points of the sum of squares of the constraints are feasible points of the MPEC. In usual formulations of MPEC all the feasible points are nonregular in the sense that they do not satisfy the Mangasarian-Fromovitz constraint qualification of nonlinear programming. Therefore, all the feasible points satisfy the classical Fritz-John necessary optimality conditions. In principle, this can cause serious difficulties for nonlinear programming algorithms applied to MPEC. However, we show that most feasible points do not satisfy a recently introduced stronger optimality condition for nonlinear programming. This is the reason why, in general, nonlinear programming algorithms are successful when applied to MPEC.

Keywords. Mathematical programming with equilibrium constraints, optimality conditions, minimization algorithms, reformulation.

AMS: 90C33, 90C30

---

*Department of Computer Science and Statistics, University of the State of S. Paulo (UNESP), C.P. 136, CEP 15054-000, São José do Rio Preto-SP, Brazil. This author was supported by FAPESP (Grant 01/05492-1) and CNPq (Grant 301115/96-6). E-mail: andreani@dcce.ibilce.unesp.br

†Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX, FAPESP (Grant 90-3724-6), CNPq and FAEP-UNICAMP. E-Mail: martinez@ime.unicamp.br

1 Introduction

Let $f : I\!R^n \times I\!R^m \to I\!R$ be a continuously differentiable function. The MPEC (mathematical programming with equilibrium constraints) problem considered in this paper consists in the minimization of f subject to

$$h(x,y)=0, \;\; g(x,y)\le 0, \tag{1}$$

$$F(x,y)+\nabla_y t(x,y)\lambda+\nabla_y s(x,y)z=0, \tag{2}$$

$$t(x,y)=0, \tag{3}$$

$$z\ge 0, \;\; s(x,y)\le 0, \;\; z_i\,[s(x,y)]_i=0, i=1,\ldots,q, \tag{4}$$

where $[s(x,y)]_i$ is the $i-$th component of $s(x,y)$, $h(x,y)\in I\!R^{m_1}, g(x,y)\in I\!R^{\ell}$, $F(x,y)\in I\!R^m$, $s(x,y)\in I\!R^q$, $t(x,y)\in I\!R^{m_2}$. We assume that h,g,F are continuously differentiable and that s,t are twice continuously differentiable.

The constraints (1) will be called *ordinary* whereas the constraints (2,3,4) are called *equilibrium constraints.* If $F(x,y)$ is the gradient (with respect to y) of some potential function $P(x,y)$ then the equilibrium constraints are the Karush-Kuhn-Tucker (KKT) conditions of

$$\text{Minimize}_y \;\; P(x,y) \quad \text{subject to } t(x,y)=0, \; s(x,y)\le 0. \tag{5}$$

When, instead of (2,3,4), the constraints are (5), we have a *bilevel programming* problem.

The applications of MPEC include the design of transportation networks, shape optimization and economic modeling. See [18, 19, 26, 33, 35, 36, 38].

MPEC is a nonlinear programming problem, so one is tempted to apply ordinary NLP algorithms for solving it. The main difficulty traditionally associated to this strategy is associated to the constraints (4). If $s_i(x,y)=0$, the gradients that correspond to this active constraint and to $z_is_i(x,y)=0$ are linearly dependent. On the other hand, if $z_i=0$, the gradient associated to the complementarity constraint $z_is_i(x,y)=0$ is dependent of the gradient associated to $z_i=0$. Since, at any feasible point, either $z_i=0$ or $s_i(x,y)=0$, it turns out that the set of gradients of active constraints are linearly dependent for all the feasible points of MPEC. Many NLP algorithms depend strongly of the hypothesis of linearly independence of the gradients of active constraints, so their performance could be poor for MPEC or, perhaps, they could be inapplicable. This motivated many authors to introduce specific algorithms for solving MPEC. See [13] and references therein.

In this paper we discuss the application of NLP algorithms to MPEC. The first problem is addressed in Section 2, where we give sufficient conditions under which "reasonable" algorithms find feasible points of MPEC. In Section 3 we discuss the lack of regularity of feasible points of MPEC.

We show that if a method converge to points that satisfy an optimality condition introduced in [31] then, its chances of converging to a solution are enhanced. Conclusions are given in Section 4.

2 NLP algorithms and feasibility

Many NLP algorithms have the property that their accumulation points are stationary (KKT) points of the squared norm of the residual of the constraints. See, for example [11]. Of course, one would like to guarantee the feasibility of accumulation points, but this is impossible unless convexity (and existence) hypotheses are introduced. In an extreme case, the set of constraints could be empty. Even in this case it is desirable to understand how the NLP algorithm behaves. Without loss of generality, assume that the constraints of an NLP problem are given by

$$H(u) = 0 \ \ u \geq 0. \tag{6}$$

(Arbitrary constraints can be transformed into this form by the introduction of suitable slack variables.) Then a stationary point of the squared residual norm is a point that satisfies the first-order optimality conditions of

$$\text{Minimize } \|H(u)\|_2^2 \quad \text{subject to} \quad u \geq 0. \tag{7}$$

Roughly speaking, in most NLP algorithms for minimizing $f(u)$ subject to (6) (see, for example [8, 11]), limit points belong to one of the following sets:

(a) Stationary (KKT) points of (7) that do not satisfy (6);

(b) Feasible points for which some constraint qualification does not hold;

(c) KKT points of the original NLP problem.

Typically, convergence theorems for nonlinear programming algorithms begin proving that limit points are KKT points of (7). For some problems (for example, if the feasible region is empty) nothing else can be proved. Therefore, it is important to characterize the situations in which KKT points of (7) are feasible points. This is the motivation of Theorem 1.

¿From now on, we write $\|\cdot\|_2 = \|\cdot\|$. Introducing slack variables $v \in I\!R^{m_2}, w \in I\!R^q$ in (2,3,4), these constraints can be formulated as

$$h(x,y) = 0, \ \ g(x,y) + v = 0, \ \ v \geq 0, \tag{8}$$

$$F(x,y) + \nabla_y t(x,y)\lambda + \nabla_y s(x,y) z = 0, \tag{9}$$

$$t(x,y) = 0, \tag{10}$$

$$s(x,y) + w = 0, \ \ w_i z_i = 0, i = 1,\ldots,q, w \geq 0, z \geq 0. \tag{11}$$

Therefore, given $r, p > 0$, solving (8)-(11) corresponds to:

$$\text{Minimize } \|h(x,y)\|^2 + \|g(x,y)+v\|^2 \tag{12}$$
$$+\|F(x,y) + \nabla_y t(x,y)\lambda + \nabla_y s(x,y)z\|^2$$
$$+\|t(x,y)\|^2 + \|s(x,y)+w\|^2 + (\sum_{i=1}^{q}(w_i z_i)^r)^p \tag{13}$$
$$\text{subject to } v \geq 0, z \geq 0, w \geq 0. \tag{14}$$

The formulation (12,13,14) corresponds to the reformulation of nonlinear complementarity problems given in [1] and [32]. Usually, one is interested in the case $r = 2, p = 1$, which corresponds to the sum of squares of the constraints, as in the formulation (7).

In the following theorem, we prove that, under some circumstances, stationary points of (12,13,14) are feasible points of the MPEC.

Theorem 1. *Assume that the following conditions hold:*

(i) $(x^*, y^*, v^*, w^*, \lambda^*, z^*)$ *is a KKT point of (12,13, 14);*

(ii) The matrix $\nabla_y F(x^*, y^*) + \sum_{i=1}^{q}[z^*]_i \nabla^2_{yy}[s(x^*, y^*)]_i$ *is positive definite in the null-space of* $\nabla_y^T t(x^*, y^*)$ *;*

(iii) $h(x^*, y^*) = 0$, $g(x^*, y^*) \leq 0$;

(iv) $t(x^*, y)$ *is an affine function (of* y*) and the functions* $[s(x^*, y)]_i$ *are convex (as functions of* y*);*

(v) There exists $\tilde{y}$ *such that* $t(x^*, \tilde{y}) = 0$ *and* $s(x^*, \tilde{y}) \leq 0$;

(vi) $p \geq 1$, $r \geq 1$ *and* $p + r > 2$.

Then, (x^*, y^*) *is a feasible point of MPEC.*

Proof. Let us define
$$u_1 = g(x^*, y^*) + v^*,$$

$$u_2 = F(x^*, y^*) + \nabla_y t(x^*, y^*)\lambda^* + \nabla_y s(x^*, y^*)z^*,$$
$$u_3 = s(x^*, y^*) + w^*,$$
$$Z^\nu = diag(([z^*]_1)^\nu, \ldots, ([z^*]_q)^\nu),$$
$$W^\nu = diag((([(w^*]_1)^\nu, \ldots, ([w^*]_q)^\nu),$$

$$H^y = \nabla_y F(x^*, y^*) + \sum_{i=1}^{q} [z^*]_i \nabla^2_{yy} [s(x^*, y^*)]_i,$$

$$H^x = \nabla_x F(x^*, y^*) + \sum_{i=1}^{q} [z^*]_i \nabla^2_{xy} [s(x^*, y^*)]_i + \sum_{i=1}^{m_2} [\lambda^*]_i \nabla^2_{xy} [t(x^*, y^*)]_i,$$

$$\theta = \sum_{i=1}^{q} ([w^*]_i [z^*]_i)^r,$$

$$e = (1, \ldots, 1).$$

Since $(x^*, y^*, v^*, w^*, \lambda^*, z^*)$ is a KKT point, then there exist α, ρ, μ such that

$$\begin{aligned} &2\nabla_x h(x^*, y^*) h(x^*, y^*) + 2\nabla_x g(x^*, y^*) u_1 + 2H^x u_2 \\ &+2\nabla_x t(x^*, y^*) t(x^*, y^*) + 2\nabla_x s(x^*, y^*) u_3 = 0, \end{aligned} \tag{15}$$

$$\begin{aligned} &2\nabla_y h(x^*, y^*) h(x^*, y^*) + 2\nabla_y g(x^*, y^*) u_1 + 2H^y u_2 \\ &+2\nabla_y t(x^*, y^*) t(x^*, y^*) + 2\nabla_y s(x^*, y^*) u_3 = 0, \end{aligned} \tag{16}$$

$$2u_1 - \alpha = 0, \tag{17}$$

$$\nabla_y^T t(x^*, y^*) u_2 = 0, \tag{18}$$

$$2\nabla_y^T s(x^*, y^*) u_2 + p(\theta)^{p-1} r W^r Z^{r-1} e - \rho = 0, \tag{19}$$

$$2u_3 + p(\theta)^{p-1} r Z^r W^{r-1} e - \mu = 0, \tag{20}$$

$$v_i^* \alpha_i = 0 \quad i = 1, \ldots, \ell, \tag{21}$$

$$[z^*]_i \rho_i = 0, \ [w^*]_i \mu_i = 0, \quad i = 1, \ldots, q, \tag{22}$$

$$v^* \geq 0, z^* \geq 0, \mu \geq 0, \rho \geq 0, \alpha \geq 0, w^* \geq 0. \tag{23}$$

Suppose that there exists k such that $[u_1]_k \neq 0$. By (17) we have that $[u_1]_k = \alpha_k > 0$. But $g(x^*, y^*) \leq 0$, so $v_k^* = \alpha_k - [g(x^*, y^*)]_k > 0$. Therefore, $\alpha_k v_k^* > 0$ which contradicts (21). Therefore, $u_1 = 0$. Thus, since $h(x^*, y^*) = 0$, (16) can be written as:

$$2H^y u_2 + 2\nabla_y t(x^*, y^*) t(x^*, y^*) + 2\nabla_y s(x^*, y^*) u_3 = 0. \tag{24}$$

Premultiplying (19) by $2u_3$ and using (20-23) we obtain:

$$4\langle u_3, \nabla_y^T s(x^*, y^*) u_2\rangle = (pr)^2(\theta)^{2p-2}\sum_{i=1}^{q}(w_i^* z_i^*)^{2r-1} + \langle \rho, \mu\rangle \geq 0. \qquad (25)$$

Since u_2 belongs to the null-space of $\nabla_y^T t(x^*, y^*)$, premultiplying (24) by $2u_2$, we get:

$$4\langle u_2, H^y u_2\rangle + 4\langle u_2, \nabla_y s(x^*, y^*) u_3\rangle = 0. \qquad (26)$$

Therefore, using (25), we have that

$$4\langle u_2, H^y u_2\rangle + (pr)^2(\theta)^{2p-2}\sum_{i=1}^{q}(w_i^* z_i^*)^{2r-1} + \langle \rho, \mu\rangle = 0. \qquad (27)$$

But H^y is positive semidefinite in the null-space of $\nabla_y^T t(x^*, y^*)$. So, using(25, 27), we obtain:

$$\langle u_2, H^y u_2\rangle = 0 \qquad (28)$$

and

$$(pr)^2(\theta)^{2p-2} r\sum_{i=1}^{q}(w_i^* z_i^*)^{2r-1} + \langle \rho, \mu\rangle = 0. \qquad (29)$$

From (23, 29), $p \geq 1$ and $r \geq 1$, it follows that

$$\theta = 0. \qquad (30)$$

Using (30) and that $p + r > 2$, we can rewrite (20) as:

$$2u_3 - \mu = 0, \quad u_3 \geq 0. \qquad (31)$$

Since H^y is positive definite in the null-space of $\nabla_y^T t$, we obtain that $u_2 = 0$. Then, by (24),

$$2\nabla_y t(x^*, y^*) t(x^*, y^*) + 2\nabla_y s(x^*, y^*) u_3 = 0. \qquad (32)$$

Since there exists $(x^*, \tilde{y})$ such that $t(x^*, \tilde{y}) = 0$, by the linearity of $t(x^*, y)$, it follows that

$$t(x^*, y^*) = t(x^*, \tilde{y}) + \nabla_y^T t(x^*, y^*)(y^* - \tilde{y}) = \nabla_y^T t(x^*, y^*)(y^* - \tilde{y}). \qquad (33)$$

Premultiplying (32) by $y^* - \tilde{y}$ and using (33), the equation (32) yields:

$$\|t(x^*, y^*)\|^2 + \langle (y^* - \tilde{y}), \nabla_y s(x^*, y^*) u_3\rangle$$

$$= \|t(x^*, y^*)\|^2 + \sum_{i=1}^{q} [u_3]_i \langle \nabla_y [s(x^*, y^*)]_i, (y^* - \tilde{y}) \rangle = 0. \tag{34}$$

Define $\mathcal{J} = \{i \in \{1, \ldots, q\} \mid w_i^* > 0\}$. If $i \in \mathcal{J}$ we get $\mu_i = 0$. So, by (31),

$$[u_3]_i = 0 \quad \text{for all} \quad i \in \mathcal{J}. \tag{35}$$

If $i \notin \mathcal{J}$, we get $w_i^* = 0$. So, by (31) and using the definition of u_3, we obtain:

$$[u_3]_i = s_i(x^*, y^*) \geq 0, \quad \text{for all} \quad i \notin \mathcal{J}. \tag{36}$$

But $[s(x^*, y)]_i$ is convex, therefore, for all $i = 1, \ldots, q$,

$$\langle \nabla_y [s(x^*, y^*)]_i, (y^* - \tilde{y}) \rangle \geq ([s(x^*, y^*)]_i - [s(x^*, \tilde{y})]_i). \tag{37}$$

Then, by (35, 37, 36), equation (34) implies that

$$0 = \|t(x^*, y^*)\|^2 + \sum_{i=1}^{q} [u_3]_i \langle \nabla_y [s(x^*, y^*)]_i, (y^* - \tilde{y}) \rangle$$

$$\geq \|t(x^*, y^*)\|^2 + \sum_{i \notin \mathcal{J}} ([s(x^*, y^*)]_i^2 - [s(x^*, y^*)]_i [s(x^*, \tilde{y})]_i). \tag{38}$$

By the hypothesis (v) we have that $s(x^*, \tilde{y}) \leq 0$. So, by (36), we get:

$$0 \geq \|t(x^*, y^*)\|^2 + \sum_{i \notin \mathcal{J}} (s_i(x^*, y^*))^2. \tag{39}$$

Therefore, by (39), we obtain that $t(x^*, y^*) = 0$ and that $[u_3]_i = s(x^*, y^*) = 0$ whenever $i \notin \mathcal{J}$.

Therefore, by (35), we deduce that $u_3 = 0$. This completes the proof. □

Remarks

(i) The derivatives with respect to x play no role in the proof of Theorem 1. Therefore, the theorem can be generalized in an obvious way with the hypothesis of stationarity with respect to y. We maintain the present formulation to stress the relationship with the class of points that are obtained by NLP algorithms.

(ii) In Theorem 1, Hypothesis (ii) cannot be relaxed. Define

$$F(y) = \begin{cases} -1 & \text{if } y \leq 1 \\ (y-1)^2 - 1 & \text{if } y > 1 \end{cases}$$

and $s(y) = -y$. The operator F if monotone and the solution of the variational inequality problem is $y = 2$. However, the point $(y, z, w) = (0, 0, 0)$ is a KKT point of

$$\text{Minimize } (-1 - z)^2 + (-y + w)^2 + (wz)^2 \text{ s. t. } z, w \geq 0,$$

wich does not correspond to a solution of the variational inequality problem.

(iii) Hypothesis (v) cannot be relaxed either. Define $F(y) = y$ and $s(y) = y^2 + 1$. Then, F is strongly monotone. But $(y, z, w) = (0, 0, 0)$ is stationary point of

$$\text{Minimize } (y + 2yz)^2 + (y^2 + 1 + w)^2 + (wz)^2 \text{ s. t. } z, w \geq 0$$

and this point does not correspond to a solution of the problem.

In the following theorem we prove that the thesis of Theorem 1 remains true assuming positive semidefiniteness of the reduced "Hessian-like" matrix if we assume that F and s are affine functions with respect to y.

Theorem 2. *Assume that the following conditions hold:*

(i) $(x^*, y^*, v^*, w^*, \lambda^*, z^*)$ *is a stationary point of (12,13, 14);*

(ii) The functions $F(x^*, y)$, $t(x^*, y)$ *and* $s(x^*, y)$ *are affine functions (w.r.t.* y*);*

(iii) The matrix $\nabla_y F(x^*, y^*)$ *is positive semidefinite in the null-space of* $\nabla_y^T t(x^*, y^*)$*;*

(iv) $h(x^*, y^*) = 0$, $g(x^*, y^*) \leq 0$*;*

(v) There exist $\tilde{y}, \tilde{\lambda}, \tilde{z} \geq 0$ *such that*

$$F(x^*, \tilde{y}) + \nabla_y t(x^*, \tilde{y})\tilde{\lambda} + \nabla_y s(x^*, \tilde{y})\tilde{z} = 0,$$

$t(x^*, \tilde{y}) = 0$ *and* $s(x^*, \tilde{y}) \leq 0$*;*

(vi) $p \geq 1$ *and* $r \geq 1$ *and* $p + r > 2$.

Then, (x^*, y^*) *is a feasible point of MPEC.*

Proof. Repeat the proof of Theorem 1 for obtaining (30). So, since $p \geq 1, r \geq 1, p + r > 2$, the point $(y^*, \lambda^*, z^*, w^*)$ is a stationary point of the convex problem:

$$\text{Minimize} \tag{40}$$

$$\|F(x^*, y) + \nabla_y t(x^*, y)\lambda + \nabla_y s(x^*, y)z\|^2$$
$$+\|t(x^*, y)\|^2 + \|s(x^*, y) + w\|^2 \tag{41}$$

$$\text{subject to } v \geq 0, z \geq 0, w \geq 0. \tag{42}$$

By Hypothesis *(v)*, there exists $\tilde{w} \geq 0$ such that $(\tilde{y}, \tilde{\lambda}, \tilde{z}, \tilde{w})$ is a solution (40,41,42) and the objective function value is zero. But $(y^*, \lambda^*, z^*, w^*)$ is a stationary point, so, since the problem is convex, the proof is complete. □

The derivatives with respect to x play no role in the proof of Theorem 1. Therefore, the theorem can be generalized in an obvious way with the hypothesis of stationarity with respect to y. We maintain the present formulation to stress the relationship with the class of points that are obtained by NLP algorithms. volver Theorems 1 and 2 establish sufficient conditions under which reasonable NLP algorithms find feasible points of MPEC. Very roughly speaking, the conditions "on the problem" say that the equilibrium constraints "look like" optimality conditions of a convex problem. The assumption $h(x^*, y^*) = 0, g(x^*, y^*) \leq 0$ has algorithmic consequences. It tells us that preserving feasibility of the ordinary constraints is important. We cannot guarantee feasibility of the ordinary constraints with the only assumption of stationarity of the sum of squares. Therefore, algorithms that maintain feasibility of the ordinary constraints at all the iterations should, perhaps, be preferred. When the ordinary constraints have a simple structure (for example, when they define boxes) to maintain their feasibility is simple, but the situation is more complicate when the constraints are highly nonlinear.

3 Optimality conditions

Assume that we have a method that converges (in some sense) to feasible points of MPEC. Essentially, a sufficient condition for that property was given in the previous section but, of course, many algorithms generate sequences that converge to feasible points, even when the hypotheses of Theorem 1 do not hold. For many nonlinear programming algorithms (see, for example [8, 11]) it has been proved the following property:

Property 1. *If a limit point is feasible and the gradients of the active constraints are linearly independent, then it satisfies the KKT conditions of the nonlinear programming problem.*

Since, in MPEC, the gradients of active constraints are linearly dependent at all the feasible points, it follows that an algorithm that satisfies Property 1 can "converge" to any feasible point. In other words, Property 1 adds nothing to the convergence properties of an algorithm, provided that convergence to feasible points takes place.

More serious than the lack of linear independence of gradients of active constraints is the fact that even the Mangasarian-Fromovitz (MF) constraint qualification [7, 24, 25, 34] is not satisfied at any feasible point of the MPEC. Therefore, even algorithms that satisfy the following *Property 2* are not satisfactory, unless something better can be proved about them.

Property 2. *If a limit point is feasible, then it satisfies the Fritz-John conditions of the nonlinear programming problem.*

Points that do not satisfy the MF constraint qualification, necessarily satisfy the Fritz-John optimality conditions. See [24, 34]. Therefore, if Property 2 holds, we have, again, a situation where all the feasible points can be cluster points of the algorithm.

This state of facts motivated many authors to devise specific algorithms (and specific optimality conditions) for MPEC. See [13] and references therein. Herskovits and Leontiev [21] observed that the inequality constraints are, in some sense, superfluous from the point of view of optimality conditions unless two complementary variables annihilate at the same point. Based on this observation, they developed an interior point algorithm for bilevel programming that can be easily generalized to MPEC.

Our objective in this paper is not to analyze specific algorithms for MPEC but to discuss the applicability of general NLP algorithms. Therefore, our question is "Are general NLP algorithms so bad that they can converge to any feasible point of MPEC?"

Fortunately, the answer is *no*. The set of points to which specific algorithms can converge is, usually, much smaller than the set of Fritz-John points. Martínez and Svaiter [31] introduced a "sequential" necessary optimality condition (AGP) and showed that, in general, the set of points that satisfy AGP is strictly smaller than the set of Fritz-John points. So, algorithms that are guaranteed to converge to AGP points are natural candidates for solving MPEC.

Let us state here the AGP condition for the special case in which the nonlinear programming problem is

$$\text{Minimize} \quad \Phi(x,y) \;\text{ subject to } \quad H(x,y)=0, \;\; y \geq 0. \tag{43}$$

Given x and $y \geq 0$, we define $\Omega(x,y)$ as the set of points (x', y') such that $y' \geq 0$ and

$$H'(x, y)(x' - x, y' - y) = 0. \tag{44}$$

For all (x, y) such that $y \geq 0$, we define

$$g_P(x, y) = P_{\Omega(x,y)}[(x, y) - \nabla\Phi(x, y)] - (x, y), \tag{45}$$

where $P_C(\cdot)$ denotes the orthogonal projection onto C. The vector $g_P(x, y)$ will be called *Approximated Gradient Projection.*

We say that a feasible point (x^*, y^*) satisfies the *AGP property* if there exists a sequence (x^k, y^k), $y^k \geq 0$ that converges to (x^*, y^*) and such that $g_P(x^k, y^k) \to 0$.

In [31] it has been proved that this property implies (and is strictly stronger than) Fritz-John. This proof is also implicit in the proof of Fritz-John conditions in Chapter 3 of [7]. The example below shows that the set of points that satisfy the AGP property can be strictly smaller than the set of Fritz-John points. Consider the following problem in two variables:

$$\text{Minimize } x^2 \text{ s.t. } y = 0, \;\; y \geq 0.$$

Here, all the feasible points are Fritz-John (because the fail to satisfy the Mangasarian-Fromovitz constraint qualification) but only the solution of the problem satisfies the AGP property.

Therefore, in general, the set of AGP points is a better approximation of the set of local minimizers than the set of Fritz-John points.

The following result has been proved in [31].

Theorem 3. *Assume that (x^*, y^*) is a local minimizer of (43). Then, (x^*, y^*) satisfies the AGP property.*

The AGP property is naturally satisfied by the recently introduced inexact-restoration algorithms [27, 28, 30] and, probably, by many other classical algorithms for nonlinear programming.

The applicability of AGP to the MPEC problem comes from the following theorem. We prove that, in MPEC, the set of points that satisfy the AGP property is considerably smaller than the set of Fritz-John points.

Consider the constraints of MPEC given in the form (8,9,10,11). In this way, the MPEC problem is a particular case of

$$\text{Minimize } \Phi(x, v, w, z) \tag{46}$$

subject to

$$H(x, v, w, z) = 0, \; w_i z_i = 0, i = 1, \ldots, q, \tag{47}$$

$$v \geq 0, w \geq 0, z \geq 0, \tag{48}$$

where $x \in I\!R^n, v \in I\!R^m, w \in I\!R^q, z \in I\!R^q, H(x, v, w, z) \in I\!R^\ell$. We use some abuse of notation to avoid an excessive number of variables and different dimensions. Remember that the lack of regularity of the feasible points of this problem is due to the simultaneous presence of the constraints $w_i z_i = 0$ and $w_i \geq 0, z_i \geq 0$. Assume that (x^*, v^*, w^*, z^*) is a feasible point of the problem (46,47,48) and that $w_i^* + z_i^* > 0$ for all $i \in I_* \subset \{1, \ldots, q\}$.

We define an associated problem to (46,47,48) which consists simply in dropping the constraints $w_i \geq 0, z_i \geq 0$ for $i \in I_*$. (This is the idea behind the analysis of Herskovits and Leontiev of bilevel programming problems [21].) So, the associated problem is

$$\text{Minimize } \Phi(x, v, w, z) \tag{49}$$

subject to

$$H(x, v, w, z) = 0, \; w_i z_i = 0, i = 1, \ldots, q \tag{50}$$

$$v \geq 0; \;\; w_i \geq 0, z_i \geq 0 \;\; \forall \;\; i \notin I_*. \tag{51}$$

Below we prove that, if (x^*, v^*, w^*, z^*) is an AGP point of (46,47,48) then, it is necessarily an AGP point of (49,50,51). This means that cluster points of algorithms that satisfy the AGP property applied to (46,47,48) are, very likely, AGP points of an associated problem where the constraints that make that all the feasible points are nonregular are not present any more. The consequence is that most feasible (Fritz-John) points of the MPEC are discarded as possible limit points of the algorithm.

Theorem 4. *Assume that (x^*, v^*, w^*, z^*) is an AGP point of (46,47,48) such that $w_i^* + z_i^* > 0 \; \forall \; i \in I_*$. Then, (x^*, v^*, w^*, z^*) is an AGP point of (49,50,51).*

Proof. Suppose that (x^k, v^k, w^k, z^k) is a sequence that satisfies the AGP property and converges to the feasible point (x^*, v^*, w^*, z^*). The names of the variables w_i and z_i can be interchanged, so we can assume that $z_i^* > 0$ for all $i \in I_*$. Let $(\bar{x}^k, \bar{v}^k, \bar{w}^k, \bar{z}^k)$ be the projection of $(x^k, v^k, w^k, z^k) - \nabla\Phi(x^k, v^k, w^k, z^k)$ on $\Omega(x^k, v^k, w^k, z^k)$. Now, $\Omega(x^k, v^k, w^k, z^k)$ is defined by a set of linear equalities and inequalities, including

$$z_i^k (w_i - w_i^k) + w_i^k (z_i - z_i^k) = 0, \;\; w_i \geq 0, \;\; z_i \geq 0, \tag{52}$$

for all $i \in I_*$. Since $z_i^* > 0$ for $i \in I_*$, $z_i^k \to z_i^*$ and, by the AGP property, $|\bar{z}_i^k - z_i^k| \to 0$, then $\bar{z}_i^k > 0$ for all $i \in I_*$ and for k large enough. Therefore, by the convexity of the subproblems that define the projections, the constraints (52) can be replaced by

$$z_i^k (w_i - w_i^k) + w_i^k (z_i - z_i^k) = 0, \quad w_i \geq 0 \tag{53}$$

for all $i \in I_*$. The rest of the proof consists on proving that, in the definition of the projection subproblem, the constraint $w_i \geq 0$ can also be dropped. Let us take a fixed index $i \in I_*$ and analyze two cases:

Case 1: There exists a natural number k_0 such that $w_i^k = 0$ for all $k \geq k_0$.
Case 2: For infinitely many indices, $w_i^k > 0$.

In Case 1, for $k \geq k_0$, (53) reduces to

$$z_i^k w_i = 0, \quad w_i \geq 0.$$

But, for k large enough, $z_i^k > 0$ so, this is obviously equivalent to $w_i = 0$. It follows that (53) reduces to

$$z_i^k (w_i - w_i^k) + w_i^k (z_i - z_i^k) = 0, \tag{54}$$

that is to say, the constraint $w_i \geq 0$ can be dropped.

In Case 2, we can assume, taking a convenient subsequence, that $w_i^k > 0$ and $z_i^k > 0$ for all k. We are going to prove that, for k large enough, $\bar{w}_i^k > 0$. In fact, if $\bar{w}_i^k = 0$, then, by (53),

$$z_i^k (-w_i^k) + w_i^k (\bar{z}_i^k - z_i^k) = 0.$$

Since $w_i^k > 0$, this implies that $\bar{z}_i^k = 2 z_i^k$. This contradicts the fact that $|\bar{z}_i^k - z_i^k| \to 0$ and $z_i^k \to z_i^* > 0$. Therefore, $\bar{w}_i^k > 0$ for k large enough. This implies that the constraint $w_i > 0$ can also be dropped in (53).

Therefore, repeating this reasoning for all $i \in I_*$, we see that, for k large enough, the subproblem that defines the projection can be defined using the constraints (54) instead of (52). But this is the same as saying that the sequence satisfies the AGP property related to the problem (49,50,51), as we wanted to prove. $\square$

Theorem 4 says that, excluding the points such that $w_i = z_i = 0$, the AGP points of the problem (46,47,48) are AGP points of (49,50,51). Therefore, the set of possible limit points of algorithms that satisfy the AGP property is restricted to the AGP points of (49,50,51). When, at a given point, the Mangasarian-Fromovitz constraint qualification of the problem

(49,50,51) is satisfied (in particular, when the gradients of the active constraints of (49,50,51) are linearly independent), the AGP property implies the KKT optimality conditions. Since the KKT conditions of (49,50,51) clearly imply those of (46,47,48), we can establish the following theorem.

Theorem 5. *Assume that (x^*, v^*, w^*, z^*) is a feasible AGP point of (46,47,48) such that $w_i^* + z_i^* > 0 \; \forall \; i = 1, \ldots, q$ and that (x^*, v^*, w^*, z^*) satisfies the MF constraint qualification of (49,50,51). Then, (x^*, v^*, w^*, z^*) is a KKT point of (46,47,48).*

4 Conclusions

If one wants to use NLP algorithms for solving MPEC, one must choose a reformulation of the constraints that makes them suitable for NLP applications. In this paper we used straightforward reformulations that consist on squaring the complementarity constraints, maintaining the positivity of the variables. See [6, 32]. We showed that, in some situations, this reformulation is good, in the sense that stationary points of the squared violation of the constraints are, necessarily, feasible points. The same study must be done for other reformulations. Reformulations based on the Penalized-Fischer-Burmeister function are especially attractive, according to the experiments in [6, 9].

We discussed the problem induced by the nonregularity of the constraints at all the feasible points of MPEC. All the feasible points of MPEC satisfy the Fritz-John optimality condition which, in consequence, is completely useless in this case. However, we observed that many algorithms have the property of converging to points that satisfy an optimality condition (introduced in [31]) which is sharper than Fritz-John. The MPEC problem is a good example of the difference between Fritz-John and the sequential optimality condition AGP, given in [31]. In fact, we proved that the fulfillment of the AGP condition implies (under a “dual nondegeneracy” assumption) the fulfillment of AGP on a problem where positivity constraints are eliminated. In the new problem, not all the feasible points are Fritz-John points. As a consequence, there are good reasons to expect that the application of algorithms that converge to AGP points will behave well in MPEC problems.

We feel that even NLP algorithms that are not guaranteed to satisfy the AGP property are unlike to converge to “very arbitrary” feasible points of MPEC. For example, there does not exist a proof that the Herskovits algorithm [20] satisfies AGP, but its application to MPEC seems to be successful [21]. We conjecture that the same is true for other interior NLP algorithms as the ones described in [8] and [37]. We think that it is worthwhile to study this aspect of NLP algorithms, not only from the point of view of their applicability to MPEC but also as an additional theoretical index of efficiency of NLP methods. Further research on this subject should be expected in the near future.

Acknowledgements. We are indebted to Luján Latorre, who read this manuscript and pointed out some misprints. We are also deeply indebted to two anonymous referees whose careful reading of the manuscript lead to considerable improvements. Both referees detected that the hypothesis $s(x^*, \tilde{y}) \leq 0$ was lacking in the first manuscript. The second counterexample that follows Theorem 1 is, essentially, due to one of the referees. The other referee motivated the first remark that follows Theorem 1 and suggested the general form of Theorem 4 presented in this revised version.

References

[1] R. Andreani, A. Friedlander and J. M. Martínez [1997], On the solution of finite-dimensional variational inequalities using smooth optimization with simple bounds, *Journal on Optimization Theory and Applications* 94, pp. 635-657.

[2] R. Andreani and J. M. Martínez [1999], Solving complementarity problems by means of a new smooth constrained nonlinear solver, in *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, edited by M. Fukushima and L. Qi, Kluwer Academic Publishers, pp. 1–24.

[3] R. Andreani and J. M. Martínez [1998], On the solution of the extended linear complementarity problem, *Linear Algebra and its Applications* 281, pp. 247-257.

[4] R. Andreani and J. M. Martínez [1999], On the reformulation of nonlinear complementarity problems using the Fischer-Burmeister function, *Applied Mathematics Letters* 12, pp. 7-12.

[5] R. Andreani and J. M. Martínez [1999], On the solution of bounded and unbounded mixed complementarity problems, to appear in *Optimization*.

[6] R. Andreani and J. M. Martínez [2000], Reformulation of variational inequalities on a simplex and compactification of complementarity problems, *SIAM Journal on Optimization* 10, pp. 878-895.

[7] D. Bertsekas [1999], *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts.

[8] R. H. Byrd, M. E. Hribar and J. Nocedal [1999], An interior point algorithm for large-scale nonlinear programming, *SIAM Journal on Optimization* 9, pp. 877-900.

[9] B. Chen, X. Chen and C. Kanzow [2000], A penalized Fischer-Burmeister NCP-function, *Mathematical Programming* 88, pp. 211-216.

[10] A. R. Conn, N. I. M. Gould and Ph. L. Toint [1991], A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Numerical Analysis* 28, pp. 545 - 572.

[11] F. M. Gomes, M. C. Maciel, J. M. Martínez [1999], Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters, *Mathematical Programming* 84, pp. 161-200.

[12] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos [2000], Augmented Lagrangian algorithms based on the spectral projected gradient for solving nonlinear programming problems, in CLAIO 2000, Workshop on Nonlinear Programming organized by J. J. Júdice, Ciudad de México.

[13] F. Facchinei, H. Jiang and L. Qi [1999], A smoothing method for mathematical programs with equilibrium constraints, *Mathematical Programming* 85, pp. 107-134.

[14] L. Fernandes, A. Friedlander, M. Guedes and J. Júdice [2001], Solution of a general linear complementarity problems using smooth optimization and its application to bilinear programming and LCP, *Applied Mathematics and Optimization* 43, pp. 1-19.

[15] A. Friedlander, J. M. Martínez and S. A. Santos [1994], On the resolution of large scale linearly constrained convex minimization problems, *SIAM Journal on Optimization* 4, pp. 331-339.

[16] A. Friedlander, J. M. Martínez and S. A. Santos [1995], Solution of linear complementarity problems using minimization with simple bounds, *Journal of Global Optimization* 6, pp. 253 - 267.

[17] A. Friedlander, J. M. Martínez and S. A. Santos [1995], A new strategy for solving variational inequalities on bounded polytopes, *Numerical Functional Analysis and Optimization* 16, pp. 653-668.

[18] P. T. Harker and J. S. Pang [1990], Finite-dimensional variational inequality and nonlinear complementarity problem: a survey of theory, algorithms and applications, *Mathematical Programming* 48, pp. 161-220.

[19] J. Haslinger and P. Neittaanmäki [1988], Finite element approximation for optimal shape design: theory and applications, Wiley, Chichester.

[20] J. Herskovits [1998], Feasible direction interior point technique for nonlinear optimization, *Journal of Optimization Theory and Applications* 99, pp. 121-146.

[21] J. Herskovits and A. Leontiev [2000], New optimality conditions and an interior point algorithm for bilevel programming, Technical Report, COPPE, Universidade Federal do Rio de Janeiro, Brazil.

[22] N. Krejić, J. M. Martínez, M. P. Mello and E. A. Pilotta [2000], Validation of an Augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of hard-spheres problems, *Computational Optimization and Applications* 16, 247-263.

[23] J. J. Júdice [1994], Algorithms for linear complementarity problems, *Algorithms for Continuous Optimization*, edited by E. Spedicato, Kluwer, pp. 435-474.

[24] O. L. Mangasarian [1969], *Nonlinear Programming*, Mc Graw Hill, New York.

[25] O. L. Mangasarian and S. Fromovitz [1967], Fritz-John necessary optimality conditions in presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications* 17, pp.37.

[26] P. Marcotte [1986], Network design problem with congestion effects: a class of bilevel programming, *Mathematical Programming* 34, pp. 142-162.

[27] J. M. Martínez [1998], Two-phase model algorithm with global convergence for nonlinear programming. *Journal of Optimization Theory and Applications* 96, pp. 397-436.

[28] J. M. Martínez [1999], Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. To appear in *Journal of Optimization Theory and Applications*.

[29] J. M. Martínez [2000], BOX-QUACAN and the implementation of Augmented Lagrangian algorithms for minimization with inequality constraints. *Computational and Applied Mathematics* 19, pp. 31-56.

[30] J. M. Martínez and E. A. Pilotta [2000], Inexact restoration algorithm for constrained optimization, *Journal of Optimization Theory and Applications* 104, pp. 135-163.

[31] J. M. Martínez and B. F. Svaiter [1999], A practical optimality condition without constraint qualifications for nonlinear programming, Technical Report, Institute of Mathematics, UNICAMP, Brazil.

[32] J. J. Moré [1996], Global methods for nonlinear complementarity problems, *Mathematics of Operations Research* 21, pp. 589–614.

[33] F. H. Murphy, H. D. Sherali and A. L. Soyster [1982], A mathematical programming approach for determining oligopolistic market equilibrium, *Mathematical Programming* 24, pp. 92-106.

[34] R. T. Rockafellar [1993], Lagrange multipliers and optimality, *SIAM Review* 35, pp. 183-238.

[35] C. Suwansirikul, C. Friesz and R. L. Tobin [1987], Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem, *Transportation Science* 21, pp. 254-263.

[36] R. L. Tobin [1992], Uniqueness results and algorithms for Stackelberg-Cournot-Nash equilibria, *Annals of Operations Research* 34, pp. 21-36.

[37] R. J. Vanderbei and D. F. Shanno [1999], An interior point algorithm for nonconvex nonlinear programming, *Computational Optimization and Applications* 13, pp. 231-252.

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