A practical optimality condition without constraint qualifications for nonlinear programming

José Mario Martínez * Benar Fux Svaiter †

January 22, 2001

Abstract

A new optimality condition for minimization with general constraints is introduced. Unlike the KKT conditions, this condition is satisfied by local minimizers of nonlinear programming problems, independently of constraint qualifications. The new condition implies, and is strictly stronger than, Fritz-John optimality conditions. Sufficiency for convex programming is proved.

Keywords. Optimality conditions, Karush-Kuhn-Tucker, minimization algorithms, constrained optimization.

AMS: 90C30

1 Introduction

In the global convergence theory of most nonlinear programming algorithms it is proved that some limit point (perhaps all the limit points) of the generated sequence satisfies a necessary condition for constrained minimization. Usually, the necessary condition says that the Karush-Kuhn-Tucker conditions are satisfied if the limit point satisfies some constraint qualification. When the constraint qualification is the one introduced by Mangasarian and Fromovitz [16], we say that the Fritz-John optimality condition is satisfied. Separation techniques are, perhaps, the most powerful tools for proving optimality conditions. See [5, 7, 11, 15, 12]. Therefore, according to that "model

*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

†Institute of Pure and Applied Mathematics (IMPA), CNPq, Estrada Dona Castorina 110, Rio de Janeiro, Brazil. This author was partially supported by CNPq Grant 301200/93-9(RN) and by PRONEX-Optimization. E-mail: benar@impa.br

global theorem", the algorithmic sequence might converge to points where the constraint qualification is not satisfied. This means that a potential failure of the nonlinear programming algorithm could be due to the presence of "nonregular" feasible attraction points that are not local minimizers.

However, convergence of good algorithms to nonregular feasible points that are not local minimizers is not frequent. This is not surprising, since algorithms are designed taking into account, at every iteration, the necessity of *minimizing* an objective function. So, convergence to points at which the objective function has no special properties should be unusual, at least for well designed methods.

On the other hand, convergence to local minimizers can be guaranteed only under restrictive second-order properties of the constraints (see [4]) or convexity-like assumptions.

Therefore, both practical experience and common sense indicates that the set of feasible limit points of good algorithms for constrained optimization, although being larger than the set of local minimizers, does not include the set of all "nonregular" feasible points. For example, 0 is a nonregular point of the problem of minimizing x subject to $-1 \leq x^3 \leq 0$ but no reasonable minimization algorithm will converge to it.

The observations above lead us to seek optimality conditions that fit better the behavior of practical methods. Another motivation for discovering "sharper than Fritz-John" optimality conditions comes from considering mathematical programming with equilibrium constraints (MPEC) problems. See [8, 9, 25] and references therein. In standard formulations of MPEC, no feasible point satisfies the Mangasarian-Fromovitz constraint qualification and, thus, all the feasible points satisfy the Fritz-John optimality condition. So, it is necessary to study optimality conditions that approximate more accurately the minimizers of the problem than Fritz-John conditions.

In this paper, we introduce an optimality condition that, roughly formulated, says that "an approximate gradient projection tends to zero". For this reason, we call it the Approximate Gradient Projection (AGP) property. We prove that the AGP property is satisfied by local minimizers of constrained optimization problems independently of constraint qualifications and using only first-order differentiability. Therefore, the AGP property is a genuine necessary optimality condition that does not use constraint qualifications at all. Moreover, we show that the set of "AGP points" is contained, and can be *strictly contained* in the set of Fritz-John points.

The reputation of KKT conditions lies, not only in being necessary optimality conditions under constraint qualifications, but also in the fact that they are sufficient optimality conditions if the nonlinear program is convex. We prove that, essentially, this is also true for the AGP property.

This paper is organized as follows. In Section 2 we state rigorously the AGP property, we prove that it is satisfied by every local minimizer of a nonlinear programming problem and that it implies Fritz-John. In Section 3 we prove that the new condition is sufficient for convex problems. Conclusions are given in Section 4.

2 The necessary condition

We consider the nonlinear programming problem

$$\text{Minimize } f(x) \quad \text{subject to} \quad x \in \Omega, \tag{1}$$

where

$$\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0, \;\; h(x) = 0\}, \tag{2}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and all the functions have continuous first partial derivatives.

Let $\gamma \in [-\infty, 0]$. For all $x \in \mathbb{R}^n$ we define $\Omega(x, \gamma)$ as the set of points $z \in \mathbb{R}^n$ that satisfy:

$$g_i(x) + g_i'(x)(z - x) \leq 0 \;\; \text{if} \;\; \gamma < g_i(x) < 0, \tag{3}$$

$$g_i'(x)(z - x) \leq 0 \;\; \text{if} \;\; g_i(x) \geq 0 \tag{4}$$

and

$$h'(x)(z - x) = 0. \tag{5}$$

The set $\Omega(x, \gamma)$ (a closed and convex polyhedron) can be interpreted as a linear approximation of the set of points $z \in \mathbb{R}^n$ that satisfy

$$h(z) = h(x), \; g_i(z) \leq g_i(x) \text{ if } g_i(x) \geq 0, \; g_i(z) \leq 0 \;\; \text{if} \;\; g_i(x) \in (\gamma, 0).$$

Observe that

$$\Omega(x, 0) = \{z \in \mathbb{R}^n \mid h'(x)(z - x) = 0, \; g_i'(x)(z - x) \leq 0 \;\; \text{if} \;\; g_i(x) \geq 0, \; i = 1, \ldots, p\}.$$

For all $x \in \mathbb{R}^n$, we define

$$d(x, \gamma) = P_{\Omega(x,\gamma)}(x - \nabla f(x)) - x, \tag{6}$$

where $P_C(y)$ denotes the orthogonal projection of y onto C for all $y \in \mathbb{R}^n$, $C \subset \mathbb{R}^n$ closed, convex. The vector $d(x, \gamma)$ will be called Approximated Gradient Projection.

We will denote $\|\cdot\| = \|\cdot\|_2$ and

$$\mathcal{B}(x, \rho) = \{z \in \mathbb{R}^n \mid \|z - x\| \leq \rho\}$$

for all $x \in \mathbb{R}^n$, $\rho > 0$. As usual, for all $v \in \mathbb{R}^n, v = (v_1, \ldots, v_n)$, we denote $v_+ = (\max\{0, v_1\}, \ldots, \max\{0, v_n\})$.

The main result of this section is proved below. It says that, if x^* is a local minimizer of (1)-(2), we can find points with arbitrary small approximate gradient projections that are arbitrarily close to x^*. The technique of this proof is similar to the one used by Bertsekas (Proposition 3.3.5 of [2]) to prove Fritz-John conditions.

Theorem 1. *Assume that x^* is a local minimizer of (1)-(2) and let $\gamma \in [-\infty, 0]$, $\varepsilon, \delta > 0$ be given. Then, there exists $x \in \mathbb{R}^n$ such that $\|x - x^*\| \leq \delta$ and $\|d(x, \gamma)\| \leq \varepsilon$.*

Proof. Let $\rho \in (0, \delta)$ be such that x^* is a global minimizer of $f(x)$ on $\Omega \cap \mathcal{B}(x^*, \rho)$. Define, for all $x \in \mathbb{R}^n$,

$$\varphi(x) = f(x) + \frac{\varepsilon}{2\rho}\|x - x^*\|^2.$$

Clearly, x^* is the unique global solution of

$$\text{Minimize } \varphi(x) \quad \text{subject to } \; x \in \Omega \cap \mathcal{B}(x^*, \rho). \tag{7}$$

Define, for all $x \in \mathbb{R}^n$, $\mu > 0$,

$$\Phi_\mu(x) = \varphi(x) + \frac{\mu}{2}[\|h(x)\|^2 + \|g(x)_+\|^2].$$

The External Penalty theory (see, for instance, [10]) guarantees that, for μ sufficiently large, there exists a solution of

$$\text{Minimize } \Phi_\mu(x) \quad \text{subject to } \quad x \in \mathcal{B}(x^*, \rho) \tag{8}$$

that is as close as desired to the global minimizer of $\varphi(x)$ on $\Omega \cap \mathcal{B}(x^*, \rho)$. So, for μ large enough, there exists a solution x_μ of (8) in the interior of $\mathcal{B}(x^*, \rho)$. Therefore,

$$\nabla \Phi_\mu(x_\mu) = 0.$$

Thus (writing, for simplicity $x = x_\mu$), we obtain:

$$0 = \nabla \Phi_\mu(x) = \nabla \varphi(x) + \mu[h'(x)^T h(x) + \sum_{g_i(x) > 0} g_i(x) \nabla g_i(x)].$$

So,

$$\nabla f(x) + \mu[h'(x)^T h(x) + \sum_{g_i(x) > 0} g_i(x) \nabla g_i(x)] + \frac{\varepsilon}{\rho}(x - x^*) = 0. \tag{9}$$

Since $x = x_\mu$ lies in the interior of the ball, we have that

$$\|x - x^*\| < \rho < \delta$$

So, by (9),

$$\|\nabla f(x) + \mu[h'(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)]\| \le \varepsilon.$$

So,

$$\|[x - \nabla f(x)] - [x + \mu[h'(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)]\| \le \varepsilon.$$

This implies, taking projections onto $\Omega(x,\gamma)$, that

$$\|P_{\Omega(x,\gamma)}(x - \nabla f(x)) - P_{\Omega(x,\gamma)}(x + \mu[h'(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)])\| \le \varepsilon. \tag{10}$$

It remains to prove that

$$P_{\Omega(x,\gamma)}(x + \mu[h'(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)]) = x.$$

To see that this is true, consider the convex quadratic subproblem

$$\text{Minimize }_y \|y - [x + \mu[h'(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)]]\|^2$$

$$\text{subject to } y \in \Omega(x,\gamma)$$

and observe that $y = x$ satisfies the sufficient KKT optimality conditions with the multipliers $\lambda = -\mu h(x)$ and $\nu_i = \mu g_i(x)$ for $g_i(x) \ge 0$, $\nu_i = 0$ else. So, by (10), $\|P_{\Omega(x,\gamma)}(x - \nabla f(x)) - x\| \le \varepsilon$ as we wanted to prove. □

The following Corollary states the AGP property in a closer way to the algorithmic framework: given a local minimizer, a sequence that satisfies $x^k \to x^*$ and $d(x^k,\gamma) \to 0$ necessarily exists. As we mentioned in the Introduction, this property is naturally satisfied by many nonlinear programming algorithms [10, 17, 18], being $\{x^k\}$ the algorithmic sequence in that case.

Corollary 1. *If x^* is a local minimizer of (1)-(2) and $\gamma \in [-\infty, 0]$, there exists a sequence $\{x^k\} \subset \mathbb{R}^n$ such that $\lim x^k = x^*$ and $\lim d(x^k,\gamma) = 0$.*

Remark. In general, $d(x^*, \gamma) = 0$ does not hold at a local minimizer x^*. For example, in the problem of minimizing x subject to $x^2 \leq 0$ we have that $d(x^*, \gamma) \neq 0$ for all $\gamma < 0$.

In the following, we prove that, for all $\gamma < 0$ the AGP conditions are equivalent. This property is not true for $\gamma = 0$. In fact, consider the problem of minimizing x subject to $x \geq 0$, with $x^k = 1/k \ \forall \ k = 0, 1, 2, \ldots$ Clearly, $d(x^k, \gamma) \to 0$ for all $\gamma < 0$ but $d(x^k, 0)$ does not tend to zero.

Property 1. *Assume that* $g(x^*) \leq 0, h(x^*) = 0$, $x^k \to x^*$ *and, for some* $\gamma \in [-\infty, 0)$, $d(x^k, \gamma) \to 0$. *Then,* $d(x^k, \gamma') \to 0$ *for all* $\gamma' < 0$.
Proof. Consider the problems

$$\text{Minimize } \|x^k - \nabla f(x^k) - y\|^2 \tag{11}$$

subject to

$$g_i'(x^k)(y - x^k) \leq 0 \quad \text{if} \;\; g_i(x^k) \geq 0, \tag{12}$$

$$g_i'(x^k)(y - x^k) \leq 0 \quad \text{if} \;\; 0 > g_i(x^k) \geq \gamma, \tag{13}$$

$$h'(x^k)(y - x^k) = 0 \tag{14}$$

and

$$\text{Minimize } \|x^k - \nabla f(x^k) - y\|^2 \tag{15}$$

subject to

$$g_i'(x^k)(y - x^k) \leq 0 \quad \text{if} \;\; g_i(x^k) \geq 0, \tag{16}$$

$$g_i'(x^k)(y - x^k) \leq 0 \quad \text{if} \;\; 0 > g_i(x^k) \geq \gamma', \tag{17}$$

$$h'(x^k)(y - x^k) = 0. \tag{18}$$

Let y^k be the solution of (11–14). By the hypothesis, we know that $\|y^k - x^k\| \to 0$. Therefore, $y^k \to x^*$. Let us show that, for k large enough, y^k satisfies the constraints (16–18). In the case of (16) and (18) this is obvious. Let us prove that the same is true for (17). We consider two cases: $g_i(x^*) = 0$ and $g_i(x*) < 0$.

If $g_i(x^*) = 0$ then, for k large enough, $g_i(x^k) > \gamma$. Therefore, if, in addition, $g_i(x^k) < 0$, the constraint (17) is satisfied.

If $g_i(x^*) < 0$, then, since $\|y^k - x^k\| \to 0$, we have that, for k large enough,

$$g_i(x^k) + g_i'(x^k)(y^k - x^k) < 0. \tag{19}$$

Therefore, the constraints (17) are also satisfied at y^k.

But y^k is a KKT point of the problem (11–14). To prove that it is also a KKT point (and, thus, a solution) of (15–18) it only remains to prove that the active constraints of (15–18) are necessarily constraints that occur in

(11–14). Again, this is obvious in the case of (16) and (18). By the analysis performed above, an active constraint of type (17) can correspond only to $g_i(x^*) = 0$ and $g_i(x^k) < 0$. In this case, $g_i(x^k) > \gamma$ for k large enough and, thus, the constraint is present in the set (13). This completes the proof. □

We define now the Strong Approximate Gradient projection (SAPG) $d_S(x, \gamma)$. The SAPG vector is the APG vector related to problem (1)-(2), when each equality constraint $h_j(x) = 0$ is transformed into two inequality constraints in the obvious way. For all $x \in I\!R^n, \gamma \in [-\infty, 0]$, the set $\Omega_S(x, \gamma)$ is defined as the set $\Omega(x, \gamma)$ related to the two-inequality reformulation. For example, $\Omega_S(x, -\infty)$ is the set of points that satisfy

$$\begin{array}{rcll} g_i'(x)(z-x) & \leq & 0, & \text{if } g_i(x) \geq 0 \\ g(x) + g_i'(x)(z-x) & \leq & 0, & \text{if } g_i(x) \leq 0 \end{array}$$

and

$$\begin{array}{ll} 0 \leq h_j(x) + h_j'(x)(z-x) \leq h_j(x) & \text{if } h_j(x) \geq 0 \\ 0 \geq h_j(x) + h_j'(x)(z-x) \geq h_j(x) & \text{if } h_j(x) \leq 0. \end{array}$$

Accordingly,

$$d_S(x, \gamma) = P_{\Omega_S(x,\gamma)}(x - \nabla f(x)) - x.$$

We say that the Strong AGP property is fulfilled when, for some sequence, $\lim d_S(x^k, \gamma) = 0$. In the following Lemma, we prove that the Strong AGP property implies the AGP property.

Lemma 1. *For all $\gamma \in [-\infty, 0]$, if $x^k \to x^*$ and $d_S(x^k, \gamma) \to 0$ then $d(x^k, \gamma) \to 0$.*

Proof. Observe that $\Omega(x, \gamma) \subset \Omega_S(x, \gamma)$ for all $x \in I\!R^n$, $\gamma \in [-\infty, 0]$. Let us call

$$w^k = P_{\Omega_S(x^k,\gamma)}(x^k - \nabla f(x^k)),$$

$$y^k = P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)).$$

Since $x^k.y^k \in \Omega(x^k, \gamma) \subset \Omega_S(x^k, \gamma)$, we have by the definition of projections for all $t \in [0, 1]$,

$$\|x^k - \nabla f(x^k) - w^k\|^2 \leq \|(x^k - \nabla f(x^k)) - (x^k + t(y^k - x^k))\|^2$$

$$\leq \|t(x^k - \nabla f(x^k) - y^k) - (1-t)\nabla f(x^k)\|^2 \leq \|\nabla f(x^k)\|^2.$$

So,

$$\frac{1}{2}\|w^k - x^k\|^2 + (w^k - x^k)^T \nabla f(x^k) \leq \frac{t^2}{2}\|y^k - x^k\|^2 + t(y^k - x^k)^T \nabla f(x^k) \leq 0$$

for all $t \in [0,1]$.

Taking limits in the above inequalities and using that $\|w^k - x^k\| \to 0$, we obtain that, for all $t \in [0,1]$,

$$\lim_{k\to\infty} \frac{t^2}{2}\|y^k - x^k\|^2 + t(y^k - x^k)^T \nabla f(x^k) = 0.$$

Therefore, for all $t \in (0,1]$,

$$\lim_{k\to\infty} \frac{t}{2}\|y^k - x^k\|^2 + (y^k - x^k)^T \nabla f(x^k) = 0. \tag{20}$$

Since $\|y^k - x^k\|$ is bounded, this implies that

$$(y^k - x^k)^T \nabla f(x^k) = 0.$$

Therefore, by (20) with $t = 1$, we obtain that $\|y^k - x^k\| \to 0$, as we wanted to prove. □

Since the SAGP comes from a reformulation of the constraints (5), Theorem 1 can be applied and the following Corollary also holds.

Corollary 2. *If x^* is a local minimizer of (1)-(2) and $\gamma \in [-\infty, 0]$, there exists a sequence $\{x^k\} \subset I\!R^n$ such that $\lim x^k = x^*$ and $\lim d_S(x^k, \gamma) = 0$.*

We finish this section proving that the AGP condition implies the Fritz-John optimality conditions. As in the case of Theorem 1, the technique of proof is similar to the one of Proposition 3.3.5 of [2]. Let us recall the equivalent formulation of the Mangasarian-Fromovitz constraint qualification (see [23]).

Mangasarian-Fromovitz constraint Qualification (MF)

If $\lambda \in I\!R^m$ and $\mu \in I\!R^p, \mu \geq 0$ are such that

$$h'(x^*)^T \lambda + g'(x^*)^T \mu = 0 \quad \textit{and} \quad \mu^T g(x^*) = 0,$$

then $\lambda = 0$ and $\mu = 0$.

Theorem 2. *Assume that x^* is a feasible point. Let $\gamma \in [-\infty, 0]$. Suppose that there exists a sequence $x^k \to x^*$ such that $d(x^k, \gamma) \to 0$ or $d_S(x^k, \gamma) \to 0$. Then, x^* is a Fritz-John point of (1)-(2).*

Proof. Recall that, by Lemma 1, $d_S(x^k, \gamma) \to 0$ implies that $d(x^k, \gamma) \to 0$, so we only need to consider this case. Define $y^k = P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k))$.

So, y^k solves

$$\text{Minimize } \frac{1}{2}\|y - x^k\|^2 + \nabla f(x^k)^T (y - x^k)$$

$$\text{subject to } \; y \in \Omega(x^k, \gamma).$$

Therefore, there exist $\lambda^k \in I\!R^m$, $\mu^k \in I\!R^p$, $\mu^k \geq 0$, such that

$$\nabla f(x^k) + (y^k - x^k) + h'(x^k)^T \lambda^k + g'(x^k)^T \mu^k = 0, \tag{21}$$

$$\mu_i^k [g_i(x^k) + g_i'(x^k)(y^k - x^k)] = 0 \;\text{ if } \gamma < g_i(x^k) < 0,$$

$$\mu_i^k [g_i'(x^k)(y^k - x^k)] = 0, \;\text{ if } g_i(x^k) \geq 0,$$

$$\mu_i^k = 0 \text{ else.}$$

Moreover, if $g_i(x^*) < 0$, we have that $g_i(x^k) < 0$ for k large enough and, since $\|y^k - x^k\| \to 0$, we also have that $g_i(x^k) + g_i'(x^k)(y^k - x^k) < 0$ in that case. Therefore, we can assume that

$$\mu_i^k = 0 \;\text{ whenever }\; g_i(x^*) < 0. \tag{22}$$

To prove Fritz-John is equivalent to prove that MF implies KKT. So, we are going to assume from now on that x^* satisfies the Mangasarian-Fromovitz constraint qualification MF.

Suppose, by contradiction, that (λ^k, μ^k) is unbounded. Define, for each k,

$$M_k = \|(\lambda^k, \mu^k)\|_\infty = \max\{\|\lambda^k\|_\infty, \|\mu^k\|_\infty\}.$$

Then $\limsup M_k = \infty$. Refining the sequence (λ^k, μ^k) and reindexing it we may suppose $M_k > 0$ for all k and

$$\lim_k M_k = +\infty.$$

Now define

$$\hat{\lambda}^k = (1/M_k)\lambda^k, \quad \hat{\mu}^k = (1/M_k)\mu^k.$$

Observe that for all k, $\|(\hat{\lambda}^k, \hat{\mu}^k)\|_\infty = 1$. Hence, the sequence $(\hat{\lambda}^k, \hat{\mu}^k)$ is bounded and has a cluster point $(\hat{\lambda}, \hat{\mu})$ satisfying

$$\hat{\mu} \geq 0, \quad \|(\hat{\lambda}, \hat{\mu})\|_\infty = 1.$$

Dividing (21) by M_k we obtain

$$(1/M_k)[\nabla f(x^k) + (y^k - x^k)] + h'(x^k)^T \hat{\lambda}^k + g'(x^k)^T \hat{\mu}^k = 0.$$

Taking the limit along the appropriate subsequence, we conclude that

$$h'(x^*)^T \hat{\lambda} + g'(x^*)^T \hat{\mu} = 0.$$

Together with (22), this contradicts the constraint qualification MF.

Now, since in (21), λ^k and μ^k are bounded, extracting a convergent subsequence, we have that $\lambda^k \to \lambda^*$ and $\mu^k \to \mu^* \geq 0$. By (22), $g(x^*)^T \mu^* = 0$ and, taking limits in (21) the KKT condition follows. □

The following corollary states that, under the Mangasarian-Fromovitz constraint qualification, the AGP condition implies the standard KKT conditions of nonlinear programming. It is a trivial consequence of Theorem 2.

Corollary 3. *If the hypotheses of Theorem 2 hold and x^* satisfies MF, then the KKT necessary conditions are fulfilled at x^*.*

Remark. The set of Fritz-John points can be strictly larger than the set of points that satisfy AGP. Consider the nonlinear program given by

$$\text{Minimize } x \quad \text{subject to} \quad x^3 \leq 0.$$

Take $x^* = 0$. Clearly, x^* satisfies Fritz-John but, for any sequence $\{x^k\}$ that converges to x^*, $d(x^k, \eta)$ does not tend to zero. Of course, x^* does not satisfy the Mangasarian-Fromovitz constraint qualification.

3 Sufficiency in the convex case

As it is well known, in convex problems the KKT conditions imply optimality. In this section, we prove that, essentially, the same sufficiency property is true for the AGP optimality condition.

Theorem 3. *Suppose that, in the nonlinear program (1)-(2), f and g_i are convex, $i = 1, \ldots, p$ and h is an affine function. Let $\gamma \in [-\infty, 0]$. Suppose that $x^* \in \Omega$ and $\{x^k\} \subset I\!R^n$ are such that $\lim x^k = x^*$, $h(x^k) = 0$ for all $k = 0, 1, 2, \ldots$ and $\lim d(x^k, \gamma) = 0$. Then, x^* is a minimizer of (1)-(2).*

Proof. Let us prove first that $\Omega \subset \Omega(x^k, \gamma)$ for all $x^k \in I\!R^n$. Assume that $z \in \Omega$. If $g_i(z) \leq 0$ and $g_i(x^k) < 0$, we have, by the convexity of g_i, that

$$0 \geq g_i(z) \geq g_i(x^k) + g_i'(x^k)(z - x^k). \tag{23}$$

Moreover, if $g_i(z) \leq 0$ and $g_i(x^k) \geq 0$,

$$0 \geq g_i(z) \geq g_i(x^k) + g_i'(x^k)(z - x^k) \geq g_i'(x^k)(z - x^k). \tag{24}$$

Therefore, by (23) and (24), $z \in \Omega$ implies that $z \in \Omega(x^k, \gamma)$, so $\Omega \subset \Omega(x^k, \gamma)$. Note that (5) holds since $h(x^k) = h(z) = 0$ and h is affine.

Let us define now $y^k = P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k))$. Let $z \in \Omega$ be arbitrary. Since $z \in \Omega(x^k, \gamma)$ and y^k minimizes $\|x^k - \nabla f(x^k) - y\|^2$ on this set, we have that

$$\nabla f(x^k)^T(z - x^k) + \frac{1}{2}\|z - x^k\|^2 \geq \nabla f(x^k)^T(y^k - x^k) + \frac{1}{2}\|y^k - x^k\|^2.$$

Taking limits on both sides of the above inequality and using that $\|y^k - x^k\| \to 0$, we obtain:

$$\nabla f(x^*)^T(z - x^*) + \frac{1}{2}\|z - x^*\|^2 \geq 0.$$

Since Ω is convex the above inequality holds replacing z by $x^* + t(z - x^*)$ for all $t \in [0, 1]$, so:

$$t\nabla f(x^*)^T(z - x^*) + t^2\frac{1}{2}\|z - x^*\|^2 \geq 0.$$

Thus,

$$\nabla f(x^*)^T(z - x^*) + t\frac{1}{2}\|z - x^*\|^2 \geq 0.$$

for all $t \in (0, 1]$. Taking limits for $t \to 0$ we obtain that $\nabla f(x^*)^T(z-x^*) \geq 0$. Since $z \in \Omega$ was arbitrary, convexity implies that x^* is minimizer of (1)-(2). $\square$

Corollary 3. *Assume that f and $g_i, i = 1, \ldots, p$ are as in Theorem 3 and that there are no equality constraints at all. Suppose that $\lim x^k = x^*$ and $d(x^k, \gamma) \to 0$. Then x^* is a minimizer of (1)-(2).*

It is easy to show that the strong AGP property is a sufficient condition for minimizers of convex problems, without the requirement $h(x^k) = 0$. This is stated in the following theorem. Therefore, the strong AGP property is a necessary and sufficient optimality condition for convex problems.

Theorem 4. *Assume that, in the nonlinear program (1)-(2), f and g_i are convex, $i = 1, \ldots, p$ and h is an affine function. Let $\gamma \in [-\infty, 0]$. Suppose that $x^* \in \Omega$ and $\{x^k\} \subset I\!R^n$ are such that $\lim x^k = x^*$, and $\lim d_S(x^k, \gamma) = 0$. Then, x^* is a minimizer of (1)-(2).*

Proof. As we mentioned in Section 2, the strong AGP direction is the AGP direction for a reformulation of the problem that does not have equality constraints. Therefore, the thesis follows from Corollary 3. □

Remarks. Theorem 3 is not true if we do not assume that $h(x^k) = 0$ for all $k = 0, 1, 2, \ldots$ Consider the convex problem

$$\text{Minimize } x_1 \quad \text{subject to} \quad h(x_1, x_2) = 0 \;, \;\; g(x_1, x_2) \leq 0,$$

where

$$h(x_1, x_2) = x_2$$

and

$$g(x_1, x_2) = [\sqrt{(x_1^2 - 1)^2 + 4x_2^2} + x_1^2 - 1]^2.$$

The feasible set of this problem is $[-1, 1] \times \{0\}$. The sequence $x^k = (-1/k, 1/k)$ satisfies the AGP property but converges to the feasible point $(0, 0)$, which is not the minimizer of the problem. However, the strong AGP property is not satisfied by that sequence.

This fact has some algorithmic consequences: On one hand, for many iterative algorithms, to assume that $h(x^k) = 0$ for all k, when h is an affine function, is not a serious restriction. In fact, for those methods, linear constraints are naturally satisfied at every iteration, so that the sufficiency property holds.

On the other hand, Theorem 3 and the counterexample above, seem to indicate that, when we have an algorithm for (1)-(2) with the AGP property, it is better to formulate the problem using a pair of inequalities replacing each equality constraint. This has a practical interpretation: the approximate projected gradient is associated to an (explicit or implicit) phase of many nonlinear programming algorithms, where we try to improve the functional value on an approximation of the feasible set. This is called "the horizontal step" in many sequential quadratic programming algorithms [6] and "optimality phase" in some inexact-restoration methods [17, 18]. The two-inequality reformulation of $h_j(x) = 0$ allows one to improve, not only the functional value, but also the feasibility when performing that phase. In fact, roughly speaking, the domain in which the functional value is improved at iteration k has, up to second order, the feasibility of $h_j(x^k)$ if we use the equality formulation. But, using two inequalities, the feasibility of the new point with respect to h_j can be (with second order error) between $-h_j(x^k)$ and $h_j(x^k)$. Even truly feasible points can be included when, in the approximate region, each equation is described by two inequalities.

Of course, there is a strong reason for not using the two-inequality reformulation: subproblems are much more difficult than when we use equalities. For example, in sequential quadratic programming algorithms with only equality constraints, the linear algebra of each iteration can be reduced to one or two matrix factorizations, whereas the combinatorial aspect cannot be avoided when using inequalities. See, for example [6].

4 Conclusions

The optimality condition introduced in this paper is oriented to the analysis of optimization algorithms. On one hand, there exist *good* nonlinear programming methods satisfy the new condition. This is a question of simple verification in some cases. The condition is explicitly proved in inexact restoration algorithms [17, 18] and it can be easily verified in augmented Lagrangian methods. In other cases [22, 24] it could demand some detailed research analysis.

We conjecture that all reliable algorithms for nonlinear programming satisfy stronger theoretical properties than the one that says that feasible accumulation points are Fritz-John. The set of AGP points is a sharper approximation to the set of local minimizers than the set of Fritz-John points but, perhaps, there can be identified other sharp approximations to the minimizers that can be linked to the convergence of good minimization algorithms. This will be the subject of future research.

The mathematical programming problem with equilibrium constraints (MPEC) is a good example of a problem where the new condition is useful. In the usual formulation, the constraints take the form $x_i g_i = 0, \ x_i \geq 0, \ g_i \geq 0$ which implies that all the feasible points are Fritz-John. However, it can be shown that few points satisfy the AGP condition. See [1] for an analysis of the use of nonlinear algorithms for solving MPEC.

Further research is necessary in order to extend the new optimality condition to nonsmooth optimization, variational inequality problems, bilevel programming and vector optimization (see [3, 13, 14, 19, 20, 21]). As in the smooth case analyzed in this paper, the guiding line comes from having in mind what efficient algorithms for those problems do, what kind of convergence results have already been proved for them and which "good and obvious" practical behavior has been lost in those results.

Acknowledgement

We are indebted to Alfredo Iusem for his helpful comments on the first draft of this paper.

References

[1] R. Andreani and J. M. Martínez. On the solution of mathematical

programming problems with equilibrium constraints using nonlinear programming algorithms. Technical Report, Department of Applied Mathematics. University of Campinas, Brazil (2000).

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

[3] G. Bigi and M. Pappalardo. Regularity conditions in vector optimization. *Journal of Optimization Theory and Applications* 102, pp. 83-96 (1999).

[4] P. T. Boggs and J. W. Tolle. A strategy for global convergence in a sequential quadratic programming algorithm. *SIAM Journal on Numerical Analysis* 26, pp. 600-623 (1989).

[5] M. Castellani. Separation of sets and optimality conditions. *Rendiconti del Circolo Matematico di Palermo. Serie II* 48, pp. 27-38 (1997).

[6] J. E. Dennis, M. El-Alem and M. C. Maciel. A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM Journal on Optimization* 5, pp. 177-207 (1995).

[7] P. H. Dien, G. Mastroeni, M. Pappalardo and P. H. Quang. Regularity conditions for constrained extremum problems via image space: the nonlinear case. *Journal of Optimization Theory and Applications* 80, pp. 19-37 (1994).

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

[9] L. Fernandes, A. Friedlander, M. Guedes and J. Júdice . Solution of generalized linear complementarity problems using smooth optimization and its application to bilinear programming and LCP. To appear in *Applied Mathematics and Optimization* (1998).

[10] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming*, John Wiley, New York, 1968.

[11] F. Giannessi. General optimality conditions via a separation scheme. In *Algorithms for continuous optimization*, edited by E. Spedicato, Kluwer, The Netherlands, pp. 1-23, 1994.

[12] F. Giannessi. On connections among separation, penalization and regularization for variational inequalities with point-to-set operators. *Rendiconti del Circolo Matematico di Palermo, Serie II* 48, pp. 137-145 (1997).

[13] F. Giannessi, G. Mastroeni and L. Pellegrini. On the theory of vector optimization and variational inequalities. Image space analysis and separation. In *Vector Variational Inequalities and Vector Equilibria. Mathematical Theories*, edited by F. Giannessi, Kluwer, Dordrecht, pp. 153-216 (2000).

[14] F. Giannessi and L. Pellegrini. Image space analysis for vector optimization and variational inequalities. Scalarization. In *Advances in Combinatorial and Global Optimization.* Edited by A. Migadalas, P. Pardalos and R. Burkard. World Scientific, River Edge, 1999.

[15] F. Giannessi and T. Rapcsak. Images, separation of sets and extremum problems. In *Recent trends in optimization theory and applications*, edited by R. P. Agarwal, World Scientific Series in Applicable Analysis, World Scientific, Singapore, 1996.

[16] O. L. Mangasarian and S. Fromovitz. The Fritz-John necessary optimality conditions in presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications* 17, pp. 37-47 (1967)

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

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

[19] P. Favati and M. Pappalardo. On the reciprocal vector optimization problem. *Journal of Optimization Theory and Applications* 47, pp. 181-193 (1985).

[20] M. Pappalardo. On stationarity in vector optimization problems. *Rendiconti del Circolo Matematico di Palermo*, Serie II, Suppl. 48, pp. 195-201 (1997).

[21] M. Pappalardo. Vector optimization. In *Encyclopedia of Optimization.* Edited by P. Pardalos and C. Floudas. Kluwer, to appear.

[22] M. L. Psiaki and K. Park. Augmented Lagrangian nonlinear programming algorithms that uses SQP and trust region techniques. *Journal of Optimization Theory and Applications* 86, pp. 311-326 (1995).

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

[24] P. Spellucci. An SQP method for general nonlinear programs using only equality constrained subproblems. *Mathematical Programming* 82, pp. 413-447 (1998).

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