ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Some Notes on Approximate Optimality Conditions in Scalar and Vector Optimization Problems

Giorgio Giorgi
(Università di Pavia)

Bienvenido Jiménez
(Universidad Nacionale de Educación a Distancia, Madrid)

Vicente Novo
(Universidad Nacionale de Educación a Distancia, Madrid)

 # 95 (10-14)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

October 2014

Some Notes on Approximate Optimality Conditions in Scalar and Vector Optimization Problems

G. Giorgi (*) B. Jiménez (**) V. Novo (***)

Abstract

We give an overview and make some remarks on the approximate optimality conditions, for a nonlinear programming problem, given by Haeser and Schuverdt (2011) and by Fiacco and McCormick (1968a). Other first-order optimality conditions in absence of constraint qualifications are examined. Finally, we extend to a Pareto problem the approximate optimality conditions of Haeser and Schuverdt.

Key words

Approximate optimality conditions. Asymptotic optimality conditions. Sequential optimality conditions. Nonlinear programming. Vector optimization problems.

Mathematics Subject Classification (2000): 90C30, 49K99.

1. Introduction

The characterization of a local solution of a constrained minimization problem has traditionally been given in terms of the functions involved in the problem, put together to form an associated Lagrangian function, whose gradient is evaluated at the solution point for a corresponding set of finite multipliers (the Lagrange multipliers, the Fritz John multipliers, the Karush-Kuhn-Tucker multipliers). Besides this classical approach, other treatments of constrained optimality conditions give a characterization of optimality in terms of appropriate sequences of points and multipliers. In this case we can speak of "approximate optimality conditions" or also "asymptotic optimality conditions" or "sequential optimality conditions". A recent paper on this second approach is due to Haeser and Schuverdt (2011). Other recent papers treating similar questions are due to Andreani, Haeser and Martinez (2011) and to Andreani, Martinez and Svaiter (2010).

* Department of Economics and Management, Via S. Felice, 5 - 27100 Pavia, (Italy). E-mail: ggiorgi@eco.unipv.it.

** Departamento de Matematica Aplicada, E.T.S.I.I. Universidad Nacionale de Educación a Distancia, Madrid, Spain. E-mail: bjimenez@ind.uned.es,

*** Departamento de Matematica Aplicada, E.T.S.I.I. Universidad Nacionale de Educación a Distancia, Madrid, Spain. E-mail: vnovo@ind.uned.es.

However, prior to these contributions, we can quote the works of Kortanek and Evans (1968), Evand and Kortanek (1970) for pseudoconcave programming, the paper of Fiacco and McCormick (1968a) and the papers of Zlobec (1970, 1971a, 1971b, 1972), where this last author generalizes in an asymptotic way, the optimality conditions given by Guignard (1969). Asymptotic versions of the Karush-Kuhn-Tucker conditions are considered also by Craven (1984) and by Trudzik (1981/82).

The present paper is organized as follows. In Section 2 we give an overview of the approach of Haeser and Schuverdt (2011); in Section 3 we follow mainly the approach of Fiacco and McCormick (1968a), by extending the first-order approximate optimality conditions also to second-order results. In Section 4 we are concerned with other optimality conditions in absence of constraint qualifications. In Section 5 we extend the results of Haeser and Schuverdt (2011) to a vector (Pareto) optimization problem.

The scalar mathematical programming problem we consider in Section 2 is:

(P) $\qquad\qquad$ *Minimize* $f(x)$, subject to $x \in S$

where

$$S = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\},$$

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $h : \mathbb{R}^n \longrightarrow \mathbb{R}^r$ are continuously differentiable functions. The associated Lagrangian function is

$$\mathcal{L}(x, u, w) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{r} w_j h_j(x).$$

We define the set of *active* (*inequality*) *constraints* at the feasible point x^0 as

$$I(x^0) = \{i : g_i(x^0) = 0\}$$

and we define the cone

$$Z(x^0) = \{z \in \mathbb{R}^n : z\nabla g_i(x^0) \leq 0,\ i \in I(x^0), z\nabla h_j(x^0) = 0,\ j = 1, ..., r,\ \text{and } z\nabla f(x^0) < 0\}.$$

The following result is well known and is a direct consequence of the Farkas theorem of the alternative.

Theorem 1. If $x^0 \in S$ and if $Z(x^0) = \varnothing$, then there exist vectors $u^0 \in \mathbb{R}^m$ and $w^0 \in \mathbb{R}^r$ such that

$$\nabla_x \mathcal{L}(x^0, u^0, w^0) = 0, \tag{1}$$

$$u_i^0 g_i(x^0) = 0,\ i = 1, ..., m, \tag{2}$$

$$u_i^0 \geq 0,\ i = 1, ..., m. \tag{3}$$

The following result is perhaps the most widely invoked theorem on nonlinear programming.

Theorem 2. (Karush-Kuhn-Tucker) If x^0 is a local solution of (P) and if some constraint qualification holds at x^0, then the hypotheses of Theorem 1 are satisfied and (1)-(3) follow.

One of the constraint qualifications more used for (P) is the Mangasarian-Fromovitz constraint qualification:

- At $x^0 \in S$ the gradients $\nabla h_1(x^0), ..., \nabla h_r(x^0)$ are linearly independent and there exists a vector $s \in \mathbb{R}^n$ such that

$$\begin{cases} s\nabla g_i(x^0) < 0, \ \forall i \in I(x^0) \\ s\nabla h_j(x^0) = 0, \ j = 1, ..., r. \end{cases}$$

This constraint qualification is necessary and sufficient for the set of the Karush-Kuhn-Tucker multipliers (u, w) satisying (1)-(3) to form a bounded set (see Gauvin (1977)). The above quoted conditions of optimality may fail for a large class of problems. This justifies the search for optimality conditions for (P), even when the Karush-Kuhn-Tucker multipliesrs do not exist.

2. The Approach of Haeser and Schuverdt

We give a short account of the sequential optimality conditions fo (P) of Haeser and Schuverdt (2011), as these conditions will be generalized to a Pareto multiobjective problem in Section 5. See also the paper of Andreani, Martinez and Svaiter (2010) and of Andreani, Haeser and Martinez (2011).

Definition 1. Let us consider (P); we say that the Approximate Karush-Kuhn-Tucker Condition (AKKT) is satisfied at a feasible point $x^0 \in S$ if, and only if, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{u^k\} \subset \mathbb{R}^m_+$, $\{w^k\} \subset \mathbb{R}^r$, such that $x^k \longrightarrow x^0$,

$$\nabla f(x^k) + \sum_{i=1}^{m} u_i^k \nabla g_i(x^k) + \sum_{j=1}^{r} w_j^k \nabla h_j(x^k) \longrightarrow 0 \tag{4}$$

and

$$g_i(x^0) < 0 \Longrightarrow u_i^k = 0 \text{ for sufficiently large } k. \tag{5}$$

This AKKT condition corresponds to the AKKT($\varnothing$) condition of Andreani, Haeser and Martinez (2011). It must be noted that AKKT implies the Karush-Kuhn-Tucker optimality conditions (1)-(3) under the *constant positive linear dependence condition* (CPLD):

- CPLD holds at $x^0 \in S$ if there exists a neighborhood $B(x^0)$ of x^0 such that for every $I \subset I(x^0)$ and every $J \subset \{1, ..., r\}$, whenever $(\{\nabla g_i(x^0)\}_{i \in I}, \{\nabla h_j(x^0)\}_{j \in J})$ is positive-linearly dependent, then $\{\nabla g_i(y)\}_{i \in I} \cup \{\nabla h_j(y)\}$ is linearly dependent for every $y \in B(x^0)$.

CPLD is a constraint qualification weaker than the Mangasarian-Fromovitz c. q.: see Andreani, Haeser, Schuverdt and Silva (2012). Moreover, CPLD is also implied by the *constant rank constraint qualification* (CRCQ) at $x^0 \in S$, which in turn implies the *relaxed constant rank* *condition* (RCR) at $x^0 \in S$. Andreani, Haeser, Schuverdt and Silva (2012) have introduced a relaxed version of CPLD, which is implied by all conditions mentioned above.

Note, moreover, that if $x^0 \in S$ is a local minimum point for (P) and any constraint qualification holds at x^0, then AKKT holds at x^0 for constant sequences $x^k = x^0$, $u^k = u^0$, $w^k = w^0$, being $u^0 \in \mathbb{R}^m_+$ and $w^0 \in \mathbb{R}^r$.

Haeser and Schuverdt (2011) prove the following result, which is a special case of a more general result of Andreani, Haeser and Martinez (2011).

Theorem 3. If $x^0 \in S$ is a local minimum for (P), then x^0 satisfies the AKKT condition (4)-(5).

The same authors then prove that a stronger version of the AKKT condition is suffient for optimality in convex programming.

Definition 2. A point $x^0 \in S$ satisfies the *strong* AKKT *condition* (SAKKT) if there exist sequences $\left\{x^k\right\} \subset \mathbb{R}^n$, $\left\{u^k\right\} \subset \mathbb{R}^m_+$, $\left\{w^k\right\} \subset \mathbb{R}^r$ such that (4) holds and

$$g_i(x^k) < 0 \Longrightarrow u_i^k = 0.$$

We note that every local minimizer for (P) satisfies also SAKKT.

Theorem 4. Let in (P) be f and g convex functions and let h be an affine function. If $x^0 \in S$ satisfies SAKKT and if the sequences $\left\{x^k\right\}, \left\{w^k\right\}$ are such that $w_j^k h_j(x^k) \geq 0$ for every $j = 1, ..., r$ and for every $k \in \mathbb{N}$, then x^0 is a solution for (P).

From the proof of Haeser and Schuverdt we can deduce that Theorem 4 holds also uinder the more general assumption of f and g pseudoconvex functions. Another approach to optimality conditions for (P), which has some aspects of similarity with the present approach, is due to Martinez and Svaiter (2003).

3. The Approach of Fiacco and McCormick

The pioneering contributions of Fiacco and McCormick (1968a) to asymptotic optimality conditions benefit from some results of their basic book on sequential unconstrained minimization techniques for mathematical programming problems (Fiacco and McCormick (1968b)). Their main results are summarized below.

Definition 3. A nonempty set $M^* \subset M \subset \mathbb{R}^n$ is called an *isolated set* of M if there exists a closed set E such that $int(E) \supset M^*$ and such that if $x \in E \diagdown M^*$, then $x \notin M$.

Fiacco and McCormick consider the following slight variant of (P):

(P_1) $\quad$ *Minimize* $f(x)$, subject to $x \in S_1$

where

$$S_1 = \{x \in \mathbb{R}^n : g(x) \geq 0, h(x) = 0\}.$$

The assumptions on (P_1) are the same made on (P); the associated Lagrangian function is

$$\mathcal{L}_1(x,u,w) = f(x) - \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{r} w_j h_j(x).$$

The Karush-Kuhn-Tucker conditions for (P_1) are obviously the same for (P), i. e. conditions (1)-(3), with $\mathcal{L}_1(x,u,w)$ instead of $\mathcal{L}(x,u,w)$.

Definition 4. Let $O(x)$ be a continuous function such that $O(x) = 0$ if $x \in S_1$ (i. e. if x satisfies the constraints of problem (P_1)) and $O(x) > 0$ otherwise.

Fiacco and McCormick (1968a) remark that the function $T(x,t) = f(x) + tO(x)$ is a penalty function for (P_1).

Definition 5. Let M be the set of local minimum points for (P_1), with local minimum value v^*.

Lemma 1. (Fiacco and McCormick (1968a, 1968b)) Let the functions involved in (P_1) be continuous; if a compact set M^* is an isolated set of M, if $T(x,t)$ is a penalty function, and if $\{t_k\}$ is an increasing unbounded positive sequence, then there exists a compact set K containing M^* in its interior and such that the unconstrained minima of $T(x,t_k)$ in $int(K)$ exist for k large enough, and every limit point of any subsequences $\{x^k\}$ of the minimizing points is in M^*. Furthermore, it follows that $T(x^k,t_k)$ and $f(x^k)$ monotonically increase to v^*, $O(x^k)$ monotonically decreases to 0, and $\lim_{k \longrightarrow \infty} t_k O(x^k) = 0$.

Several examples of penalty functions had been considered in the literature. Fiacco and McCormick (1968a) choose the following one:

$$T(x,t) = f(x) + t\sum_{i=1}^{m} g_i^2(x) H(g_i) + t\sum_{j=1}^{r} h_j^2(x) \tag{6}$$

where $H(g_i) = 0$ if $g_i(x) \geq 0$ and $H(g_i) = 1$ if $g_i(x) < 0$.

This penalty function has been rather extensively studied by various authors: see, e. g., Butler and Martin (1962), Fiacco and McCormick (1967, 1968b), Pietrzykowski (1962), Zangwill (1967).

It may be easily verified that (6) satisfies the definition of a penalty function as given above. Based on the above concepts and results, Fiacco and McCormick (1968a) present the following optimality conditions for problem (P_1).

Theorem 5. Let us suppose that there exists a compact set M^* which is an isolated set of M (set of local minima of (P_1) associated with local minimum value v^*). Then there exists a triplet (x^k,u^k,w^k) such that $x^k \longrightarrow x^* \in M^*$, $f(x^k) \longrightarrow v^*$, $u_i^k \geq 0$ and $u_i^k g_i(x^k) \longrightarrow 0$ for $i = 1,...,m$, $w_j^k h_j(x^k) \longrightarrow 0$ for $j = 1,...,r$, and $\nabla\mathcal{L}_1(x^k,u^k,w^k) \equiv 0$.

Proof. From Lemma 1, it follows that an unconstrained minimizing point x^k for any penalty function $T(x,t_k)$ exists with the property that $x^k \longrightarrow x^* \in M^*$ and $f(x^k) \longrightarrow v^*$.

Numerous realizations of $T(x,t)$ can be used to obtain the desired results. Let us consider the example of the penalty function given in (6). A necessary condition for unconstrained minimization implies that

$$\nabla_x T(x^k,t_k) \equiv \nabla f(x^k) + 2t_k \sum_{i=1}^{m} g_i(x^k).H\left[g_i(x^k)\right]\nabla g_i(x^k) + 2t_k \sum_{j=1}^{r} h_j(x^k)\nabla h_j(x^k) = 0. \quad (7)$$

If we define

$$u_i^k \equiv -2t_k g_i(x^k)H\left[g_i(x^k)\right],\ i = 1,...,m \quad (8)$$

and

$$w_j^k \equiv 2t_k h_j(x^k),\ j = 1,...,r, \quad (9)$$

then it follows from (7) that

$$\nabla_x T(x^k,t_k) = \nabla_x \mathcal{L}_1(x^k,u^k,w^k) \equiv 0 \quad (10)$$

where $\mathcal{L}_1(x,u,w)$ is the Lagrangian function associated to (P_1) and previously defined. Note that $u_i^k \geq 0$ for all i. Further, invoking the last conclusion of Lemma 1, i. e. $\lim_{k\longrightarrow\infty} t_k O(x^k) = 0$ and interpreting this in terms of our present penalty function (6) leads immediately to the fact that $u_i^k g_i(x^k) \longrightarrow 0$ and $w_j^k h_j(x^k) \longrightarrow 0$ for all i,j. □

The next results are essentially the approximate necessary optimality conditions for (P_1) given by Haeser and Schuverdt (2011).

Corollary 1. If $x^* \in S_1$ is a local minimum for (P_1), then the conclusions of Theorem 5 hold, with the exception that the identical vanishing of the gradient of Lagrangian function along the sequence must be substituted with the condition

$$\nabla_x \mathcal{L}_1(x^k,u^k,w^k) \longrightarrow 0.$$

Proof. Consider problem (P_1) with the objective function replaced by $f(x) + \frac{1}{2} \parallel x - x^* \parallel^2$, and denote this perturbed problem by (P_1^*). Since x^* is a local minimum of problem (P_1), it follows that x^* is an isolated (i. e. unique in a neighborhoood) local minimum of problem (P_1^*). The conclusions of the corollary now follow immediatedly from Theorem 5 applied to problem (P_1^*), noting that

$$\nabla_x \mathcal{L}_1(x^k,u^k,w^k) + (x^k - x^*) \equiv 0,$$

so that $\nabla_x \mathcal{L}_1(x^k,u^k,w^k) \longrightarrow 0$. □

It is worthwhile to remark that if a solution $\bar{x}$ of problem (P_1) is isolated, i. e. locally unique, then the final conclusions of Corollary 1 can be strengthened to $\nabla_x \mathcal{L}_1(x^k,u^k,w^k) \equiv 0$. This follows from the fact that the set of local minima M^*, as in Theorem 5, can be selected such that $M^* = \{x^*\}$.

Fiacco and McCormick remark also that Corollary 1 leads easily to the Fritz John optimality conditions for (P_1). Define the generalized Lagrangian function

$$\tilde{\mathcal{L}}_1(x,u,w) = \mu_0 f(x) - \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{r} \omega_j h_j(x).$$

Corollary 2. If $x^* \in S_1$ is a local minimum point for (P_1), then there exists a pair $(\mu^*, \omega^*) \neq 0$ such that $\mu_i^* \geq 0$, $i = 0, 1, ..., m$, $\mu_i^* g_i(x^*) = 0$, $i = 1, ..., m$, and $\nabla \tilde{\mathcal{L}}_1(x^*, u^*, \omega^*) = 0$.

Proof. If $\{\mu^k, \omega^k\}$ has a finite limit point μ^*, ω^*, the conclusion follows immediately from the conclusions of Corollary 1, with $\mu_0^* = 1$, $\mu_i^* = u_i^*$, $i = 1, ..., m$, and $\omega_j^* = w_j^*$, $j = 1, ..., r$. Otherwise, define $v^k = \sum_i \mu_i^k + \sum_j \mid \omega_j^k \mid$. We can assume $v^k > 0$ for every k and $v^k \longrightarrow +\infty$. Let $\mu_0^k = 1/v^k$, $\mu_i^k = u_i^k/v^k$, $i = 1, ..., m$, and $\omega_j^k = w_j^k/v^k$, $j = 1, ..., r$. There must exist a subsequence which we still denote by $\{\mu^k, \omega^k\}$, and a pair (μ^*, ω^*) such that $\mu_0^k \longrightarrow \mu_0^* = 0$, $\mu_i^k \longrightarrow \mu_i^*$, $i = 1, ..., m$, and $\omega_j^k \longrightarrow \omega_j^*$, $j = 1, ..., r$. The conclusions now follow by dividing the necessary limiting relations of Corollary 1 by v^k and passing to the limit along the indicated convergent subsequence. □

Corollary 2 shows that, all things considered, the main optimality result os Haeser and Schuverdt is not, beyond its algorithmic relevance, more general than the classical optimality theorem of Fritz John. We will make other similar considerations in the next Section.

If (P_1) is a convex programming problem, i. e. if f is convex, g is concave and h is linear affine, by exploiting some results in Fiacco and McCormick (1968b, Theorem 28), it is possible to state also necessary and sufficient asymptotic optimality conditions for (P_1), under the above assumptions. In particular, the sufficient conditions are quite similar to the sufficient conditions obtained by Haeser and Schuverdt.

Theorem 6. If (P_1) is a convex programming problem and the set M^* of its solutions is nonempty and bounded, then there exist vectors (x^k, u^k, w^k) such that $x^k \longrightarrow x^* \in M^*$, $f(x^k) \longrightarrow f(x^*)$, $u_i^k \geq 0$ and $u_i^k g_i(x^k) \longrightarrow 0$ for $i = 1, ..., m$, $w_j^k h_j(x^k) \longrightarrow 0$ for $j = 1, ..., r$, and $\mathcal{L}_1(x^k, u^k, w^k) \leq \mathcal{L}_1(x, u^k, w^k)$ for all k.

Theorem 7. Necessary and sufficient conditions that x^* be a solution of the convex programming (P_1) are that there exists (x^k, u^k, w^k) such that $x^k \longrightarrow x^* \in S_1$, $u_i^k \geq 0$ and $u_i^k g_i(x^k) \longrightarrow 0$ for $i = 1, ..., m$, $w_j^k h_j(x^k) \longrightarrow 0$ for $j = 1, ..., r$, and $\liminf_k \mathcal{L}_1(x^k, u^k, w^k) \leq \liminf_k \mathcal{L}_1(x, u^k, w^k)$.

Theorem 8. Necessary and sufficient conditions that x^* be a solution of the convex programming problem (P_1), are that there exists (x^k, u^k, w^k) such that $x^k \longrightarrow x^* \in S_1$, $u_i^k \geq 0$ and $u_i^k g_i(x^k) \longrightarrow 0$ for $i = 1, ..., m$, $w_j^k h_j(x^k) \longrightarrow 0$ for $j = 1, ...r$, and $\nabla_x \mathcal{L}_1(x^k, u^k, w^k) \longrightarrow 0$.

Next, Fiacco and McCormick (1968a) give also asymptotic second order conditions for (P_1), under the assumptions that the functions involved are twice continuously differentiable. For the reader's convenience we report their main results. To obtain the second order necessary conditions these authors utilize the following exterior penalty function, which is everywhere twice continuously differentiable:

$$T(x,t) = f(x) + t\sum_{i=1}^{m}[\min(0, g_i(x)]^4 + t\sum_{j=1}^{r} h_j^4(x). \tag{11}$$

The first order necessary condition that this function has an unconstrained minimum is that

$$\nabla T(x,t) = 0. \tag{12}$$

The second order condition is that the Hessian matrix $\nabla^2 T(x,t)$ is positive semidefinite, i.e.

$$z^\top \nabla^2 T(x,t) z \geq 0, \text{ for all } z. \tag{13}$$

Suppose x^* is the limit of x^k, a sequence of unconstrained minima of (11) corresponding to $\{t_k\}$, where $0 < t_k \longrightarrow +\infty$. Expanding $\nabla^2 T(x^k, t_k)$, with $f^k \equiv f(x^k)$, $g_i^k \equiv g_i(x^k)$, etc., yields

$$\begin{aligned}&\nabla^2 f^k - \textstyle\sum_{i\in F^k} u_i^k \nabla^2 g_i^k + \sum_{j=1}^{r} w_j^k \nabla^2 h_j^k + \sum_{i\in F^k} \nabla g_i^k u_i^k \left(\frac{3}{g_i^k}\right)\left(\nabla g_i^k\right)^\top + \\ &+ \textstyle\sum_{j=1}^{r} \nabla h_j^k w_j^k \left(\frac{3}{h_j^k}\right)\left(\nabla h_j^k\right)^\top\end{aligned} \tag{14}$$

where $u_i^k = -4t_k g_i^3(x^k)$, $i \in F^k \equiv \{i \mid g_i(x^k) < 0\}$ and $w_j^k = 4t_k h_j(x^k)$, $j = 1, ..., r$.

When t_k is large enough, all constraints where $g_i(x^*) > 0$ will be strictly satisfied and hence do not enter into the penalty function. From the fact that (14) must be a positive semidefinite matrix, it follows that

$$\left(z^k\right)^\top \nabla^2 \mathcal{L}_1(x^k, u^k, w^k) z^k \geq 0 \tag{15}$$

for all $z^k \in Z^k \equiv \{z \mid z^\top \nabla h_j^k = 0, \text{ all } j, \text{ and } z^\top \nabla g_i^k = 0, \text{ all } i \in F^k\}$.

On the grounds of the above considerations, Fiacco and McCormick (1968a) obtain the following asymptotic second-order necessary conditions for an isolated compact set of local minima for (P_1).

Theorem 9. If in (P_1) the functions are twice continuously differentiable, and if a compact set M^* is an isolated set of M, then there exists (x^k, u^k, w^k) such that $x^k \longrightarrow x^* \in M^*$, $f(x) \longrightarrow v^*$, $u_i^k \geq 0$ and $u_i^k g_i(x^k) \longrightarrow 0$ for $i = 1, ..., m$, $w_j^k h_j(x^k) \longrightarrow 0$ for $j = 1, ...r$, $\nabla_x \mathcal{L}_1(x^k, u^k, w^k) = 0$ and

$$\left(z^k\right)^\top \nabla^2 \mathcal{L}_1(x^k, u^k, w^k) z^k \geq 0$$

for all z^k where $(z^k)^\top \nabla h_j^k = 0$, $j = 1, ..., r$ and $(z^k)^\top \nabla g_i^k = 0$ for all $i \in I(x^*) = \{i \mid g_i(x^*) = 0\}$.

Proof. The first part of this theorem duplicated Theorem 5. The final part follows from (15) and the fact that for k large enough, $F^k \subset I(x^*)$. □

Corollary 3. If $x^* \in S_1$ is a local (not necessarily isolated) minimum of (P_1), and if the functions involved in (P_1) are twice continuously differentiable, then the conclusions of Theorem 9 follow, except that the vanishing of the gradient of $\mathcal{L}_1$ is replaced by the condition

$$\nabla_x \mathcal{L}_1(x^k, u^k, w^k) \longrightarrow 0,$$

and the second-order conditions are replaced by

$$\liminf_{k\longrightarrow\infty} \left(z^k\right)^\top \nabla^2\mathcal{L}_1(x^k,u^k,w^k)z^k \geq 0$$

for all z^k where $(z^k)^\top\nabla h_j^k = 0$, $j = 1,...,r$, and $(z^k)^\top\nabla g_i^k = 0$ for all $i \in I(x^*)$.

Proof. The proof follows by applying Theorem 9 to problem (P_1), modified by adding the term $\sum_{j=1}^n (x_j - x_j^*)^2$ to the objective function. This makes x^* an isolated local minimum of the modified problem, hence making Theorem 9 directly applicable. Since the gradient and second order partial derivative matrix of the perturbation vanish as $x^k \longrightarrow x^*$, the appropriate conclusions can be replaced by limiting statements. □

Finally, we note that Fiacco and McCormick (1968a) obtain also asymtotic second-order sufficient conditions for a strict local minimum of problem (P_1).

4. Other Necessary First-Order Optimality Conditions in Absence of Constraint Qualifications

Perhaps the most known necessary first-order optimality conditions for (P), in absence of constraint qualifications, are the Fritz John conditions. With reference to (P) these conditions have been obtained by Mangasarian and Fromovitz (1967); see also Birbil, Frenk and Still (2007) and Giorgi (2011) for other more elementary and short proofs of the Fritz John conditions for (P).

With reference to a mathematical programming problem with inequality only, i. e. to problem

(P_0) $$\underset{x\in S_0}{Minimize} f(x),$$

where $S_0 = \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i = 1,...,m\}$, Elster and Götz (1972) have given a generalization of the Kuhn-Tucker conditions, with disregards of constraint qualifications. We recall the classical definition of *contingent cone* or *Bouligand tangent cone* at a point $x^0 \in cl(M)$, $M \subset \mathbb{R}^n$:

$$T(M,x^0) = \left\{x \in \mathbb{R}^n : \exists \{x^n\} \subset M,\ x^n \longrightarrow x^0, \exists \{\lambda_n\} \subset \mathbb{R}_+ \text{ such that } \lim_{n\longrightarrow+\infty} \lambda_n(x^n - x^0) = x\right\}.$$

We recall that the *polar cone* of a cone $C \subset \mathbb{R}^n$ is the convex cone

$$C^* = \{x \in \mathbb{R}^n : xy \leq 0,\ \forall y \in C\}.$$

Theorem 10 (Elster and Götz). Let x^0 be a local solution of (P_0), where f and every g_i, $i = 1,...,m$, are differentiable at x^0. Then, for each number $\delta > 0$ there exist vectors $\varepsilon^i \in T^*(S_0,x^0)$, with $\| \varepsilon^i \| < \delta$, $i = 1,...,m$, and $u = u(\varepsilon^1,...,\varepsilon^m) \in \mathbb{R}^m$, such that

$$\nabla f(x^0) + \sum_{i=1}^m u_i \left[\nabla g_i(x^0) + \varepsilon^i\right] = 0$$

$$u_i \geq 0,\ u_i g_i(x^0) = 0,\ i = 1,...,m.$$

Furthermore, if a constraint qualification holds at x^0, then the above conditions collapse to the classical Karush-Kuhn-Tucker conditions.

On the above result we make the following three remarks.

Remark 1. The conditions of Theorem 10 seem more precise than the Fritz John conditions, but from a strictly mathematical point of view they are in fact equivalent to the Fritz John assertion. Indeed, since the vectors ε^i are essentially any multiple of the vector $\nabla f(x^0)$, the condition can be rewritten as

$$\nabla f(x^0) + \sum u_i(\nabla g_i(x^0) + (...)\nabla f(x^0)) = (...)\nabla f(x^0) + \sum u_i \nabla g_i(x^0) = 0.$$

Conversely, from the Fritz John condition

$$u_0 \nabla f(x^0) + \sum u_i \nabla g_i(x^0) = 0$$

e. g. with $u_0 + \sum u_i = 1$, we get

$$\nabla f(x^0) + \sum u_i(\nabla g_i(x^0) + \nabla f(x^0)) = 0$$

so that the conditions of the theorem are fulfilled with $\varepsilon^i = \nabla f(x^0)$.

Remark 2. The proof of Elster and Götz (1972) can be simplified. The crucial point is the fact that for an optimal solution x^0 the inequality system

$$y^\top \nabla f(x^0) < 0,\ y^\top \nabla g_i(x^0) < 0,\ i \in I(x^0),$$

has no solution (this is the "Abadie linearization lemma"; by using the Gordan theorem of the alternative we obtain directly the Fritz John conditions). That means that for each y with $y^\top \nabla f(x^0) < 0$ there exists an index $i \in I(x^0)$ such that $y^\top \nabla g_i(x^0) \geq 0$ and especially

$$y^\top(\nabla g_i(x^0) + \frac{1}{M}\nabla f(x^0)) > 0,$$

with $M > 0$ (this constant is not essential). Hence, by contraposition $y^\top(\nabla g_i(x^0) + \frac{1}{M}\nabla f(x^0)) \leq 0, \forall i \in I(x^0)$, implies $y^\top \nabla f(x^0) \geq 0$ and the Farkas theorem provides the conditions of Theorem 10 with $\varepsilon^i = \frac{1}{M}\nabla f(x^0)$.

Remark 3. The result of Elster and Götz can be extended to problem (P) using, as for the Fritz John conditions, the implicit function theorem and assuming thet the functions h_j, defining the equality constraints $h_j(x) = 0$, $j = 1, ..., r$, are continuously differentiable around $x^0 \in S$.

Yet another approach to optimality conditions for (P_0), in absence of constraint qualifications, has been presented by Gould and Tolle (1972). Let us denote by C_0 the *linearizing cone* at $x^0 \in S_0$, i. e.

$$C_0 = \left\{x \in \mathbb{R}^n : x\nabla g_i(x^0) \leq 0,\ \forall i \in I(x^0)\right\}.$$

It is well known (see, e. g., Abadie (1967), Bazaraa and Shetty (1976)) that it holds $T(S_0, x^0) \subset C_0$, i. e. $C_0^* \subset T^*(S_0, x^0)$. As $T^*(S_0, x^0)$ is a convex cone, from the relation

$$T^*(S_0, x^0) = C_0^* \cup (T^*(S_0, x^0) \diagdown C_0^* \cup \{0\})$$

we obtain

$$T^*(S_0, x^0) = C_0^* + (T^*(S_0, x^0) \diagdown C_0^* \cup \{0\})$$

i. e.

$$T^*(S_0, x^0) = B_0^* + (T^*(S_0, x^0) \diagdown C_0^* \cup \{0\}),$$

where $B_0^* = \left\{ x \in \mathbb{R}^n : x = \sum_{i \in I(x^0)} \lambda_i \nabla g_i(x^0), \ \lambda_i \geq 0 \right\}$ is the *cone of gradients.*

Taking into account the well known necessary optimality condition of Gould and Tolle (1971), Guignard (1969), Varaiya (1967), that is, if $x^0 \in X$ is a local solution of the problem $\underset{x \in X}{Min} f(x)$, with $X \subset \mathbb{R}^n$ and f differentiable at x^0, then $-\nabla f(x^0) \in T^*(X, x^0)$, we have therefore the following result, due to Gould and Tolle (1972).

Theorem 11. If x^0 is a local solution of (P_0), where the functions are differentiable at x^0, then there exist scalars $u_i \geq 0$, $i \in I(x^0)$, such that

$$-\left[\nabla f(x^0) + \sum_{i \in I(x^0)} u_i \nabla g_i(x^0) \right] \in T^*(S_0, x^0) \diagdown C_0^* \cup \{0\}\,.$$

Remark 4.

i) If it holds

$$C_0^* = T^*(S_0, x^0),$$

i. e. if the Guignard-Gould-Tolle constraint qualification holds at x^0, the the previous optimality conditions collapse to the usual Karush-Kihn-Tucker conditions.

ii) Theorem 11 can immediately be fitted to (P), by defining the cone

$$D_0 = \left\{ x \in \mathbb{R}^n : x \nabla h_j(x^0) = 0, \ \forall j = 1, ..., r \right\}$$

and the linearizing cone $E_0 = C_0 \cap D_0$. We have then the necessary optimality condition:

$$-\left[\nabla f(x^0) + \sum_{i \in I(x^0)} u_i \nabla g_i(x^0) + \sum_{j=1}^{r} w_j \nabla h_j(x^0) \right] \in T^*(S, x^0) \diagdown E_0^* \cup \{0\}\,,$$
$$u_i \geq 0, i \in I(x^0), w_j \in \mathbb{R}, j = 1, ..., r.$$

5. Approximate Karush-Kuhn-Tucker Conditions in Pareto Multiobjective Problems

In the present Section we extend the main result of Haeser and Schuverdt (2011), i. e. Theorem 3, to a Pareto optimization problem. We consider the following vector optimization problem

(VP) $$\underset{x\in S}{Minimize} f(x)$$

where $S = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$, $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $h : \mathbb{R}^n \longrightarrow \mathbb{R}^r$ are continuously differentiable functions. We recall that, given a problem

$$Minimize \{f(x) : x \in M\},$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ and $M \subset \mathbb{R}^n$, a point $x^0 \in M$ is said to be a *weak Pareto minimum,* denoted $x^0 \in WMin(f, M)$, if there is no $x \in M$ such that $f(x) < f(x^0)$. If the previous conditions are required to hold on a neighborhood of the point x^0, then x^0 is a *local* weak Pareto minimum point. In this Section we follow, in general, the notations of Haeser and Schuverdt (2011).

Definition 6. We say that the Approximate Karush-Kuhn-Tucker conditions are satisfied for (VP) at a feasible point $x^0 \in S$ if and only if there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p_+$, $\{u^k\} \subset \mathbb{R}^m_+$, and $\{w^k\} \subset \mathbb{R}^r$, with $x^k \longrightarrow x^0$ and such that

$$\sum_{\ell=1}^{p} \lambda_\ell \nabla f_\ell(x^k) + \sum_{i=1}^{m} u_i^k \nabla g_i(x^k) + \sum_{j=1}^{r} w_j^k \nabla h_j(x^k) \longrightarrow 0 \tag{16}$$

$$\sum_{\ell=1}^{p} \lambda_\ell = 1 \tag{17}$$

$$g_i(x^0) < 0 \Longrightarrow u_i^k = 0 \text{ for sufficiently large } k. \tag{18}$$

We define $\phi : \mathbb{R}^p \longrightarrow \mathbb{R}$ by

$$\phi(y) = \underset{1\leq i\leq p}{Max} \{y_i\}.$$

This function will be used to scalarize (VP). The following properties of ϕ are clear:

(*i*) $\phi(y) \leq 0 \Longleftrightarrow y \in -\mathbb{R}^p_+$,

(*ii*) $\phi(y) < 0 \Longleftrightarrow y \in -int\mathbb{R}^p_+$.

Also the following result is well known. We prove it for the reader's convenience.

Lemma 2. If $x^0 \in WMin(f, M)$, then $x^0 \in Min(\phi(f(\cdot) - f(x^0)), M)$.

Proof. Suppose that $x^0 \notin Min(\phi(f(\cdot) - f(x^0)), M)$, then there exists $x^1 \in M$ such that $\phi(f(x^1) - f(x^0)) < \phi(f(x^0) - f(x^0)) = 0$. It follows that $f(x^1) - f(x^0) \in -int\mathbb{R}^p_+$, which contradicts the assumption. □

The main result of the present Section is the following

Theorem 12. If $x^0 \in S$ is a local weak solution of problem (*VP*), then x^0 satisfies the AKKT conditions.

Proof. By assumption, there exists $\delta > 0$ such that $x^0 \in WMin(f, S \cap B(x^0, \delta))$, where $B(x^0, \delta)$ is the open ball centered at x^0 and with radius δ. By Lemma 2 we deduce that $x^0 \in Min(\phi(f(\cdot) - f(x^0)), S \cap B(x^0, \delta))$. In consequence, we may suppose that x^0 is the unique solution of the problem

$$Minimize(\phi(f(x) - f(x^0)) + \frac{1}{2} \parallel x - x^0 \parallel^2, \text{ subject to } x \in S, \ \parallel x - x^0 \parallel \leq \delta \tag{19}$$

(choosing a small δ, if necessary).

We define, for each $k \in \mathbb{N}$,

$$\Psi_k(x) = \phi(f(x) - f(x^0)) + \frac{1}{2} \parallel x - x^0 \parallel^2 + \sum_{i=1}^{m} kg_i^2(x)_+ + \sum_{j=1}^{r} kh_j^2(x),$$

where. if $v \in \mathbb{R}^m$, we denote $v_+ = (\max\{v_1, 0\}, ..., \max\{v_m, 0\})^\top$.

Let x^k be a solution of the problem

$$Minimize \Psi_k(x) \text{ subject to } \parallel x - x^0 \parallel \leq \delta. \tag{20}$$

Let us observe that x^k exists because $\Psi_k(x)$ is continuous and $\bar{B}(x^0, \delta)$ (closure of $B(x^0, \delta)$) is compact. Let z be an accumulation point of $\{x^k\}$. We may suppose that $x^k \longrightarrow z$, choosing a subsequence if necessary. On one hand, we have

$$\phi(f(x^k) - f(x^0)) \leq \Psi_k(x^k) \tag{21}$$

because $\Psi_k(x^k) - \phi(f(x^k) - f(x^0)) = \frac{1}{2} \parallel x^k - x^0 \parallel^2 + \sum_{i=1}^m kg_i^2(x^k)_+ + \sum_{j=1}^r kh_j^2(x^k) \geq 0$.

On the other hand, as x^0 is a feasible point of problem (20) and x^k is a solution, one has

$$\Psi_k(x^k) - \phi_k(x^0) = 0, \ \forall k \in \mathbb{N} \tag{22}$$

since $x^0 \in S$.

Let us prove that z is a feasible point for problem (19). Indeed, first, as $\parallel x - x^0 \parallel \leq \delta$ it follows that $\parallel z - x^0 \parallel \leq \delta$. Second, suppose that $\sum_{i=1}^m g_i^2(z)_+ + \sum_{j=1}^r h_j^2(z) > 0$. Then, there exists $c > 0$ such that $\sum_{i=1}^m g_i^2(x^k)_+ + \sum_{j=1}^r h_j^2(x^k) > c$ for all k large enough, by continuity and because $x^k \longrightarrow z$.

Now, as

$$\Psi_k(x^k) = \phi(f(x^k) - f(x^0)) + \tfrac{1}{2} \parallel x^k - x^0 \parallel^2 + k(\textstyle\sum_{i=1}^m g_i^2(x^k)_+ \\ + \textstyle\sum_{j=1}^r h_j^2(x^k)) > \phi(f(x^k) - f(x^0)) + kc,$$

taking the limit we obtain $\Psi_k(x^k) \longrightarrow +\infty$, which contradicts (22). Therefore we have $\sum_{i=1}^m g_i^2(z)_+ + \sum_{j=1}^r h_j^2(z) = 0$, and this implies that $z \in S$.

From (22) one has

$$\Psi_k(x^k) = \phi(f(x^k) - f(x^0)) + \frac{1}{2} \parallel x^k - x^0 \parallel^2 + \sum_{i=1}^m k g_i^2(x^k)_+ + \sum_{j=1}^r k h_j^2(x^k) \leq 0$$

and, as $\sum_{i=1}^m k g_i^2(x^k)_+ + \sum_{j=1}^r k h_j^2(x^k) \geq 0$, we get $\phi(f(x^k) - f(x^0)) + \frac{1}{2} \parallel x^k - x^0 \parallel^2 \leq 0$. Taking the limit we have

$$\phi(f(z) - f(x^0)) + \frac{1}{2} \parallel z - x^0 \parallel^2 \leq 0.$$

As x^0 is the unique solution of problem (19), with value 0, we conclude that $z = x^0$. Therefore, $x^k \longrightarrow x^0$ and $\parallel x^k - x^0 \parallel \leq \delta$ for all k sufficiently large.

Now, as x^k is a solution of the nonsmooth problem (20) and it is an interior point of the feasible set, for k large enough, it follows that $0 \in \partial_C \Psi_k(x^k)$, where $\partial_C \Psi(x)$ is the *Clarke subdifferential* of a locally Lipschitz function Ψ at x (we recall that all the functions of problem (VP) are continuously differentiable, and therefore locally Lipschitz). By applying some calculus rules (see Clarke (1983)), we have

$$0 \in co\left(\bigcup_{1 \leq \ell \leq p} \{\nabla f_\ell(x^k)\} \right) + x^k - x^0 + \sum_{i=1}^m k g_i(x^k)_+ \cdot \nabla g_i(x^k) + \sum_{j=1}^r k h_j(x^k) \cdot \nabla h_j(x^k).$$

Hence, there exists $\lambda_\ell^k \geq 0$, $\ell = 1, ..., p$, such that $\sum_{\ell=1}^p \lambda_\ell^k = 1$ and

$$\sum_{\ell=1}^p \lambda_\ell^k \nabla f_\ell(x^k) + \sum_{i=1}^m k g_i(x^k)_+ \cdot \nabla g_i(x^k) + \sum_{j=1}^r k h_j(x^k) \cdot \nabla h_j(x^k) = x^0 - x^k \longrightarrow 0.$$

Choosing $u_i = k g_i(x^k)_+ \geq 0$ and $w_j = k h_j(x^k)$ we see that x^0 satisfies AKKT (condition (18) is clearly satisfied, by the continuity of g_i) and this ends the proof. $\square$

References

J. ABADIE (1967), On the Kuhn-Tucker theorem; in J. Abadie (Ed.), Nonlinear Programming, North Holland, Amsterdam, 19-36.

R. ANDREANI, J. M. MARTINEZ, B. F. SVAITER (2010), A new sequential optimality condition for constrained optimization and algorithmic consequences, SIAM J. Optim., 20, 3533-3554.

R. ANDREANI, G. HAESER, J. M. MARTINEZ (2011), On sequential optimality conditions for smooth constrained optimization, Optimization, 60, 627-641.

R. ANDREANI, G. HAESER, M. L. SCHUVERDT, J. S. SILVA (2012), A relaxed constant positive linear dependence constraint qualification and applications, Math. Programming, Ser. A, 135, 255-273.

M. S. BAZARAA, C. M. SHETTY (1976), Foundations of Optimization, Springer Verlag, Berlin.

S. I. BIRBIL, J. B. G. FRENK, G. J. STILL (2007), An elementary proof of the Fritz John and Karush-Kuhn-Tucker conditions in nonlinear programming, European J. of Operational Research, 180, 479-484.

T. BUTLER, A. V. MARTIN (1962), On a method of Courant for minimizing functionals, J. Math. and Physics, 41, 291-299.

F. H. CLARKE (1983), Optimization and Nonsmooth Analysis, John Wiley & Sons, New York.

B. D. CRAVEN (1984), Modified Kuhn-Tucker conditions when a minimum is not attained, Oper. Res. Letters, 3, 47-52.

K.-H. ELSTER, R. GÖTZ (1972), Ueber eine Erweiterung des Satzes von Kuhn and Tucker, Wiss. Z. TH. Ilmenau, 18, 37-41.

J. P. EVANS, K. O. KORTANEK (1970), Asymptotic Lagrange regularity: infimum not assumed, Cahiers Centre d'Etudes de Recherche Opérationnelle, 12, 76-94.

A. V. FIACCO, G. P. McCORMICK (1967), The slacked unconstrained minimization technique for convex programming, SIAM J. Appl. Math., 15, 505-515.

A. V. FIACCO, G. P. McCORMICK (1968a), Asymptotic conditions for constrained minimization, RAC-PTP-340, Research Analysis Corporation, McLean, Virginia.

A. V. FIACCO, G. P. McCORMICK (1968b), Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley & Sons, New York.

J. GAUVIN (1977), A necessary and sufficient condition to have bounded multipliers in nonconvex programming, Math. Programming, 12, 136-138.

G. GIORGI (2011), Remarks on Fritz John conditions for problems with inequality and equality constraints, International Journal of Pure and Applied Mathematics, 71, 643-657.

F. J. GOULD, J. W. TOLLE (1971), A necessary and sufficient qualification for constrained optimization, SIAM J. Appl. Math., 20, 164-172.

F. J. GOULD, J. W. TOLLE (1972), Geometry of optimality conditions and constraint qualifications, Math. Programming, 2, 1-18.

M. GUIGNARD (1969), Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space, SIAM J. Control, 7, 232-241.

G. HAESER, M. L. SCHUVERDT (2011), On approximate KKT condition and its extension to continuous variational inequalities, J. Optim. Theory Appl., 149, 528-539.

K. O. KORTANEK, J. P. EVANS (1968), Asymptotic Lagrange regularity for pseudoconcave programming with weak constraint qualification, Operations Res., 16, 849-857.

O. L. MANGASARIAN, S. FROMOVITZ (1967), The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, J. Math. Anal. Appl., 17, 37-47.

M. MARTINEZ, B. F. SVAITER (2003), A practical optimality condition without constraint qualification for nonlinear programming, J. Optim. Theory Appl., 118, 117-133.

T. PIETRZYKOWSKI (1962), Application of the steepest descent method to concave programming; in Proceedings of the IFIPS Congress, Munich 1962, North Holland Company, Amsterdam, 185-189.

L. I. TRUDZIK (1981/82), Asymptotic Kuhn-Tucker conditions in abstract spaces, Numer. Funct. Anal. and Optim., 4. 355-369.

P. VARAIYA (1967), Nonlinear programming in Banach space, SIAM J. Appl. Math., 15, 284-293.

W. I. ZANGWILL (1967), Non-linear programming via penalty functions, Management Science, 13, 344-358.

S. ZLOBEC (1970), Asymptotic Kuhn-Tucker conditions for mathematical programming problems in a Banach space, SIAM J. Control, 8, 505-512.

S. ZLOBEC (1971a), Optimality conditions for mathematical programming problems, Glasnik Matematicki, 6, 187-191.

S. ZLOBEC (1971b), Extensions of asymptotic Kuhn-Tucker conditions in mathematical programming, SIAM J. Appl. Math., 21, 448-460.

S. ZLOBEC (1972), A class of optimality conditions for mathematical programming problems in Banach spaces, Glasnik Matematicki, 7, 127-137.