Iranian Journal of Operations Research
Vol. 5, No. 1, 2014, pp. 15-28

On Sequential Optimality Conditions without Constraint Qualifications for Nonlinear Programming with Nonsmooth Convex Objective Functions

S. Ahmadi[1,*], N. Movahedian[2]

Sequential optimality conditions provide adequate theoretical tools to justify stopping criteria for nonlinear programming solvers. Here, nonsmooth approximate gradient projection and complementary approximate Karush-Kuhn-Tucker conditions are presented. These sequential optimality conditions are satisfied by local minimizers of optimization problems independently of the fulfillment of constraint qualifications. It is proved that nonsmooth complementary approximate Karush-Kuhn-Tucker conditions are stronger than nonsmooth approximate gradient projection conditions. Sufficiency for differentiable generalized convex programming is established.

Keywords*: Optimization problems, Necessary optimality conditions, Constraint qualification, Necessary optimality conditions, Nonsmooth analysis.*

Manuscript was received on 25/08/2013, revised on 12/05/2014 and accepted for publication on 19/09/2014.

1. Introduction

Here, we study sequential optimality conditions for nonlinear programming with nonsmooth convex objective functions.

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 the constraint qualification, then it satisfies Karush-Kuhn-Tucker (KKT) conditions".

A constraint qualification is a property of the feasible points of a nonlinear programming problem that, if satisfied by a local minimizer, then the KKT conditions hold; e.g. see [3].

Practical methods for solving constrained optimization problems are iterative. At every iteration, one must decide, whether it is sensible to terminate the algorithm or not. Since testing 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. These facts lead one to study a different type of optimality conditions.

In [2, 6], sequential optimality conditions were introduced for nonlinear programs, where the authors observed that the new conditions are satisfied by local minimizers of constrained optimization problems independently of constraint qualifications and using only first-order differentiability.

* Corresponding Author.
1 Department of Mathematics, University of Isfahan, Isfahan, Iran. Email: s68.ahmady@gmail.com
2 Department of Mathematics, University of Isfahan, Isfahan, Iran. Email: n.movahedian@sci.ui.ac.ir

Therefore, those conditions are genuine necessary optimality conditions which do not use constraint qualifications at all.

Here, we assume that the objective function is not smooth, but rather a convex function. Then, we will formulate new nonsmooth sequential optimality conditions for the problem. We will prove that, similar to the smooth case, the new conditions are necessary optimality conditions independently of the fulfillment of constraint qualifications. Our alternative for the classical derivative $\nabla f(x)$ is the subdifferential defined for convex functions. From the optimization perspective, the subdifferential $\partial f(.)$ of a convex function f carries many of the useful properties of the derivative.

The remainder of our work is as follows.
In Section 2, we define *n*-AGP property and prove that it is satisfied by every local minimizer of a (convex) nonlinear programming problem and that it implies the Fritz-John conditions. In Section 3, we present *n*-CAKKT condition and show its satisfaction by local minimizers. In Section 4, we prove that *n*-CAKKT is a stronger condition than *n*-AGP. In Section 5, we show that CAKKT is a sufficient optimality condition for smooth convex-like problems.

Notations:

- For $h : \mathbb{R}^n \to \mathbb{R}^m$, we denote $\nabla \mathrm{h} = (\nabla h_1, \ldots, \nabla h_m)^T$.
- $\mathbb{R}_+ = \{x \in IR | x \geq 0\}$.
- If $\in \mathbb{R}^n$, we denote $v_+ = (\max\{v_1; 0\}, \ldots, \max\{v_n; 0\})^T$.
- The symbol $\|.\|$ denotes $\|.\|_2$.
- $\mathrm{B}(x, \delta) = \{z \in \mathbb{R}^n \mid \|z - x\| \leq \delta\}$.
- $P_\Omega(x)$ is the Euclidean projection of x on Ω.

2. Nonsmooth Approximate Gradient Projection Conditions

(Smooth) Approximate Gradient Projection (AGP) conditions were introduced in [6], where the authors observed that AGP is the optimality condition that fits the natural stopping criterion for inexact restoration methods. In [6], it was also proved that AGP implies the Fritz-John conditions. Here, we will define this condition for a nonsmooth convex problem.

Consider the following nonlinear programming problem:

$$\text{Minimize } f(x) \text{ s.t. } x \in \Omega, \tag{1}$$

where

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

The function $f: \mathbb{R}^n \to \mathbb{R}$ is convex (not necessarily smooth) and $h: \mathbb{R}^n \to \mathbb{R}^m$ and $g: \mathbb{R}^n \to \mathbb{R}^p$ are convex and continuously differentiable.

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

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

$$\nabla g_i(x)^T(z-x) \le 0, \text{ if } g_i(x) \ge 0, \tag{4}$$

and

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

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

$$\begin{array}{lll} h(z) = h(x) & & \\ g_i(z) \le g_i(x), & \text{if} & g_i(x) \le 0 \\ g_i(z) \le 0, & \text{if} & g_i(x) \in (\gamma, 0). \end{array}$$

Observe that

$$\Omega(x,0) = \{z \in \mathbb{R}^n | \nabla h(x)(z-x) = 0, \nabla g_i(x)^T(z-x) \le 0, \ \text{ if } \ g_i(x) \ge 0,$$
$$\nabla g_i(x)^T(y-x) + g_i(x) \le 0, \ \text{ if } \ g_i(x) < 0, i = 1, \dots, p\}.$$

The attractiveness of (smooth) AGP is that it does not involve Lagrange multipliers estimates. Instead, an approximate projected gradient of the objective function is used. Here, we will extend this concept to nonsmooth convex problem (1). For all $x \in \mathbb{R}^n$ and $\xi \in \partial f(x)$, we define $d(x,\gamma,\xi) \in \mathbb{R}^n$ as

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

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 and convex. The vector $d(x,\gamma,\xi)$ will be called approximate gradient projection(AGP) . Now, we are ready to define the nonsmooth version of AGP.

Definition 2.1. Let $\gamma \in (-\infty, 0]$. We say that a feasible point x^* of (2) satisfies the nonsmooth-AGP $(\gamma)\big(n - AGP(\gamma)\big)$ condition when there are sequences $x^k \to x^*$ and $\xi^k \in \partial f\big(x^k\big)$ satisfying

$$\lim_{k\to\infty} \big\|d\big(x^k, \gamma, \xi^k\big)\big\| = 0\,. \tag{7}$$

It is worth mentioning that, similar to the smooth case, n-AGP(γ) is equivalent to n-AGP($\gamma\,'$) for all $\gamma, \gamma\,' \in (-\infty, 0)$. For this reason, we will always write n-AGP instead of AGP(γ). This fact can be proved similar to the one of [6, Property 1].

The main result of this section is proved below. It says that if x^* is a local minimizer of (1) and f is convex, thenwe can find points with sufficiently small approximate gradient projections that are arbitrarily close to x^*.

Theorem 2.2. Assume that x^* is a local minimizer of (1), f is convex and *h, g* are continuously differentiable and convex. Let $\gamma \in (-\infty, 0]$ and $\varepsilon\,, \delta \in (0,\infty)$ be given. Then, there exist $x \in \mathbb{R}^n$ and $\xi \in \partial f(x)$ such that $\|x - x^*\| \le \delta$ and $\|d(x,\gamma,\xi)\| \le \varepsilon$.

Proof. Let $\rho \in (0, \delta)$ be such that x^* is a global minimizer of $f(x)$ on $\Omega \cap B(x^*, \rho)$. 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

$$\text{Minimize} \quad \varphi\,(x) \quad \text{s.t.} \quad x \in \Omega \cap B(x^*,\ \rho).$$

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

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

The external penalty theory (see Theorem 9.2.2 in [3]) guarantees that, for μ sufficiently large, there exists a solution of

$$\text{Minimize} \quad \varphi_\mu(x) \qquad \text{s.t.} \quad x \in\ \Omega \cap B(x^*,\ \rho) \tag{8}$$

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

$$0 \in \partial\phi_\mu\big(x_\mu\,\big).$$

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

$$0\ \in\ \partial\phi_\mu(x)\ \subseteq\ \partial\varphi(x) +\ \mu\left[\nabla h(x)^T h(x) + \sum_{g_i(x)>0} \nabla g_i(x) g_i(x)\right].$$

Since f is convex, $\partial f(x) \neq \emptyset$ and there exists $\xi \in \partial f(x)$ such that

$$\xi +\ \mu\left[\nabla h(x)^T h(x) + \sum_{g_i(x)>0} g_i(x)\nabla g_i(x)\right] + \frac{\varepsilon}{\rho}(x - x^*) = 0 \tag{9}$$

Since $x =\ x_\mu$ lies in the interior of the ball, we have that $\|x - x^*\|\ <\ \rho\ <\ \delta$ and by (9),

$$\left\|\xi +\ \mu\left[\nabla h(x)^T h(x) + \sum_{g_i(x)>0} \nabla g_i(x) g_i(x)\right]\right\| \leq \varepsilon.$$

So,

$$\left\|[x - \xi] - \left[x +\ \mu\left[\nabla h(x)^T h(x) + \sum_{g_i(x)>0} \nabla g_i(x) g_i(x)\right]\right]\right\| \leq \varepsilon.$$

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

$$\left\|P_{\Omega(x,\gamma)}(x - \xi) -\ P_{\Omega(x,\gamma)}\left(x +\ \mu\left[\nabla h(x)^T h(x) + \sum_{g_i(x)>0} \nabla g_i(x) g_i(x)\right]\right)\right\| \leq \varepsilon. \tag{10}$$

It remains to prove that

$$P_{\Omega(x,\gamma)}\left(x+\ \mu\left[\nabla h(x)^T h(x)+\ \sum_{g_i(x)>0}\nabla g_i(x)g_i(x)\right]\right)=x.$$

To see this, consider the convex quadratic subproblem

$$\text{Minimize}_y \quad \left\|y-\left[x+\ \mu\left[\nabla h(x)^T h(x)+\ \sum_{g_i(x)>0}\nabla g_i(x)g_i(x)\right]\right]\right\|^2$$

subject to $y\in\Omega(x,\gamma)$ and observe that $y=x$ satisfies the sufficient KKT optimality conditions with the multipliers $\lambda=2\mu h(x)$ and $v_i=2\mu g_i(x)$, for $g_i(x)\geq 0$, and $v_i=0$ otherwise.
So, by (10), $\left\|P_{\Omega(x,\gamma)}(x-\ \xi)-x\right\|\leq\varepsilon$ as desired.

The following corollary states the *n*-AGP property suited for use in an algorithmic framework: given a local minimizer, a sequence exists that satisfies $x_k\ \rightarrow\ x^*$ and $d(x_k,\gamma,\xi)\ \rightarrow 0$.

Corollary 2.3. If x^* is a local minimizer of (1), f is convex, h , g are continuously differentiable and convex and $\gamma\ \in(-\infty,0]$, thenthere exist sequences $\{x^k\}\subset\mathbb{R}^n$ and $\xi^k\ \in\ \partial f\left(x^k\right)$ such that $\lim_{k\rightarrow\infty}x^k=\ x^*$ and $\lim_{k\rightarrow\infty}d\left(x^k,\gamma,\xi^k\right)=0$.

Finally, we will prove that the *n*-AGP condition implies the Fritz-John optimality conditions. First, let us recall the equivalent formulation of the Mangasarian-Fromovitz Constraint qualification (see [7]). We say that the feasible point x^* of (1) satisfies the Mangasarian-Fromovitz Constraint Qualification (MFCQ) if for any vectors $\lambda\in\mathbb{R}^n$ and $\mu\in\mathbb{R}^p$, with $\mu\geq 0$, the following implication holds true:

$$[\nabla h(x^*)^T\lambda+\nabla g(x^*)^T\,\mu=\ 0\ \text{ and }\ \mu^T g(x^*)=\ 0]\ \Longrightarrow\ [\lambda=0 \text{ and } \mu=0].$$

To continue, let us define the Fritz-John and Karush-Kuhn-Tucker conditions for problem (1).

Definition 2.4. Consider the nonlinear programming problem (1), where f is convex and g , h are continuously differentiable and convex.

- We say that $x^*\ \in\ \mathbb{R}^n$ fulfills the nonsmooth-Fritz-John (*n*-FJ) conditions if
$$h(x^*)=0,g(x^*)\leq 0,$$
and there exist $\mu_0\in\mathbb{R}_+$, $\mu\ \in\ \mathbb{R}^p{}_+$ and $\lambda\ \in\ \mathbb{R}^m$ such that
$$0\ \in\ \mu_0\partial f(x^*)+\ \nabla h(x^*)^T\lambda+\ \nabla g(x^*)^T\mu,$$
$$\mu_i g_i(x^*)=0\quad ,i=1,\dots,p,$$
$$\left(\mu_0,\vec{\mu},\vec{\lambda}\right)\ \notin\ \left(0,\vec{0},\vec{0}\right).$$

- We say that $x^*\ \in\ \mathbb{R}^n$ fulfills the nonsmooth-Karush-Kuhn (*n*-KKT) conditions if
$$h(x^*)=0\ ,g(x^*)\leq 0,$$
and there exist $\mu\ \in\ \mathbb{R}^p{}_+$ and $\lambda\ \in\ \mathbb{R}^m$ such that
$$0\ \in\ \partial f(x^*)+\ \nabla h(x^*)^T\lambda+\ \nabla g(x^*)^T\mu,$$
$$\mu_i g_i(x^*)=0\quad ,i=1,\dots,p.$$

Now, we are ready to state the final result of this section which shows that *n*-FJ conditions hold at any feasible point with *n*-AGP property.

Theorem 2.5. Assume that x^* is a feasible point, f is convex and h , g are continuously differentiable and convex. Let $\gamma \in (-\infty, 0]$. Suppose that there are sequences $x^k \to x^*$ and $\xi^k \in \partial f(x^k)$ such that $d(x^k , \gamma , \xi^k) \to 0$. Then, x^* is a nonsmooth-Fritz-John point of (1).

Proof. Define, for each k, $y^k = P_{\Omega(x^k,\gamma)}(x^k - \xi^k)$. Obviously, y^k solves the following problem:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\left\|y - x^k\right\|^2 + (\xi^k)^T(y - x^k) \\ \text{Subject to} & y \in \Omega(x^k, \gamma). \end{array}$$

The above gives us $\lambda^k \in \mathbb{R}^m$ and $\mu^k \in \mathbb{R}^p$ such that $\mu^k \geq 0$ and

$$\xi^k + (y^k - x^k) + \nabla h(x^k)^T \lambda^k + \nabla g(x^k)^T \mu^k = 0 \tag{11}$$

$$\begin{array}{ll} \mu_i{}^k\left[g_i(x^k) + \nabla g_i^T(x^k)(y^k - x^k)\right] = 0, & \text{if} \quad \gamma < g_i(x^k) < 0, \\ \mu_i{}^k\left[\nabla g_i^T(x^k)(y^k - x^k)\right] = 0, & \text{if} \quad g_i(x^k) \geq 0. \\ \mu_i{}^k = 0, & \text{otherwise} \end{array}$$

Moreover, if $g_i(x^*) < 0$, thenwe have that $g_i(x^k) < 0$, for k large enough, and since $\left\|y^k - x^k\right\| \to 0$, we also have that $\nabla g_i(x^k)^T(y^k - x^k) < 0$. Therefore, we can assume

$$\mu_i{}^k = 0, \quad \text{whenever} \quad g_i(x^*) < 0. \tag{12}$$

To establish nonsmooth-Fritz-John conditions is equivalent to establish that MFCQ implies *n*-KKT conditions. So, from now on we are going to assume that x^* satisfies the MFCQ.

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

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

we have lim sup $M_k = \infty$. Refining the sequence (λ^k , μ^k) and reindexing it, we may suppose that $M_k > 0$ and $lim_k\ M_k = +\infty$. Now, define

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

Observing that $\left\|(\hat{\lambda}^k , \hat{\mu}^k)\right\|_\infty = 1$, the sequence $(\hat{\lambda}^k , \hat{\mu}^k)$ is bounded and has a cluster point $(\hat{\lambda} , \hat{\mu})$ satisfying

$$\hat{\mu} \geq 0, \qquad \left\|(\hat{\lambda} , \hat{\mu})\right\|_\infty = 1. \tag{13}$$

On the other hand, the convexity of f implies that it is locally Lipschitz on $\mathbb{R}^n$ (see [4, Theorem 4.1.1]). Thus, we can assume without generality that for each k , $\|\xi_k\| \leq L$, where $L > 0$ is the

Lipschitz constant of f near x^*. Therefore, with a similar argument, we may suppose that $\xi_k \to \hat{\xi} \in \partial f(x^*)$. Dividing (11) into M_k, we obtain:

$$(1/M_k)\left[\xi^k + \left(y^k - x^k\right)\right] + \nabla h\left(x^k\right)^T \hat{\lambda}^k + \nabla g\left(x^k\right)^T \hat{\mu}^k = 0.$$

Taking the limit along the appropriate subsequence, we conclude that

$$\nabla h(x^*)^T \hat{\lambda} + \nabla g(x^*)^T \hat{\mu} = 0.$$

Together with (13), this contradicts the MFCQ.

Now, since in (11), λ^k and μ^k are bounded, extracting convergent subsequences, we have that $\lambda^k \to \lambda^*$ and $\mu^k \to \mu^* \geq 0$. By (12), $g(x^*)^T \mu^* = 0$ and, taking limits in (11), we get

$$\begin{aligned} 0 &= \hat{\xi} + \nabla h(x^*)^T \hat{\lambda} + \nabla g(x^*)^T \hat{\mu} \\ &\in \partial f(x^*) + \nabla h(x^*)^T \hat{\lambda} + \nabla g(x^*)^T \hat{\mu}, \end{aligned}$$

to complete the proof.

Corollary 2.6. If the hypotheses of Theorem 2.5 hold and x* satisfies MFCQ, then the x^* is an *n*-KKT point.

3. Nonsmooth Complementary Approximate Karush-Kuhn-Tucker Conditions

Here, we will formulate new nonsmooth sequential optimality conditions called *nonsmooth Complementary Approximate Karush-Kuhn-Tucker* (*n*-CAKKT) conditions. This concept was introduced by Andreani, Martinez and Svaiter [2] for the smooth case of problem (1). In fact, we will generalize their notion to nonsmooth problems. We will also prove that, similar to the smooth case, *n*-CAKKT conditions are necessary optimality conditions independently of the fulfillment of constraint qualifications.

Definition 3.1. We say that $x^* \in \mathbb{R}^n$ satisfies the nonsmooth Complementary Approximate Karush-Kuhn-Tucker (*n*-CAKKT) conditions for problem (1) if

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

and there exists a sequence $x_k \to x^*$ such that

- for all $k \in \mathbb{N}$, there exist $\lambda^k \in \mathbb{R}^m, \mu^k \in \mathbb{R}^p{}_+, \xi^k \in \partial f\left(x^k\right)$ such that

$$\lim_{k \to \infty} \left\| \xi^k + \nabla h\left(x^k\right)^T \lambda^k + \nabla g\left(x^k\right)^T \mu^k \right\| = 0, \tag{14}$$

$$\lim_{k \to \infty} \sum_{i=1}^{m} \left|\lambda_i{}^k h_i\left(x^k\right)\right| + \sum_{i=1}^{p} \left|\mu_i{}^k g_i\left(x^k\right)\right| = 0. \tag{15}$$

The points satisfying *n*-CAKKT conditions will be called *n*-CAKKT points. Clearly, *n*-KKT conditions imply *n*-CAKKT conditions. A useful property of *n*-CAKKT is that they are also satisfied at local minimizers when, due to the lack of fulfillment of constraint qualifications, the *n*-KKT conditions do not hold. The following result shows that *n*-CAKKT conditions are necessary optimality conditions without any constraint qualification.

Theorem 3.2. Let x^* be a local minimizer of (1). Then, x^* satisfies *n*-CAKKT conditions.

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

$$\text{Minimize } f(x) + \|x - x^*\|_2{}^2 \text{ s.t. } h(x) = 0, g(x) \leq 0, x \in B(x^*, \delta). \tag{16}$$

Clearly, x^* is the unique solution of (16). Now, take the sequence $\rho_k > 0$ such that $\rho_k \to \infty$ and suppose that x^k is a solution of the following problem:

$$\begin{gathered}\text{Minimize } f(x) + \|x - x^*\|_2{}^2 + \tfrac{\rho_k}{2}\left[\left\|h(x^k)\right\|_2{}^2 + \textstyle\sum_{i=1}^{p} g_i(x^k)_+{}^2\right] \\ \text{s.t. } x \in B(x^*, \delta).\end{gathered} \tag{17}$$

By the compactness of $B(x^*, \delta)$, x^k is well defined for all k. Moreover, since $x^* \in B(x^*; \delta)$, $h(x^*) = 0$ and $g(x^*)_+ = 0$, we have

$$f(x^k) + \|x^k - x^*\|_2{}^2 + \frac{\rho_k}{2}\left[\left\|h(x^k)\right\|_2{}^2 + \sum_{i=1}^{p} g_i(x^k)_+{}^2\right] \leq f(x^*). \tag{18}$$

Now, let us prove that $\lim_{k\to\infty} x^k = x^*$. Since $\{x^k\} \subset B(x^*, \delta)$, the compactness of $B(x^*, \delta)$ gives us some $\tilde{x} \in B(x^*, \delta)$ such that $x^k \to \tilde{x}$. If we assume that $\tilde{x}$ is a feasible point for (1) , then we easily get

$$f(x^*) \leq f(\tilde{x}) + \|\tilde{x} - x^*\|_2{}^2 \leq f(x^*),$$

which implies that $\tilde{x} = x^*$.
Now, suppose that $\tilde{x}$ is not a feasible point. Hence, we have

$$a = \left[\|h(\tilde{x})\|_2{}^2 + \sum_{i=1}^{p} g_i(\tilde{x})_+{}^2\right] > 0.$$

From (18), we obtain:

$$\frac{\rho_k}{2}\left[\left\|h(x^k)\right\|_2{}^2 + \sum_{i=1}^{p} g_i(x^k)_+{}^2\right] \leq f(x^*) - f(x^k).$$

Since f is continuous, (again using [4, Theorem 4.1.1]), $f(x^k) \to f(x^*)$ and $\{f(x^*) - f(x^k)\}$ is bounded. On the other hand, since $\left[\left\|h(x^k)\right\|_2{}^2 + \sum_{i=1}^{p} g_i(x^k)_+{}^2\right] = 0$, the sequence $\{\rho_k\}$ is bounded, which contradicts $\rho_k \to \infty$.
Thus, we have $\lim_{k\to\infty} x^k = x^*$. Therefore, by (18) and the continuity of f, we have

$$\lim_{k\to\infty}\left\|x^k - x^*\right\|_2^{\,2} + \frac{\rho_k}{2}\left[\left\|h\left(x^k\right)\right\|_2^{\,2} + \sum_{i=1}^{p} g_i\left(x^k\right)_+^{\,2}\right] = 0.$$

Thus,

$$\lim_{k\to\infty}\left[\sum_{i=1}^{m}\rho_k h_i\left(x^k\right)^2 + \sum_{i=1}^{p}\rho_k g_i\left(x^k\right)_+^{\,2}\right] = 0.$$

Defining the vectors

$$\lambda^k = \rho_k h\left(x^k\right), \qquad \mu^k = \rho_k g_i\left(x^k\right)_+, \tag{19}$$

we obtain:

$$\lim_{k\to\infty}\left[\sum_{i=1}^{m}\left|\lambda_i^{\,k} h_i\left(x^k\right)\right| + \sum_{i=1}^{p}\mu_i^{\,k} g_i\left(x^k\right)_+\right] = 0. \tag{20}$$

Furthermore, by (19), we conclude that if $g_i\left(x^k\right) < 0$, then $g_i\left(x^k\right)_+ = 0$ and $\mu_i^{\,k} = 0$. Therefore, using (20), we obtain:

$$\lim_{k\to\infty}\left[\sum_{i=1}^{m}\left|\lambda_i^{\,k} h_i\left(x^k\right)\right| + \sum_{i=1}^{p}\left|\mu_i^{\,k} g_i\left(x^k\right)\right|\right] = 0. \tag{21}$$

Thus, (15) follows from (19) and (21). For k large enough, we have $\left\|x^k - x^*\right\| < \delta$. Consequently, x^k is a local optimal solution of the following unconstrained problem:

$$\text{Minimize } f(x) + G_k(x) \qquad \text{s.t.} \quad \|x - x^*\| < \delta, \tag{22}$$

,where

$$G_k(x) = \frac{\rho_k}{2}\left[\left\|h\left(x^k\right)\right\|_2^{\,2} + \sum_{i=1}^{p} g_i\left(x^k\right)_+^{\,2}\right].$$

Thus, we get

$$0 \in \partial(f + G_k)\left(x^k\right) \subseteq \partial f\left(x^k\right) + \nabla G_k\left(x^k\right).$$

The above gives us the sequence $\xi^k \in \partial f\left(x^k\right)$ such that for all k,

$$\xi^k + \rho_k\left[\sum_{i=1}^{m} h_i\left(x^k\right)\nabla h_i\left(x^k\right) + \sum_{i=1}^{p} g_i\left(x^k\right)_+\nabla g_i\left(x^k\right)\right] = 0. \tag{23}$$

Using (19) and taking limit as $k \to \infty$, we arrive at

$$\lim_{k\to\infty} \xi^k + \left[\sum_{i=1}^{m} \lambda_i{}^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k)\right] = 0,$$

to complete the proof.

4. Strength of *n*-CAKKT Conditions

Necessary optimality conditions should be as strong as possible. In this section we will see that *n*-CAKKT conditions are indeed strong. We will prove that the fulfillment of *n*-CAKKT conditions imply the fulfillment of *n*-AGP conditions. In Section 3, we proved that *n*-AGP conditions are strong optimality conditions in the sense that they imply *n*-KKT conditions or the absence of a weak constraint qualification. These results suggest that the points satisfying *n*-CAKKT conditions are more likely to be local minimizers than points merely verifying *n*-AGP conditions, or the points not satisfying weak (nonsmooth) constraint qualifications. Therefore, the points that approximate an *n*-CAKKT point have more chances to be close to local minimizers than the points that approximate an *n*-AGP point.
The next theorem shows that *n*-CAKKT conditions imply *n*-AGP conditions.

Theorem 4.1. Assume that x^* is a feasible *n*-CAKKT point of (1). Then, x^* satisfies the *n*-AGP conditions.

Proof. Assume that $\{x_k\} \subset \mathbb{R}^n$ converges to x^* and satisfies (14) and (15). Since f is convex, $\partial f(x^k) \neq \emptyset$. Therefore, there exists $\xi^k \in \partial f(x^k)$. Let y^k be the solution of

$$\begin{aligned} &\text{Minimize } \left\|[x^k - \xi^k] - y\right\|_2^2 \\ &\text{s.t. } y \in \Omega(x^k, 0), \end{aligned} \tag{24}$$

where $\Omega(x^k, 0)$ is the set of points $y \in \mathbb{R}^n$ satisfying

$$\begin{aligned} &\nabla h_i(x^k)^T (y - x^k) = 0, && \text{for } i = 1, \ldots, m \\ &\nabla g_i(x^k)^T (y - x^k) \leq 0, && \text{if } g_i(x^k) \geq 0 \\ &\nabla g_i(x^k)^T (y - x^k) + g_i(x^k) \leq 0, && \text{if } g_i(x^k) < 0. \end{aligned}$$

The objective function in (24) is a strictly convex quadratic function and $\Omega(x^k, 0)$ is defined by linear constraints. Since $x^k \in \Omega(x^k, 0)$, $\Omega(x^k, 0)$ is nonempty, y^k exists and is the unique solution of the problem. We wish to show that $\lim_{k\to\infty} \|y^k - x^k\| = 0$. Since the constraints of (24) are linear, the KKT conditions are fulfilled at y^k. Therefore, there exist $\{\hat{\lambda}^k\} \subset \mathbb{R}^m$, $\{\hat{\mu}^k\} \subset \mathbb{R}^p{}_+$ and $\xi^k \in \partial f(x^k)$ such that

$$[y^k - x^k] + \xi^k + \nabla h(x^k)^T \hat{\lambda}^k + \nabla g(x^k)^T \hat{\mu}^k = 0, \tag{25}$$

$$\nabla h_i(x^k)^T (y^k - x^k) = 0, \quad i = 1, \ldots, m, \tag{26}$$

$$\nabla g_i(x^k)^T (y^k - x^k) \leq 0, \quad \text{if } g_i(x^k) \geq 0, \tag{27}$$

$$g_i(x^k) + \nabla g_i(x^k)^T (y^k - x^k) \leq 0, \quad \text{if } g_i(x^k) < 0. \tag{28}$$

$$\hat{\mu}_i{}^k \nabla g_i(x^k)^T (y^k - x^k) = 0, \quad \text{if } g_i(x^k) \geq 0, \tag{29}$$

and

$$\hat{\mu}_i{}^k g_i(x^k) + \hat{\mu}_i{}^k \nabla g_i(x^k)^T (y^k - x^k) = 0\,, \qquad \text{if } g_i(x^k) < 0\,. \tag{30}$$

Pre-multiplaying (25) by $(y^k - x^k)^T$ and using (26 - 29), we obtain:

$$\|y^k - x^k\|_2{}^2 + (\xi^k)^T (y^k - x^k) + \sum_{g_i(x^k)<0} \hat{\mu}_i{}^k \nabla g_i(x^k)^T (y^k - x^k) = 0\,. \tag{31}$$

If $g_i(x^k) < 0$, then by (30) we have

$$\hat{\mu}_i{}^k \nabla g_i(x^k)^T (y^k - x^k) = -\hat{\mu}_i{}^k g_i(x^k).$$

Therefore, applying (31), we arrive at

$$\|y^k - x^k\|_2{}^2 + (\xi^k)^T (y^k - x^k) = \sum_{g_i(x^k)<0} \hat{\mu}_i{}^k g_i(x^k)^T\,.$$

Since $\hat{\mu}^k \geq 0$, we have

$$\|y^k - x^k\|_2{}^2 \leq -(\xi^k)^T (y^k - x^k). \tag{32}$$

On the other hand, using (14), we can find sequences $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p{}_+, \xi^k \in \partial f(x^k)$ and $\{v^k\} \subset \mathbb{R}^n$ such that

$$\xi^k + \sum_{i=1}^{m} \lambda_i{}^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k) = v^k \to 0\,.$$

Therefore,

$$\begin{aligned}
-(\xi^k)^T (y^k - x^k) \\
&= \sum_{i=1}^{m} \lambda_i{}^k \nabla h_i(x^k)^T (y^k - x^k) \\
&\quad + \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k\,.
\end{aligned}$$

Thus, by (26),

$$-(\xi^k)^T (y^k - x^k) = \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k\,.$$

$$= \sum_{g_i(x^k)<0} \mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) + \sum_{g_i(x^k)\geq 0} \mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k.$$

If $g_i(x^k) \geq 0$, then (27) together with the fact that $\mu^k \geq 0$ lead to $\mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) \leq 0$, and hence

$$-(\xi^k)^T(y^k - x^k) \leq -\sum_{g_i(x^k)\geq 0} \mu_i{}^k \nabla g_i(x^k)^T (y^k - x^k) - (y^k - x^k)^T v^k$$
$$= \sum_{g_i(x^k)<0} \mu_i{}^k \left[g_i(x^k) + \nabla g_i(x^k)^T (y^k - x^k)\right] - \sum_{g_i(x^k)\geq 0} \mu_i{}^k g_i(x^k) - (y^k - x^k)^T v^k.$$

Thus, by (28) we have

$$\begin{aligned} -(\xi^k)^T(y^k - x^k) &\leq -\sum_{g_i(x^k)<0} \mu_i{}^k g_i(x^k) - (y^k - x^k)^T v^k \\ &\leq -\sum_{g_i(x^k)\geq 0} \mu_i{}^k g_i(x^k) + \|v^k\|_2 \|y^k - x^k\|_2 \\ &\leq \sum_{g_i(x^k)<0} \left|\mu_i{}^k g_i(x^k)\right| + \|v^k\|_2 \|y^k - x^k\|_2 . \end{aligned}$$

Finally, the above inequality and the one in (32) give us

$$\|y^k - x^k\|_2 \left(\|y^k - x^k\|_2 - \|v^k\|_2\right) \leq \sum_{g_i(x^k)<0} \left|\mu_i{}^k g_i(x^k)\right| . \tag{33}$$

Now, if for all k , $\|y^k - x^k\|_2 \leq \|v^k\|_2 \downarrow 0$, then there remains nothing to prove. Now, suppose that for all k, $\|y^k - x^k\|_2 > \|v^k\|_2$. Then, by (15) we get

$$0 \leq \|y^k - x^k\|_2 \left(\|y^k - x^k\|_2 - \|v^k\|_2\right) \leq \sum_{g_i(x^k)<0} \left|\mu_i{}^k g_i(x^k)\right| \downarrow 0 ,$$

which means that $\lim_{k\to\infty} \|y^k - x^k\| = 0$, as desired. Therefore, x^* satisfies the *n*-AGP conditions.

5. Sufficient Optimality Condition

In this section, we will show that under some generalized convexity assumptions, CAKKT conditions are sufficient optimality conditions for global minimizers.

Theorem 5.1. Assume that, for the problem (1) all the functions are continuously differentiable, the objective function f is pseudoconvex, the inequality constraints $g_i(x)$, for $i = 1, \dots, p$, are quasiconvex and the equality constraints $h_i, i = 1, \dots, m$, are affine. Let x^* be a feasible point that satisfies the CAKKT conditions. Then, x^* is a global minimizer of (1).

Proof. Assume that x^k , λ^k , μ^k are given by (14) and (15). Let z be a feasible point and define

$$I = \{i \in 1, \dots, p | g_i(x^*) = 0\},$$
$$I_k = \{i \in 1, \dots, p | g_i(x^k) = 0\}.$$

Thus, for all $i \in I_k$, we have $g_i(x^k) = 0$ and by the feasiblity of z, we have $g_i(z) \le g_i(x^k)$. The quasiconvexity of g_i implies that $\nabla g_i(x^k)^T(z - x^k) \le 0$. Also, since the h_i are affine, we have

$$h_i(z) = h_i(x^k) + \nabla h_i(x^k)^T(z - x^k) = 0, \qquad i = 1, \dots, m.$$

Thus, we get

$$\begin{aligned}\nabla f(x^k)^T(z - x^k) &\\ &+ \sum_{i=1}^{m} \lambda_i{}^k \left[h_i(x^k) + \nabla h_i(x^k)^T(z - x^k)\right] + \sum_{i \in I_k} \mu_i{}^k \left[\nabla g_i(x^k)^T(z - x^k)\right] \\ &\le \nabla f(x^k)^T(z - x^k)\end{aligned} \tag{34}$$

Observe that for all $i \in I_k$, we have $g_i(x^k) = 0$, and by (15), $\mu_i{}^k > 0$. In addition, for all $i \notin I_k$, we have $g_i(x^k) < 0$, and by (15), $\mu_i{}^k = 0$. Hence, we have

$$\sum_{i \in I_k} \mu_i{}^k \left[\nabla g_i(x^k)^T(z - x^k)\right] = \sum_{i=1}^{p} \mu_i{}^k \left[\nabla g_i(x^k)^T(z - x^k)\right] \tag{35}$$

Also, since $g_i(x^k) = 0$ and $\mu_i{}^k \ge 0$, we obtain:

$$\sum_{i=1}^{p} \mu_i{}^k g_i(x^k) \le 0. \tag{36}$$

Therefore, by (34), (35) and (36) we have

$$\begin{aligned}&\left[\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i{}^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k)\right]^T (z - x^k) \\ &\qquad + \sum_{i=1}^{m} \lambda_i{}^k h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k g_i(x^k) \le \nabla f(x^k)^T (z - x^k).\end{aligned}$$

Thus,

$$\begin{aligned}&\lim_{k \to \infty} \left[\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i{}^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k \nabla g_i(x^k)\right]^T (z - x^k) \\ &\qquad + \lim_{k \to \infty} \left[\sum_{i=1}^{m} \lambda_i{}^k h_i(x^k) + \sum_{i=1}^{p} \mu_i{}^k g_i(x^k)\right] \le \lim_{k \to \infty} \left[\nabla f(x^k)^T(z - x^k)\right].\end{aligned}$$

Taking the limit, as $k \to \infty$, continuous differentiability of the functions and assumptions (14) and (15), we get

$$\nabla f(x^k)^T(z - x^*) \ge 0.$$

Since f is pseudoconvex, we get

$$f(z) \geq f(x^*),$$

and this complete the proof.

References

[1] Andreani, R., Haeser, G. and Martínez, J.M. (2011), On sequential optimality conditions for smoothconstrained optimization, *Optimization*, 60, 11-19.

[2] Andreani, R., Martínez, J.M. and Svaiter, B.F. (2010), A new sequential optimality condition for constrained optimization and algorithmic consequences, *SIAM Journal on Optimization*, 20, 3533-3554.

[3] Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993), Nonlinear Programming: Theory and Algorithms,Wiley, New York.

[4] Borwein, J.M. and Lewis, A.S. (2000), Convex Analysis and Nonlinear Optimization, Theory and Examples, CMS Books in Mathematics, Springer-Verlag, New York.

[5] Fiacco, A.V. and McCormick, G.P. (1968), Nonlinear Programming, John Wiley, New York.

[6] Martínez, J.M. and Svaiter, B.F. (2003), A practical optimality condition without constraint qualifications for nonlinear programming, *Journal of Optimization Theory and Applications*, 118, 117-133.

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