

On Sequential Optimality Conditions *

Roberto Andreani [†] Gabriel Haeser [‡] J. M. Martínez [§]

June 11, 2009.

Abstract

Sequential optimality conditions provide adequate theoretical tools to justify stopping criteria for nonlinear programming solvers. Approximate KKT and Approximate Gradient Projection conditions are analyzed in this work. These conditions are not necessarily equivalent. Implications between different conditions and counter-examples will be shown. Algorithmic consequences will be discussed.

Key words: Nonlinear Programming, Optimality Conditions, Constraint Qualifications, Stopping criteria.

AMS Subject Classification: 90C30, 49K99, 65K05.

1 Introduction

In this paper we study sequential first-order optimality conditions for Nonlinear Programming.

Necessary optimality conditions must be satisfied by the minimizers of optimization problems. Usually, the theorems that support an optimality condition are of the form: “If the local minimizer x satisfies CQ, then it satisfies KKT”, where KKT represents the Karush-Kuhn-Tucker conditions and CQ is a constraint qualification [6, 13, 27]. In other words, usual first-order necessary optimality conditions are of the form “KKT or not-CQ”.

On the other hand, practical methods for solving constrained optimization problems are iterative. One must decide, at each iteration, whether it is sensible to terminate the execution of the algorithm or not. Since testing true optimality is very difficult, the obvious idea is to terminate when a necessary optimality condition is approximately satisfied. However, most popular numerical optimization solvers do not test constraint qualifications at all, although (approximate) KKT conditions are always tested. Many practitioners seem not to be aware that constraint qualifications exist. This computational practice may be justified by a theoretical property of local minimizers: Roughly speaking, a local minimizer might not be KKT, but it can always be approximated by a sequence of “approximate-KKT” points.

This state of facts leads one to study a different type of optimality conditions. We say that x satisfies the “sequential optimality condition” defined by the mathematical proposition \mathcal{P} when

*This work was supported by PRONEX-CNPq/FAPERJ Grant E-26/171.164/2003 - APQ1, FAPESP Grants 05/02163/8 and 06/53768-0, and CNPq.

[†]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: andreani@ime.unicamp.br

[‡]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: ghaeser@ime.unicamp.br

[§]Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil. E-mail: martinez@ime.unicamp.br

there exists a sequence $\{x^k\}$ that converges to x and satisfies $\mathcal{P}(\{x^k\})$. Usually, a sequential optimality condition is associated with some quantity ϵ_k that must tend to zero. The natural termination criterion that corresponds to the sequential optimality condition indicates to stop the execution of the algorithm when ϵ_k is sufficiently small.

Sequential necessary optimality conditions are subject to the same requirements as ordinary (punctual) optimality conditions: They must be satisfied by the minimizers of the problem, and they should be as strong as possible. Moreover, useful (or algorithmically-oriented) optimality conditions should be associated with some practical algorithm.

This paper is organized as follows. Approximate KKT conditions are addressed in Section 2. In Section 3 we present AGP conditions and we state the relationships between these conditions and the ones analyzed in Section 2. Conclusions are given in Section 4.

Notation.

- $\mathbb{N} = \{0, 1, 2, \dots\}$.
- $\|\cdot\|$ denotes an arbitrary norm.
- If $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ we denote $\nabla h = (\nabla h_1, \dots, \nabla h_m)$.
- $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$.
- If $v \in \mathbb{R}^n$, we denote $v_+ = (\max\{v_1, 0\}, \dots, \max\{v_n, 0\})^T$.
- If $v \in \mathbb{R}^n$, we denote $v_- = (\min\{v_1, 0\}, \dots, \min\{v_n, 0\})^T$.
- $A \subset B$ means that the set A is contained in B .
- $B(x, \delta) = \{z \in \mathbb{R}^n \mid \|z - x\| \leq \delta\}$.
- $P_\Omega(x)$ is the Euclidean projection of x on Ω .

2 Approximate-KKT conditions

We consider the Nonlinear Programming problem in the form

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, g(x) \leq 0, \quad (1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are smooth.

Let $I \subset \{1, \dots, p\}$. We say that I satisfies the Sufficient Interior Property if for all feasible point x there exists a sequence of feasible points z^k such that $z^k \rightarrow x$ and $g_i(z^k) < 0$ for all $i \in I$. (Note that $I = \emptyset$ satisfies this property.) For simplicity, we will write $\text{AKKT} = \text{AKKT}(\emptyset)$ from now on.

Assume that I satisfies the Sufficient Interior Property. We say that the feasible point x^* fulfills the Approximate-KKT condition associated with I ($\text{AKKT}(I)$) if there exists a sequence $\{x^k\}$ that converges to x^* and satisfies:

- For all $k \in \mathbb{N}$ there exist $\lambda^k \in \mathbb{R}^m$, $\mu^k \in \mathbb{R}_+^p$ such that

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0, \quad (2)$$

$$\mu_i^k = 0 \text{ for all } i \text{ such that } g_i(x^*) < 0, \quad (3)$$

and

$$g_i(x^k) < 0 \text{ for all } i \in I. \quad (4)$$

Lemma 2.1. *A feasible point x^* satisfies AKKT(I) if, and only if, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$, $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $x^k \rightarrow x^*$, $\varepsilon_k \rightarrow 0$,*

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon_k, \quad (5)$$

$$\mu_i^k = 0 \text{ for all } i \text{ such that } g_i(x^k) < -\varepsilon_k, \quad (6)$$

and

$$g_i(x^k) < 0 \text{ for all } i \in I. \quad (7)$$

Proof. Assume first that $x^k \rightarrow x^*$ and $\{x^k\}, \{\lambda^k\}, \{\mu^k\}$ satisfy (2,3,4). Define

$$J = \{j \in \{1, \dots, p\} \mid g_j(x^*) = 0\}.$$

Define, for all $k \in \mathbb{N}$,

$$\varepsilon_k = \max\{\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\|, -g_j(x^k), j \in J\}. \quad (8)$$

By (2), the continuity of g and the definition of J , we have that $\varepsilon_k \rightarrow 0$ and (5) holds. Moreover, (4) is identical to (7).

Now, if $g_i(x^k) < -\varepsilon_k$, we have that $-g_i(x^k) > \varepsilon_k$ and, by (8), $-g_i(x^k) > -g_j(x^k)$ for all $j \in J$. Therefore, $i \notin J$, so $g_i(x^*) < 0$. Then, by (3), $\mu_i^k = 0$ and, so, (6) holds. This completes the first part of the proof.

Now, let us assume that there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$, $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $x^k \rightarrow x^*$, $\varepsilon_k \rightarrow 0$ and (5,6,7) take place. By (5) and (7), we have that (2) and (4) hold. Suppose now that $g_i(x^*) < 0$. Then, for k large enough, since $x^k \rightarrow x^*$ and g_i is continuous, we have that $g_i(x^k) < -\varepsilon_k$. Then, by (6), $\mu_i^k = 0$. Thus, (3) also holds. \square

Lemma 2.1 provides a natural stopping criterion associated with AKKT(I). Given small positive tolerances $\varepsilon_{feas}, \varepsilon_{opt}, \varepsilon_{comp}$ associated with feasibility, optimality (5) and complementarity (6), an algorithm that aims to solve (1) preserving $g_i(x^k) < 0$ for all $i \in I$ (7) should be stopped when

$$\|h(x^k)\| \leq \varepsilon_{feas}, \|g(x^k)_+\| \leq \varepsilon_{feas}, \quad (9)$$

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon_{opt}, \quad (10)$$

and

$$\mu_i^k = 0 \text{ whenever } g_i(x^k) < -\varepsilon_{comp}, \quad (11)$$

for suitable multipliers $\lambda^k \in \mathbb{R}^m$, $\mu^k \in \mathbb{R}_+^p$. This is the stopping criterion employed in Algencan [1]¹. and other Augmented Lagrangian algorithms [8]. Note that (11) can be forced, imposing that $\mu_i^k = 0$ whenever $g_i(x^k) < -\varepsilon_{comp}$ even if $\mu_i^k = 0$ is not the Lagrange multiplier suggested by

¹Algencan is the nonlinear programming solver available in the Tango project www.ime.usp.br/~egbirgin/tango.

the algorithm. After forcing $\mu_i^k = 0$ in those cases, we must test (10). Stopping should be decided when both (9) and (10) hold.

Clearly, if $I_{small} \subset I_{big}$ one has that $\text{AKKT}(I_{big})$ implies $\text{AKKT}(I_{small})$. The reciprocal is not true, as the following counter-example shows.

Counter-example 2.1. $\text{AKKT}(I_{small})$ does not imply $\text{AKKT}(I_{big})$.

Consider the problem

$$\text{Minimize } x \text{ subject to } g_1(x) \equiv -x^2 \leq 0, g_2(x) \equiv x \leq 0$$

and the point $x^* = 0$. Define $I_{small} = \{1\}, I_{big} = \{1, 2\}$. Clearly, both I_{small} and I_{big} satisfy the Sufficient Interior Property. The condition $\text{AKKT}(I_{small})$ is satisfied: Take $x^k = 1/k, \mu_1^k = k/2, \mu_2^k = 0$ for all $k \in \mathbb{N}$.

However, $\text{AKKT}(I_{big})$ does not hold. In fact, if $g_2(x^k) < 0$ one has that $x^k < 0$. So, $-2x^k \mu_1^k + \mu_2^k \geq 0$ and $1 + (-2x^k) \mu_1^k + \mu_2^k$ cannot tend to zero.

2.1 AKKT(I) is an optimality condition

We are going to prove that, if x^* is a local minimizer of (1) and I satisfies the Sufficient Interior Property, then x^* satisfies $\text{AKKT}(I)$. The proof is based on the convergence properties of the Internal-External Penalty method given below.

Lemma 2.2. *Let $\{\rho_k\}$ be a positive sequence that tends to infinity. Let $\Omega \subset \mathbb{R}^n$ be closed. Assume that, for all $k \in \mathbb{N}$, x^k is a global solution of*

$$\text{Minimize } f(x) + \rho_k \left[\sum_{i=1}^m h_i(x)^2 + \sum_{i \notin I_1} g_i(x)_+^2 \right] - \frac{1}{\rho_k} \sum_{i \in I_1} \frac{1}{g_i(x)}$$

subject to

$$g_i(x) < 0 \quad \forall i \in I_1, x \in \Omega.$$

Consider the problem

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0, g(x) \leq 0, x \in \Omega \tag{12}$$

and assume that I_1 is such that for every global minimizer of (12), there exists a feasible sequence $\{z^k\} \subset \Omega$ that converges to x and satisfies $g_i(z^k) < 0$ for all $i \in I_1$. Then, every limit point of x^k is a global minimizer of $f(x)$ subject to $h(x) = 0, g(x) \leq 0, x \in \Omega$.

Proof. The proof of this lemma is a slight variation of the proof of Theorem 4.3.10 of [11]. □

It is interesting to observe that the thesis of Lemma 2.2 does not hold under a weaker hypothesis on I_1 . If we only assume that every feasible point can be approximated by a sequence such that $g_i(x^k) < 0$ for all $i \in I_1$ (x^k not necessarily feasible) the convergence of the interior-exterior penalty method may not occur. Take, for example, the problem of minimizing x subject to $x(x-1) = 0, -x^3 \leq 0$. The global minimizer is $x^* = 0$ but the interior-exterior penalty method converges to 1. See [19] for details.

Theorem 2.1. *Let x^* be a local minimizer of (1) and assume that $I \subset \{1, \dots, p\}$ satisfies the Sufficient Interior Property. Then, x^* satisfies AKKT(I).*

Proof. Let $\delta > 0$ be such that $f(x^*) \leq f(x)$ for all feasible x such that $\|x - x^*\| \leq \delta$. Consider the problem

$$\text{Minimize } f(x) + \|x - x^*\|_2^2 \text{ subject to } h(x) = 0, g(x) \leq 0, x \in B(x^*, \delta). \quad (13)$$

Clearly, x^* is the unique solution of (13). Let x^k be a solution of

$$\text{Minimize } f(x) + \|x - x^*\|_2^2 + \rho_k (\|h(x)\|_2^2 + \sum_{i \notin I} g_i(x)_+^2) - \frac{1}{\rho_k} \sum_{i \in I} \frac{1}{g_i(x)}$$

subject to $g_i(x) < 0$ for all $i \in I$ and $x \in B(x^*, \delta)$. By the compactness of $B(x^*, \delta)$ and standard arguments of barrier methods [11], x^k is well defined for all k . By the Sufficient Interior Property, the hypotheses of Lemma 2.2 are fulfilled. Therefore, the sequence $\{x^k\}$ converges to x^* . For k large enough, one has that $\|x^k - x^*\| < \delta$, therefore, the gradient of the objective function must vanish. Thus,

$$\nabla f(x^k) + 2(x^k - x^*) + \sum_{i=1}^m 2\rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{i \notin I} 2\rho_k g_i(x^k)_+ \nabla g_i(x^k) + \sum_{i \in I} \frac{1}{\rho_k g_i(x^k)^2} \nabla g_i(x^k) = 0$$

and

$$g_i(x^k) < 0 \text{ for all } i \in I, k \in \mathbb{N}. \quad (14)$$

Let us write $\lambda_i^k = 2\rho_k h_i(x^k)$ for all $i = 1, \dots, m$, $\mu_i^k = 2\rho_k g_i(x^k)_+$ if $i \notin I$, $\mu_i^k = \frac{1}{\rho_k g_i(x^k)^2}$ if $i \in I$.

By (14) we have that (4) holds.

Since $\|x^k - x^*\| \rightarrow 0$ we have that

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k) + \nabla h(x^k) \lambda^k + \nabla g(x^k) \mu^k\| = 0. \quad (15)$$

If $g_i(x^*) < 0$, then, for k large enough, $g_i(x^k) < 0$. Then, if $i \notin I$, we have that $g_i(x^k)_+ = 0$ and $\mu_i^k = 0$. If $g_i(x^*) < 0$ and $i \in I$ we have that $\frac{1}{\rho_k g_i(x^k)^2} \rightarrow 0$. Thus, in (15) one has that $\mu_i^k \rightarrow 0$. By the continuity of g_i , (15) continues to be true re-defining $\mu_i^k = 0$ for $i \in I$, $g_i(x^*) < 0$. \square

2.2 AKKT(I) is a strong optimality condition

In this section we prove that AKKT(I) implies ‘‘KKT or not-CPLD’’, where CPLD is the Constant Positive Linear Dependence constraint qualification.

We say that the feasible point x fulfills the CPLD condition if the following property is satisfied: If $i_1, \dots, i_q \in \{1, \dots, m\}$ and $j_1, \dots, j_r \in \{1, \dots, p\}$ are such that $g_{j_\ell}(x) = 0, \ell = 1, \dots, r$ and the gradients $\nabla h_{i_1}(x), \dots, \nabla h_{i_q}(x), \nabla g_{j_1}(x), \dots, \nabla g_{j_r}(x)$ are linearly dependent with nonnegative coefficients corresponding to the gradients of inequalities, then there exists a neighborhood V of x such that $\nabla h_{i_1}(z), \dots, \nabla h_{i_q}(z), \nabla g_{j_1}(z), \dots, \nabla g_{j_r}(z)$ are linearly dependent for all $z \in V$. The CPLD condition was introduced in [28] and its status as a constraint qualification was elucidated in [5].

Every local minimizer that satisfies CPLD necessarily fulfills the KKT conditions [5]. This means that ‘‘KKT or not-CPLD’’ is a necessary optimality condition. This condition is fulfilled by any feasible limit point of Algencon [1]. Since CPLD is weaker than the Mangasarian-Fromovitz

constraint qualification (MFCQ) [23], the optimality condition “KKT or not-CPLD” is stronger than the Fritz-John conditions, which can be expressed in the form “KKT or not-MFCQ” [29].

Here we are going to prove that $\text{AKKT}(I)$ is strictly stronger than “KKT or not-CPLD”. Before proving that $\text{AKKT}(I)$ implies “KKT or not-CPLD” let us show that the reciprocal is not true.

Counter-example 2.2. “KKT or not-CQ” does not imply AKKT.

Recall that every local minimizer x^* that satisfies a constraint qualification necessarily fulfills the KKT conditions. Consider a nonlinear programming problem whose feasible set is $\{x \in \mathbb{R}^2 \mid x_1^2 = 0\}$. No feasible point satisfies any constraint qualification. To verify this, consider the objective function $f_1(x_1, x_2) = x_1$. Although all the feasible points are minimizers, the gradient of f_1 is never a linear combination of the constraint gradient, therefore the local minimizers are not KKT points. This means that, independently of the objective function and the constraint qualification, all the feasible points satisfy “KKT or not-CQ”. Now, consider the objective function defined by $f(x_1, x_2) = x_2$. Since $\nabla f(x) = (0, 1)^T$ for all x and $\nabla h(x)$ is a multiple of $(1, 0)^T$ for all x , it turns out that $\|\nabla f(x) + \lambda \nabla h(x)\|$ is bounded away from zero for all x . Therefore, $\|\nabla f(x^k) + \lambda^k \nabla h(x^k)\|$ cannot tend to zero. Thus, no feasible point satisfies AKKT.

Theorem 2.2. $\text{AKKT}(I)$ implies “KKT or not-CPLD”.

Proof. Assume that x^* satisfies $\text{AKKT}(I)$ and CPLD. Therefore, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$, $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $x^k \rightarrow x^*$, $\varepsilon_k \rightarrow 0$ satisfying (5) and (6). (We do not need to use (7) at all for this proof.) Therefore, $x^k, x^*, \lambda^k, \mu^k$ satisfy the conditions used in Theorem 4.5 of [1] in the context of proving KKT for Algencan. Thus, we can reproduce the arguments of that theorem to prove that x^* satisfy KKT. \square

3 Approximate Gradient Projection conditions

Approximate Gradient Projection (AGP) conditions were introduced in [26], where the authors observed that AGP is the optimality condition that fits the natural stopping criterion for Inexact Restoration methods [7, 12, 16, 17, 21, 22, 24, 25].

Let $\gamma \in (0, \infty]$. We say that a feasible point x^* of (1) satisfies the $\text{AGP}(\gamma)$ condition introduced in [26] when there exists a sequence $\{x^k\}$ that tends to x^* and satisfies

$$\lim_{k \rightarrow \infty} \|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0, \quad (16)$$

where Ω_k is the set of points $x \in \mathbb{R}^n$ defined by:

$$\nabla h_i(x^k)^T(x - x^k) = 0 \text{ for all } i = 1, \dots, m, \quad (17)$$

$$\nabla g_j(x^k)^T(x - x^k) \leq 0 \text{ for all } j \mid g_j(x^k) \geq 0, \quad (18)$$

and

$$g_j(x^k) + \nabla g_j(x^k)^T(x - x^k) \leq 0 \text{ for all } j \mid -\gamma < g_j(x^k) < 0. \quad (19)$$

Martínez and Svaiter [26] proved that $\text{AGP}(\gamma)$ is an optimality condition (every local minimizer satisfy it) and that $\text{AGP}(\gamma)$ is equivalent to $\text{AGP}(\gamma')$ for all $\gamma, \gamma' \in (0, \infty]$. For this reason we will always write AGP instead of $\text{AGP}(\gamma)$. In [26] it was also proved that AGP implies the Fritz-John condition (KKT or not-MFCQ). The stronger result that AGP implies “KKT or not-CPLD” seems to be proved for the first time in [16].

If $\{x^k\}$ is a sequence generated by an optimization algorithm, the natural stopping criterion associated with AGP is given by (9) and

$$\|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\| \leq \varepsilon_{opt}.$$

It is easy to prove that AGP implies AKKT [30]. Surprisingly, the reciprocal is not true, as the following counter-example shows. Therefore, AGP is a stronger optimality condition than AKKT.

Counter-example 3.1. AKKT does not imply AGP.

Consider the problem

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

where

$$\begin{aligned} f(x_1, x_2) &= -x_2, \\ h(x_1, x_2) &= x_1 x_2, \\ g(x_1, x_2) &= -x_1. \end{aligned}$$

Define $x^* = (0, 1)^T$. Let us show first that x^* does not satisfy AGP.

Assume that $x^k \rightarrow x^*$. If $x_1^k > 0$, the set Ω_k defined by (17,18,19) is the intersection of the half-space $x_1 \geq 0$ with the tangent line to $h(x_1, x_2) = h(x_1^k, x_2^k)$ that passes through x^k . This line tends to be vertical when x^k approaches x^* . Therefore, $P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k$ tends to $(0, 1)^T$. Analogously, if $x_1^k < 0$, the set Ω_k is the half-space $x_1 \geq x_1^k$ intersected with the tangent line to $h(x_1, x_2) = x_1^k x_2^k$ that passes through x^k . So, $P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k$ tends to $(0, 1)^T$. Therefore, for any sequence $x^k \rightarrow x^*$, $\|P_{\Omega_k}(x^k - \nabla f(x^k)) - x^k\|$ cannot tend to zero. As a consequence, x^* does not satisfy AGP.

Now, let us show that x^* satisfies AKKT. Define $x^k = (1/k, 1)^T$, $\lambda^k = \mu^k = k$, therefore, for all $k \in \mathbb{N}$ we have $\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k = (0, -1)^T + (1, 1/k)^T k + (-1, 0)^T k = 0$. Since $g(x^*) = 0$, we have that (2-4) hold, so x^* satisfies AKKT.

A generalization of the AGP condition has been used in the literature in the context of Inexact Restoration methods [12, 17, 24] and mathematical programming with complementarity constraints [2] without mentioning that this generalization is not equivalent to the original AGP condition given in [26]. Roughly speaking, the generalization consists on including some of the linear constraints of the problem (1) in the definition of Ω_k , imposing that these constraints must be satisfied by x^k for all $k \in \mathbb{N}$. During several years, authors seemed to believe that all these sequential optimality conditions (including AKKT) were equivalent. We will see here that this is not the case.

3.1 C-AGP condition

Assume that the functions g_{q+1}, \dots, g_p in (1) are convex and h_{r+1}, \dots, h_m are affine. Therefore, the set Ω defined by $g_j(x) \leq 0$ for $j = q+1, \dots, p$ and $h_i(x) = 0, i = r+1, \dots, m$ is closed and convex. We say that x^* satisfies the C-AGP condition if there exists a sequence $\{x^k\} \subset \Omega$ that tends to x^* and satisfies

$$\lim_{k \rightarrow \infty} \|P_{\Omega_k \cap \Omega}(x^k - \nabla f(x^k)) - x^k\| = 0, \quad (20)$$

where Ω_k is defined by (17,18,19) for $i = 1, \dots, r, j = 1, \dots, q$.

If an optimization algorithm generates iterates $x^k \in \Omega$, the natural stopping criterion associated with C-AGP requires (9) and

$$\|P_{\Omega_k \cap \Omega}(x^k - \nabla f(x^k)) - x^k\| \leq \varepsilon_{opt}.$$

Let us prove that, under the condition that the constraints in Ω satisfy some constraint qualification, C-AGP is an optimality condition.

Theorem 3.1. *Let x^* be a local minimizer of (1) and assume that, for all $x \in \Omega$, the constraints $g_i(x) \leq 0, i = q + 1, \dots, p, h_i(x) = 0, i = r + 1, \dots, m$ satisfy some constraint qualification. Then, x^* satisfies C-AGP.*

Proof. We use the technique employed in [26] for proving that AGP is an optimality condition. Let $\delta > 0$ be such that x^* is a global minimizer of (1) with the additional constraint $\|x - x^*\| \leq \delta$. Therefore, x^* is the unique global minimizer of $f(x) + \frac{1}{2}\|x - x^*\|_2^2$ subject to the constraints of (1) and $\|x - x^*\| \leq \delta$. Assume that $\rho_k \rightarrow \infty$ and let x^k be a global minimizer of $f(x) + \frac{1}{2}\|x - x^*\|_2^2 + \rho_k(\sum_{i=1}^r h_i(x)^2 + \sum_{i=1}^q g_i(x)_+^2)$ subject to $x \in \Omega$ and $\|x - x^*\| \leq \delta$. By the theory of convergence of external penalty methods, since x^* is the unique global minimizer, it turns out that $x^k \rightarrow x^*$. So, $\|x^k - x^*\| < \delta$ for k large enough. Since Ω satisfies a constraint qualification, it turns out that the KKT conditions of the subproblem must hold. Let

$$A_k = \{i \in \{q + 1, \dots, p\} \mid g_i(x^k) = 0\}.$$

Therefore, for k large enough there exist $\{\mu_i^k, i \in A_k\}, \lambda_i^k, i = r + 1, \dots, m$ such that

$$\begin{aligned} \nabla f(x^k) + 2\rho_k \left[\sum_{i=1}^r h_i(x^k) \nabla h_i(x^k) + \sum_{i=1}^q g_i(x^k)_+ \nabla g_i(x^k) \right] + (x^k - x^*) \\ + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A_k} \mu_i^k \nabla g_i(x^k) = 0. \end{aligned}$$

Thus,

$$\begin{aligned} (x^k - \nabla f(x^k)) - [x^k + 2\rho_k \left[\sum_{i=1}^r h_i(x^k) \nabla h_i(x^k) + \sum_{i=1}^q g_i(x^k)_+ \nabla g_i(x^k) \right] \\ + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A_k} \mu_i^k \nabla g_i(x^k)] = x^k - x^*. \end{aligned}$$

By the non-expansion property of projections, we deduce that:

$$\begin{aligned} \|P_{\Omega_k \cap \Omega}(x^k - \nabla f(x^k)) - P_{\Omega_k \cap \Omega}[x^k + 2\rho_k \left[\sum_{i=1}^r h_i(x^k) \nabla h_i(x^k) + \sum_{i=1}^q g_i(x^k)_+ \nabla g_i(x^k) \right] \\ + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A_k} \mu_i^k \nabla g_i(x^k)]\|_2 \leq \|x^k - x^*\|_2. \end{aligned}$$

But, by the definition of x^k , writing the optimality conditions of the projection minimization problem, we obtain:

$$\begin{aligned} P_{\Omega_k \cap \Omega}[x^k + 2\rho_k \left[\sum_{i=1}^r h_i(x^k) \nabla h_i(x^k) + \sum_{i=1}^q g_i(x^k)_+ \nabla g_i(x^k) \right] \\ + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A_k} \mu_i^k \nabla g_i(x^k)] = x^k. \end{aligned}$$

Therefore,

$$\|P_{\Omega_k \cap \Omega}(x^k - \nabla f(x^k)) - x^k\|_2 \leq \|x^k - x^*\|_2 \rightarrow 0.$$

This completes the proof. \square

Our next result shows that C-AGP is a strong optimality condition in the sense that it implies “KKT or not-MFCQ”.

Theorem 3.2. *Assume the feasible point x^* satisfies C-AGP and the Mangasarian-Fromovitz constraint qualification. Then, x^* satisfies the KKT conditions.*

Proof. In order to simplify the notation, we consider here $r = m$. The case $r < m$ follows straightforwardly. By the C-AGP condition, there exists a sequence $\{x^k\}$ such that $x^k \rightarrow x^*$ and $\|y^k - x^k\| \rightarrow 0$, where y^k is the solution of

$$\text{Minimize } \frac{1}{2} \|y - x^k + \nabla f(x^k)\|_2^2 \quad (21)$$

subject to

$$\nabla h(x^k)^T (y - x^k) = 0, \quad (22)$$

$$g_i(x^k)_- + \nabla g_i(x^k)^T (y - x^k) = 0, \quad i = 1, \dots, q, \quad (23)$$

$$g_j(y) \leq 0, \quad j = q + 1, \dots, p. \quad (24)$$

Observe first that, since $x^k \rightarrow x^*$ and $\|y^k - x^k\| \rightarrow 0$, we have:

$$\lim_{k \rightarrow \infty} y^k = x^*. \quad (25)$$

Assume that $i \leq q$ is such that $g_i(x^*) < 0$. Then, there exists $c > 0$ such that $g_i(x^k)_- < -c < 0$ for k large enough. Thus, since $\|y^k - x^k\| \rightarrow 0$, for k large enough we have:

$$g_i(x^k)_- + \nabla g_i(x^k)^T (y^k - x^k) < -c/2 < 0. \quad (26)$$

Assume now that $j \in \{q + 1, \dots, p\}$ is such that $g_j(x^*) < 0$. Then, by (25),

$$g_j(y^k) < 0 \quad (27)$$

for k large enough.

By (26) and (27), for k large enough, the indices of the active constraints of (21–24) at the y^k are contained in the set of indices of active constraints of (1) at x^* .

Assume now that $\underline{\lambda}^k \in \mathbb{R}^m$, $\underline{\mu}_1^k, \dots, \underline{\mu}_p^k \in \mathbb{R}_+$ are such that

$$\nabla h(y^k) \underline{\lambda}^k + \sum_{i=1}^q \underline{\mu}_i^k \nabla g_i(x^k) + \sum_{j=q+1}^p \underline{\mu}_j^k \nabla g_j(y^k) = 0, \quad (28)$$

with $\underline{\mu}_i^k = 0$ if the constraint (23) is not active at y^k and $\underline{\mu}_j^k = 0$ if the constraint (24) is not active at y^k . Assume, moreover, that for infinitely many indices k at least one of the coefficients $\underline{\lambda}_1^k, \dots, \underline{\lambda}_m^k, \underline{\mu}_1^k, \dots, \underline{\mu}_p^k$ is non-null. Then, dividing (28) by the maximum modulus of the coefficients, we may assume, without loss of generality, that the maximum modulus of the coefficients in (28) is 1 for all k . Then, using compactness and taking limits for $k \rightarrow \infty$ in (28) we obtain

$$\nabla h(x^*) \underline{\lambda} + \sum_{i=1}^q \underline{\mu}_i \nabla g_i(x^*) + \sum_{j=q+1}^p \underline{\mu}_j \nabla g_j(x^*) = 0,$$

where $\underline{\mu}_1, \dots, \underline{\mu}_p \geq 0$, at least one of the coefficients is non-null and, for all $i = 1, \dots, p$, $\underline{\mu}_i = 0$ if $g_i(x^*) < 0$. This is not possible since, by hypotheses, x^* satisfies MFCQ. Therefore, the existence of $\underline{\lambda}^k \in \mathbb{R}$, $\underline{\mu}_1^k, \dots, \underline{\mu}_p^k \in \mathbb{R}_+$ satisfying (28) is not possible. This means that, for all k large enough, y^k satisfies the Mangasarian-Fromovitz constraint qualification corresponding to the problem (21–24). It turns out that, for all k large enough, the KKT conditions of (21–24) are fulfilled.

Therefore, for k large enough, there exist $\lambda^k \in \mathbb{R}^m$, $\mu^k \in \mathbb{R}_+^p$ such that

$$y^k - x^k + \nabla f(x^k) + \nabla h(x^k)\lambda^k + \sum_{i \in I_k} \mu_i^k \nabla g_i(x^k) + \sum_{j \in J_k} \mu_j^k \nabla g_j(y^k) = 0, \quad (29)$$

where I_k and J_k are the indices of active inequality constraints at y^k . Above we proved that $I_k \subset I_*$ and $J_k \subset J_*$, where I_*, J_* are the indices of active inequality constraints at x^* for problem (1). If the sequences $\{\lambda^k\}$ and $\{\mu^k\}$ are bounded, then, taking convergent subsequences and taking limits in (29) we arrive to the KKT conditions at x^* .

If at least one of the sequences $\{\lambda^k\}$, $\{\mu^k\}$ is unbounded, the maximum element M_k of $|\lambda_i^k|$, $i = 1, \dots, m$, μ_j^k , $j = 1, \dots, p$ tends to infinity along some subsequence. So, dividing both members of (29) by M_k , we get:

$$(y^k - x^k + \nabla f(x^k))/M_k + \nabla h(x^k)\lambda^k/M_k + \sum_{i \in I_k} \mu_i^k/M_k \nabla g_i(x^k) + \sum_{j \in J_k} \mu_j^k/M_k \nabla g_j(y^k) = 0, \quad (30)$$

Taking limits along convergent subsequences in (30) we obtain:

$$\nabla h(x^*)\lambda^* + \nabla g(x^*)\mu^* = 0,$$

where $\mu^* \geq 0$ and $\|\lambda^*\| + \|\mu^*\| > 0$. This means that x^* does not satisfy MFCQ, which contradicts the hypothesis. \square

Counter-example 3.2. AGP does not imply C-AGP.

Consider the problem of minimizing x_2 subject to the constraints of the counter-example 3.1. Let us show that the point $x^* = (0, 1)^T$ does not satisfy C-AGP. If $\{x^k\}$ is such that $x_1^k \geq 0$ for all k , the same argument used in counter-example 3.1 may be used to show that (20) cannot hold.

However, x^* satisfies AGP. To see this, consider the sequence $x^k = (-1/k, 1)^T$. In this case, the projection of $x^k - \nabla f(x^k)$ on the set defined by $\nabla h(x^k)^T(x - x^k) = 0$ and $\nabla g(x^k)^T(x - x^k) \leq 0$ is equal to x^k for all $k \in \mathbb{N}$. Therefore, AGP holds.

The result above encourages one to conjecture that C-AGP also satisfies “KKT or not-CPLD”, as is the case of the original AGP condition. Surprisingly, this is not true, as the following example shows.

Counter-example 3.3. C-AGP does not imply “KKT or not-CPLD”.

Consider the problem (1) with $n = 2, p = 2, q = 1$, $f(x_1, x_2) = x_1$, $g_1(x_1, x_2) = -x_1^2 - x_2$, $g_2(x_1, x_2) = x_1^2 + x_2$. The function g_2 is obviously convex and the set of points such that $g_2(x_1, x_2) \leq 0$ clearly satisfy standard constraint qualifications. It is easy to see that the CPLD condition is fulfilled at $x^* = (0, 0)^T$ since $\nabla g_1(x) + \nabla g_2(x) = 0$ for all x . On the other hand, the KKT conditions do not hold at x^* . Let us show, however, that C-AGP is satisfied. We define, for all $k \in \mathbb{N}$, $x^k = x^*$. Then, $\nabla f(x^k) = (1, 0)^T$ for all k and $x^k - \nabla f(x^k) = (-1, 0)^T$. Now, the set Ω_k is $\{x \in \mathbb{R}^2 \mid \nabla g_2(x^k)^T(x - x^k) \leq 0\}$, so Ω_k is the half-plane $x_2 \geq 0$. This implies that $\Omega_k \cap \Omega = \{(0, 0)^T\} = \{x^k\}$ for all k . Therefore, $\|P_{\Omega \cap \Omega_k}(x^k - \nabla f(x^k)) - x^k\| = 0$ for all k . This

means that the C-AGP condition is satisfied at x^* . (Recall that this is not in contradiction with Theorem 3.2 because x^* does not satisfy MFCQ.)

Counter-example 3.4. C-AGP does not imply AGP.

The counter-example that shows that C-AGP does not imply “KKT or not-CPLD” may also be used to show that C-AGP does not imply AGP. In fact, if the point x^* satisfies AGP, since AGP implies “KKT or not-CPLD”, this optimality condition would hold at x^* .

3.2 L-AGP Condition

The independence of C-AGP with respect to AGP and the fact that C-AGP does not imply “KKT or not-CPLD” induces one to think that AGP is, essentially, the strongest sequential optimality condition that can be achieved by numerical optimization algorithms. However, a stronger AGP-like condition may be used in a very common situation: When some of the constraints that define the feasible set are linear. We say that a feasible point x^* satisfies the L-AGP condition if it satisfies the C-AGP condition in the case that g_{q+1}, \dots, g_p are affine functions. The status of L-AGP can be deduced from observations already made in this paper. The counter-example 3.2 may be used to show that AGP does not imply L-AGP. On the other hand, if a point x^* satisfies L-AGP, the corresponding sequence $\{x^k\}$ may be used to show that AGP also holds. In other words, L-AGP is strictly stronger than AGP. This supports the point of view that, if an optimization problem possesses linear constraints, it is sensible to preserve feasibility with respect to them, declaring convergence when the AGP criterion holds with some tolerance. On the other hand, using the same criterion with respect to general convex constraints does not seem to have special advantages.

It is worth mentioning that, in [2], the L-AGP condition was used (with the name AGP) in connection to mathematical programming problems with equilibrium constraints. In that paper it was shown that, if an algorithm that theoretically converges to L-AGP points goes to a feasible nondegenerate point, then this point is KKT.

4 Final remarks

Assume that one has a nonlinear programming problem with a set I_{lin} of linear constraints, a different set I_{conv} of convex constraints and a third set I_{gen} of general constraints. Let us say that C-AGP is satisfied when (20) holds defining $\{r+1, \dots, m\} \cup \{q+1, \dots, p\} = I_{conv}$ and that L-AGP holds when one defines $\{r+1, \dots, m\} \cup \{q+1, \dots, p\} = I_{lin}$. Then, the main results of this paper can be visualized in Figure 1. In this figure CQ represents any constraint qualification weaker than CPLD. Roughly speaking, Figure 1 shows that L-AGP is the strongest first-order optimality condition that may be used in practical algorithms.

The results of this paper may have a practical application in the design of novel optimization algorithms. We have in mind recent sequential quadratic programming methods [15, 31] whose implementation details are not consolidated, as well as inexact restoration methods and methods based on filters.

Moreover, in the last 15 years many algorithms appeared aiming to solve new optimization-like problems (equilibrium, multiobjective, bilevel, order-value and many others) [3, 4, 9, 10, 14, 18, 20]. Punctual necessary optimality conditions have been encountered for many of these problems but, frequently, their algorithmic consequences are not clear. We believe that the sequential optimality analysis may be useful in these cases both from the theoretical and the practical point of view.

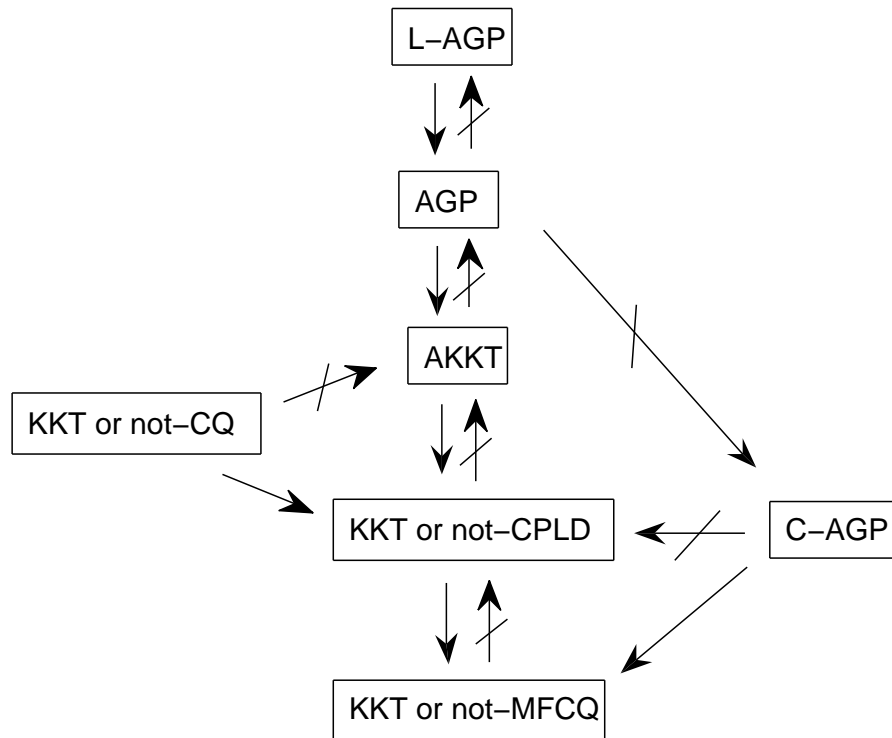


Figure 1: Punctual and Sequential Optimality Conditions

References

- [1] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian Methods with general lower-level constraints, *SIAM Journal on Optimization* 18, pp. 1286–1309, 2007.
- [2] R. Andreani and J. M. Martínez, On the solution of mathematical programming problems with equilibrium constraints, *Mathematical Methods of Operations Research* 54, pp. 345–358, 2001.
- [3] R. Andreani, J. M. Martínez, L. Martínez and F. Yano, Continuous Optimization Methods for Structure Alignments, *Mathematical Programming* 112, pp. 93–124, 2008.
- [4] R. Andreani, J. M. Martínez, L. Martínez and F. S. Yano, Low Order-Value Optimization and applications, *Journal of Global Optimization* 43, pp. 1–10, 2009.
- [5] R. Andreani, J. M. Martínez and M. L. Schuverdt, On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification, *Journal of Optimization Theory and Applications* 125, pp. 473–485, 2005.
- [6] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Mass., 1999.

- [7] E. G. Birgin and J. M. Martínez, Local convergence of an Inexact-Restoration method and numerical experiments, *Journal of Optimization Theory and Applications* 127, pp. 229–247, 2005.
- [8] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.
- [9] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.
- [10] F. Facchinei and J-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* volumes I and II, Springer, New York, 2003.
- [11] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.
- [12] A. Fischer and A. Friedlander, A new line search inexact restoration approach for nonlinear programming, to appear in *Computational Optimization and Applications*.
- [13] R. Fletcher, *Practical Methods of Optimization*, Academic Press, London, 1987.
- [14] R. Gárciga Otero and B. F. Svaiter, A new condition characterizing solutions of variational inequality problems. *Journal of Optimization Theory and Applications* 137, pp. 89-98, 2008.
- [15] F. A. M. Gomes, A sequential quadratic programming algorithm that combines merit function and filter ideas, *Computational and Applied Mathematics* 26, pp. 337–379, 2007.
- [16] M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos, Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints, *SIAM Journal on Scientific Computing* 31, pp. 1628-1652, 2009.
- [17] C. C. Gonzaga, E. W. Karas and M. Vanti, A globally convergent filter method for Nonlinear Programming, *SIAM Journal on Optimization* 14, pp. 646–669, 2003.
- [18] L. M. Graña Drummond and B. F. Svaiter, A steepest descent method for vector optimization, *Journal of Computational and Applied Mathematics* 175, pp. 395-414, 2005.
- [19] G. Haeser, *Condições sequenciais de otimalidade*, Tese de Doutorado, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, Brazil, 2009.
- [20] A. N. Iusem and M. Nasri, Inexact proximal point methods for equilibrium problems in Banach spaces, *Numerical Functional Analysis and Optimization* pp. 1279-1308, 2007.
- [21] E. W. Karas, E. A. Pilotta and A. A. Ribeiro, Numerical comparison of merit function with filter criterion in inexact restoration algorithms using Hard-Spheres Problems, to appear in *Computational Optimization and Applications*.
- [22] C. Y. Kaya and J. M. Martínez, Euler discretization and Inexact Restoration for Optimal Control. *Journal of Optimization Theory and Applications* 134, pp. 191-206, 2007.
- [23] 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.

- [24] J. M. Martínez, Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications* 111, pp. 39-58 ,2001.
- [25] J. M. Martínez and E. A. Pilotta. Inexact restoration algorithms for constrained optimization, *Journal of Optimization Theory and Applications* 104, pp. 135-163, 2000.
- [26] J. M. Martínez and B. F. Svaiter, A practical optimality condition without constraint qualifications for nonlinear programming, *Journal of Optimization Theory and Applications* 118, pp. 117–133, 2003.
- [27] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
- [28] L. Qi and Z. Wei, On the constant positive linear dependence condition and its application to SQP methods, *SIAM Journal on Optimization* 10, pp. 963–981, 2000.
- [29] R. T. Rockafellar, Lagrange multipliers and optimality, *SIAM Review* 35, pp. 183-238, 1993.
- [30] M. L. Schuverdt, *Métodos de Lagrangiano Aumentado com convergência usando a condição de dependência linear positiva constante*, Tese de Doutorado, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, 2006.
- [31] C. Shen, W. Xue and D. Pu, A filter SQP algorithm without a feasibility restoration phase, to appear in *Computational and Applied Mathematics*, 2009.