OPITMALITY CONDITIONS WITH CONSTRAINT QUALIFICATION FOR NON LINEAR PROGRAMMING PROBLEMS SOLVE BY A CONE APPROACH ON THE KARUSH – KUHN – TUCKER CONCEPTS

Seifu Endris, Department of mathematics, College of Natural & Computational Sciences, Debre Berhan University, Ethiopia

Prof. A.N.Mohamad, Department of Mathematical Modeling, College of Natural & Computational Sciences, Debre Berhan university, Ethiopia

---

Abstract: *We establish the existence of optimization problem specially for Non Linear Programming (NLP), we introduce "A cone Approach on the Karush Kuhn Tucker Optimality Conditions and Constraint qualification" and we discussed in this paper some of the constraint qualifications as well as some relation between them and also to show and observe the weakest of these constraint qualifications with respect to the concept of cones and their polars*

Key Words: *Polar Cones, the Tangent Cone, Slater Constraint Qualification and the Quasi regularity constraints*

INTRODUCTION:

In mathematics, Non Linear Programming (NLP) is the process of solving an optimization problem defined by a system of equalities and inequalities, collectively termed constraints, over a set of unknown real variables, along with an objective function to be maximized or minimized, where some of the constraints or the objective functions are nonlinear [3]

In optimization, the Karush – Kuhn – Tucker (KKT) conditions (also known as Kuhn – Tucker – Conditions) are first order necessary conditions for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. Allowing inequality constraints, the KKT approach to nonlinear programming generalizes the method of Lagrange multipliers, which allows only equality constraints. The system of equations corresponding to the KKT conditions is usually not solved directly, except in the few special case where a closed – form solution can be derived analytically. In general, many optimization algorithms can be interpreted as methods for numerically solving the KKT system of equations [3], [5]

The KKT conditions were originally named after Harold W. Kuhn, and Albert W. Tucker, who first published the conditions [9]. Later scholars discovered that that the necessary conditions for this problem had been stated by William Karush in his master's thesis and Kjeldsen (2005) [7], [8]

Under differentiability and constraint qualifications, the KKT conditions provide necessary conditions for a solution to be optimal. Under convexity, these conditions are also sufficient. If some of the functions are non – differentiable, sub differential versions of KKT conditions are available [9]

Consider the standard NLP:

$$\left.\begin{array}{ll}\text{Min:} & f(x) \\ \text{Subject to} & g_j(x) \leq 0, \forall\, j = 1, 2, \dots, p \\ & h_i(x) = 0, \forall\, i = 1, 2, \dots, m\end{array}\right\} \longrightarrow \text{(P)}$$

Where the functions $f(x){:}\,\mathbb{R}^n \longrightarrow \mathbb{R}$; $g_j(x){:}\,\mathbb{R}^n \longrightarrow \mathbb{R}^p$ *and* $h_i(x){:}\,\mathbb{R}^n \longrightarrow \mathbb{R}^m$, they are conditionally differentiable

The feasible set of problem (P) will be denoted by Ω i.e.

$$\Omega = \left\{x \in \mathbb{R}^n \middle| g_j(x) \leq 0 \ \ and\ h_i(x) = 0\right\}$$

The classical KKT condition at given $\overline{x} \in \Omega$, and then there exists Lagrangian multipliers $\lambda \in \mathbb{R}^m$ *and* $\mu \in \mathbb{R}^p$ Such that

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\overline{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\overline{x}) = 0 \ \ where$$

$$\mu_j \geq 0$$

$$\mu_j \,.\, g_j(\overline{x}) = 0$$

In order to have an optimal solution to the given NLP problem, the KKT necessary condition has to be satisfied. The above optimality criteria has been used to formulate algorithms that solve (P) in the presence of any constraint qualification. These algorithms use the cones of directions of constancy. However, if x solves (P), but x, it is not a Kuhn – Tucker point (KKP), i.e. the KKT conditions do not hold at x, and then the program (P) is "Unstable" i.e. the "Perturbation" function, which is the optimal value of (P) as a function of perturbations of its right handed side, may decrease infinitely steeply in some direction. Thus, though we may solve (P), in practice our solution may be nowhere near the true solution. It is therefore of interest to know beforehand whether or not x, it is a KKP. Now, if a constraint qualification holds at x, and then x, it is necessarily a KKP for all objective functions which achieves a constrained minimum at x: Program (P) is therefore "Stable" at x for all such objective functions

If the problem is unconstraint, then the KKT conditions reduces to $\nabla f(\overline{x}) = 0$, which is a necessary optimality condition; however, this will not always be true, see the following counter example:

COUNTER EXAMPLE:

Consider the problem (P) with $h(x): \mathbb{R}^2 \longrightarrow \mathbb{R};\ g(x): \mathbb{R}^2 \longrightarrow \mathbb{R}^2\ for\ j = 1, 2$ defined by

$h(x) = -x_1$, and

$g(x) = [x_2 - (1 - x_1)^3, -x_2]^T$

Note that $x^* = (1, 0)^T$, it is minimizers of the problem but the KKT condition do not hold

In this paper we prove the KKT condition supposing the equality between the polar of the tangent cone and the polar of the first order feasible variations cone. Although this condition is the weakest assumption, it is extremely difficult to be verified. Therefore, other constraint qualifications, which are easier to be verified, will be discussed as: Slater's Linear independence of Gradient, Mangasarian – Fromovitz's and quasi – regularity. In general, we call a property of the feasible set a constraint qualification if it guarantees the KKT conditions to hold at a local minimizer.

Several mathematicians obtained different constraint qualifications. In this research, we will discuss many of them as well as some relations between them. A special interest is devoted to show the weakest such qualification

Notation:

Given $\overline{x} \in \Omega$, and the set $A(\overline{x})$, it denotes the set of inequality active constraint indices that is

$$A(\overline{x}) = \{j | g_j(\overline{x}) = 0, 1 \leq j \leq p\} \longrightarrow \quad (1)$$

1. Some Mathematical Preliminaries:

We need some preliminary definitions, results and relatively important cones which are quite relevant to prove KKT theorem or conditions

The **cones** of the direction of constancy are used to derive: New as well as known optimality conditions, weakest constraint qualifications; and regularization techniques, for the NLP problem. In addition the "Badly behaved set" of constraints, i.e. the set of constraints which causes problem in the KKT theory, is isolated and a computational procedure for checking whether a feasible point is regular or not is presented.

Cones:

Polar Cones:

Definition – 1:

A sub set $S\ of\ \mathbb{R}^n$, it is called a **convex set** if two points $x, y \in X\ \ and\ \lambda \in [0,1]$, such that $(1-\lambda)x + \lambda y \in S$

Definition – 2:

A subset $C\ of\ \mathbb{R}^n$, it is a **cone** when $td \in C, for\ all\ t \geq 0\ and\ d \in C$

Definition – 3:

Given a set $S \subset \mathbb{R}^n$, the **polar of** $\boldsymbol{S}$, it is given by $P(S) = \{p \in \mathbb{R}^n | p^T x \leq 0, for\ all\ x \in S\}$

Note that for any $S \subset \mathbb{R}^n\ then\ P(S)$, it is a cone and also $S \subset P\big(P(S)\big)$. This holds with equality if S, it is a closed, convex cone, as established by Farkas' Lemma [6] as shown below

Farkas' Lemma – 1:

Let $C \subset \mathbb{R}^n$, it is a closed convex cone, and then $P\big(P(C)\big) = C$

Proof:

For any $x \in C$, we have $x^T y \leq 0, for\ all\ y \in P(C) \Longrightarrow x \in P\big(P(C)\big) \Longrightarrow C \subset P\big(P(C)\big)$

Conversely, take $z \in P\big(P(C)\big)\ and\ let\ \hat{z} = proj_C^{(z)} \in C \Longrightarrow (z-\hat{z})^T(x-\hat{z}) \leq 0, for\ all\ x \in C,$

$Taking\ x = 0\ \ and\ x = 2\hat{z}$, we obtain $(z-\hat{z})^T\hat{z} = 0 \Longrightarrow (z-\hat{z})^T x \leq 0, for\ all\ x \in C$

$\Longrightarrow (z-\hat{z}) \in P(C)\ \ and\ since\ z \in P\big(P(C)\big)$, we have $(z-\hat{z})^T\hat{z} = 0 \Longrightarrow \|z-\hat{z}\|^2 = 0$

$\Longrightarrow z = \hat{z}\ \ and\ z \in C \Longrightarrow P\big(P(C)\big) \subset C \Longrightarrow C = P\big(P(C)\big)$

Definition – 4:

Let S, it is a nonempty set in $\mathbb{R}^n$, and let $\overline{x} \in \Omega$. The **cone of feasible direction** of $S\ at\ \overline{x}$ it is denoted by V, and given by

$V(\overline{x}) = \{d \in \mathbb{R}^n | \overline{x} + \lambda d \in S, for\ all\ \lambda \in (0,\delta]\ for\ some\ \delta > 0\}$

Each nonzero vector $d \in V$, it is called a **feasible direction**.

Definition – 5:

Given a function $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, **the cone of descent directions** at $\overline{x}$ it is denoted by F, and given by $F(\overline{x}) = \{d \in \mathbb{R}^n | f(\overline{x} + \lambda d) < f(\overline{x}), for\ all\ \lambda \in (0,\delta]\ for\ some\ \delta > 0\}$

Each direction $d \in F$, it is called a **descent direction** of $f\ at\ \overline{x}$

Lemma – 2:

Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, it is a differentiable function at a point $\overline{x} \in \mathbb{R}^n$, and then

a. $\nabla f(\overline{x})^T d \leq 0\ , for\ all\ d \in\ F(\overline{x})$
b. If $d \in \mathbb{R}^n$ satisfies $\nabla f(\overline{x})^T d < 0\ , and\ then\ d \in\ F(\overline{x})$, we get the set and denoted by
   $F_0(\overline{x}) = \{d \in \mathbb{R}^n | \nabla f(\overline{x})^T d < 0\} \longrightarrow (2)$

Proof of (a):

Let $d \in F(\overline{x})$ and for $some$ $\delta > 0$ and for all $\lambda \in [0, \delta]$, we have

$f(\overline{x} + \lambda d) < f(\overline{x})$ $i.e. f(\overline{x} + \lambda d) - f(\overline{x}) < 0 \Longrightarrow \nabla f(\overline{x}) = \lim_{\lambda \to 0^+} \frac{f(\overline{x}+\lambda d) - f(\overline{x})}{\lambda} \leq 0$

$\Longrightarrow \nabla f(\overline{x})^T d \leq 0$

Proof of (b) obviously true by the definition of cone of descent directions

Definition – 6:

The **set of first order feasible variations** at a point $\overline{x} \in \Omega$, it is the set

$D(\overline{x}) = \{d \in \mathbb{R}^n | \nabla h_i(\overline{x})^T d = 0, \forall\, i = 1, \dots, m$ and $\nabla g_j(\overline{x})^T d \leq 0, \forall\, j \in A(\overline{x})\} \longrightarrow$ (3)

$where$ $A(\overline{x})$, it is the active set defined by (1)

Note that $D(\overline{x})$, it is a nonempty closed, convex cone; it is often said that this cone is a linear approximation of the feasible set

Again, given $\overline{x} \in \Omega$, define the cone

$G(\overline{x}) = \{\sum_{i=1}^{m} \lambda_i \nabla h_i(\overline{x}) + \sum_{j \in A(\overline{x})} \mu_j \nabla g_j(\overline{x}) | \mu_j \geq 0, \forall\, j \in A(\overline{x})\} \longrightarrow$ (4)

To further proceed we need some properties of this cone, and we need the classical result, named Caratheodory's Lemma

Caratheodory's Lemma – 3:

Let $u_1, u_2, \dots, u_r$ they are nonzero vectors in $\mathbb{R}^n$ for $m < r$ and $x \in \mathbb{R}^n$ such that

$x = \sum_{i=1}^{r} \gamma_i u_i$ $with$ $\gamma_i \geq 0, for$ all $i > m$, and then there exist indices subset

$I \subset \{1, \dots, m\}$ and $J \subset \{m+1, \dots, r\}$ and $scalars$ $\gamma_i^{'}$ $where$ $i \in I \cup J$ $with$ $\gamma_i^{'} \geq 0, for$ $i \in J$

Such that $x = \sum_{i \in I \cup J} \gamma^{'} u_i$, and the vectors u_i for $i \in I \cup J$, they are linearly independent

Proof:

Suppose the vectors $u_1, u_2, \dots, u_r$, they are linearly independent, there is nothing to prove

Assume that $u_1, u_2, \dots, u_r$, they are linearly dependent, so, there exist scalars $\alpha_1, \dots, \alpha_r$, not all

$\alpha_i = 0$, such that $\sum_{i=1}^{r} \alpha_i u_i = 0 \Longrightarrow for$ all $t \in \mathbb{R}$ $then$ $x = \sum_{i=1}^{r} (\gamma_i - t\alpha_i) u_i$

Define $\overline{t}$, as t of minimum absolute value that vanish one of the coefficients $(\gamma_i - t\alpha_i)$. Then

$x = \sum_{i=1}^{r} (\gamma_i - \overline{t}\alpha_i) u_i$, $with$ $(\gamma_i - \overline{t}\alpha_i) \geq 0, for$ all $i > m$

$\Longrightarrow x$, it is written as a linear combination using of no more $r - 1$vectors.

We can repeat this process until that all vectors of the linear combination are linearly independent

Lemma – 4:

For any $\overline{x} \in \Omega$ then $\mathrm{G}(\overline{x})$, it is a closed convex cone

Proof:

To prove $\mathrm{G}(\overline{x})$ which is defined in (4), it is convex, consider $\mathrm{A}(\overline{x}) = \{1, \dots, \mathrm{q}\}$ for $x_1, x_2 \in \mathrm{G}(\overline{x})$ and $t \in [0,1]$, and then there exists $\lambda, \alpha \in \mathbb{R}^m$ *and* $\mu, \beta \in \mathbb{R}_+^q$ *such that*

$$x_1 = \sum_{i=1}^{m} \lambda_i \nabla h_i(\overline{x}) + \sum_{j=1}^{q} \mu_j \nabla g_j(\overline{x}) \ \ and \ \ x_2 = \sum_{i=1}^{m} \alpha_i \nabla h_i(\overline{x}) + \sum_{j=1}^{q} \beta_j \nabla g_j(\overline{x})$$

$$\Longrightarrow tx_1 + (1-\mathrm{t})x_2 = \sum_{i=1}^{m}(t\lambda_i + (1-\mathrm{t})\alpha_i) \nabla h_i(\overline{x}) + \sum_{j=1}^{q}\left(t\mu_j + (1-\mathrm{t})\beta_j\right) \nabla g_j(\overline{x})$$

Since $(t\lambda_i + (1-\mathrm{t})\alpha_i) \geq 0$ *and* $\left(t\mu_j + (1-\mathrm{t})\beta_j\right) \geq 0 \Longrightarrow tx_1 + (1-\mathrm{t})x_2 \geq 0$

$\Longrightarrow x_1, x_2 \in \mathrm{G}(\overline{x})$, and hence $\mathrm{G}(\overline{x})$, it is convex

To prove that $\mathrm{G}(\overline{x})$, it is closed, consider $(s^k) \subset \mathrm{G}(\overline{x})$, satisfying $s^k \longrightarrow s^* \in \mathbb{R}^n$

It has to be proved that $s^* \in \mathrm{G}(\overline{x})$, for suitable matrices B and C, we have

$\mathrm{G}(\overline{x}) = \{B\lambda + C\rho | \rho \geq 0\}$, by the Caratheodory's lemma we can assume that $D = (BC)$, it has linearly independent columns, so that $D^T D$, it is a non singular matrix

Since $(s^k) \subset \mathrm{G}(\overline{x})$, and then there exists $\gamma^k = \begin{pmatrix} \lambda^k \\ \rho^k \end{pmatrix}$ *with* $\rho^k \geq 0$, such that

$$s^k = D\gamma^k \longrightarrow (5)$$

Since $D^T D$, it is a non singular matrix $\Longrightarrow \gamma^k = (D^T D)^{-1} D^T s^k$, taking the limit $k \longrightarrow \infty$, we get

$$\begin{pmatrix} \lambda^* \\ \rho^* \end{pmatrix} = \gamma^* = \lim_{k \to \infty} \gamma^k = (D^T D)^{-1} D^T s^* \ with\ \rho^* \geq 0$$

Again, taking the limit $k \longrightarrow \infty$, in (5) we get $\lim_{k \to \infty} s^k = s^* = D\gamma^* \in \mathrm{G}(\overline{x})$,

And hence $\mathrm{G}(\overline{x})$, it is closed

Lemma – 5:

For any $\overline{x} \in \Omega$, and then $\mathrm{D}(\overline{x}) = P\left(\mathrm{G}(\overline{x})\right)$

Proof:

By the lemma – 1 and 4 we need to prove that $\mathrm{D}(\overline{x}) = P\left(\mathrm{G}(\overline{x})\right)$ *where* $\mathrm{D}(\overline{x})$ and $\mathrm{G}(\overline{x})$, they are defined in (3) and (4) respectively

Consider $d \in \mathrm{D}(\overline{x})$, and given $s \in \mathrm{G}(\overline{x})$, and then we have

$$d^T s = \sum_{i=1}^{m} d^T \lambda_i \nabla h_i(\overline{x}) + \sum_{j \in A(\overline{x})} \mu_j\, d^T \nabla g_j(\overline{x}) \longrightarrow (6)$$

By the definition of $\mathrm{D}(\overline{x})$ and since $\mu_j \geq 0 \Longrightarrow d^T s \leq 0 \Longrightarrow d \in P\left(\mathrm{G}(\overline{x})\right)$

Conversely, consider $d \in P\left(\mathrm{G}(\overline{x})\right)$ $i.e.\, d^T s \leq 0$ *for all* $s \in \mathrm{G}(\overline{x})$

In particular, since $\pm\nabla h_i(\overline{x}) \in \mathrm{G}(\overline{x})$, for all $\mathrm{i} = 1, .., \mathrm{m}$, we get $d^T \nabla h_i(\overline{x}) = 0$

Furthermore, since $\nabla g_j(\overline{x}) \in \mathrm{G}(\overline{x})$, for all $j \in A(\overline{x})$, we have $d^T \nabla g_j(\overline{x}) \leq 0$

And hence $d \in \mathrm{D}(\overline{x})$ and then $\mathrm{D}(\overline{x}) = P\big(\mathrm{G}(\overline{x})\big)$

The Tangent Cone:

Definition – 7:

A vector $d \in \mathbb{R}^n$, it is called tangent direction to $\Omega \subset \mathbb{R}^n$ $from$ $\overline{x} \in \Omega$, when either $d = 0$, or there exists a sequence of feasible points $(x^k) \subset \Omega$ $such\ that\ x^k \longrightarrow \overline{x}$, and also

$$\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} \longrightarrow \frac{d}{\|d\|}$$

Clearly, the set $\mathrm{T}(\overline{x})$ of the tangent directions to Ω $from$ $\overline{x}$, it is a cone. This set is said to be a tangent cone.

Example – 1:

Let $S = \{(x, y) \in \mathbb{R}^2 | x^2 - \mathrm{y} = 0\}$, find the tangent cone at $(0, 0)$

Solution:

Let $(x_k, y_k) \longrightarrow (0, 0), i.e. x_k \longrightarrow 0\ \ and\ y_k = x_k^2$

$$\Longrightarrow \|(x_k, y_k) - (0, 0)\| = \sqrt{x_k^2 + (y_k)^2} = |x_k|\sqrt{x_k^2 + 1}$$

$$\Longrightarrow \lim_{x_k \longrightarrow 0^+} \frac{x_k}{|x_k|\sqrt{x_k^2+1}} = 1\ \ and\ \lim_{x_k \longrightarrow 0^+} \frac{y_k}{|x_k|\sqrt{x_k^2+1}} = 0$$, and also

$$\lim_{x_k \longrightarrow 0^-} \frac{x_k}{|x_k|\sqrt{x_k^2+1}} = -1\ \ and\ \lim_{x_k \longrightarrow 0^-} \frac{y_k}{|x_k|\sqrt{x_k^2+1}} = 0$$

$$\Longrightarrow T(0, 0) = \{(-1, 0), (1, 0)\}$$

Lemma – 6:

For any $\overline{x} \in \Omega$, then $T(\overline{x})$, it is closed where $T(\overline{x})$ it is the tangent direction to Ω

Proof:

Consider $(d^k) \subset T(\overline{x})$ $with\ d^k \longrightarrow d$

To prove that $d \in T(\overline{x})$

When $d = 0$, we get $d \in T(\overline{x})$

Assume that $d \neq 0$, and supposes that with loss of generality that $d^k \neq 0, for\ all\ k \in \mathbb{N}$

Fixed $k \in \mathbb{N}$ $and\ since\ d^k \in T(\overline{x})$, and then there exists,

$$\left(x^{k,j}\right)_{j\in\mathbb{N}} \subset \Omega\ then x^{k,j} \xrightarrow{j} \overline{x}\ \ and\ q^{k,j} = \frac{x^{k,j} - \overline{x}}{\|x^{k,j} - \overline{x}\|} \xrightarrow{j} \frac{d^k}{\|d^k\|}$$

$\Longrightarrow$There exists $j_k \in \mathbb{N}$ such that $\|x^k - \overline{x}\| < \frac{1}{k}$ and $\left|q^k - \frac{d^k}{\|d^k\|}\right| < \frac{1}{k}$ $where$

$x^k = x^{k,j_k}$ $and\ q^{k,j_k} = q^k$, taking the limit $k \longrightarrow \infty$, we get $x^k \longrightarrow \overline{x}$, and also

$\left|q^k - \frac{d}{\|d\|}\right| \le \left|q^k - \frac{d^k}{\|d^k\|}\right| + \left|\frac{d^k}{\|d^k\|} - \frac{d}{\|d\|}\right| \longrightarrow 0$

$\Longrightarrow \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = q^k \longrightarrow \frac{d}{\|d\|} \Longrightarrow d \in T(\bar{x})$

$\Longrightarrow T(\bar{x})$, it is closed where $T(\bar{x})$ it is the tangent direction to Ω

Remark:

We have presented two different linear approximations of the feasible set at a point $\bar{x}$

a. The first order feasible variations cone $D(\bar{x})$, and

b. The tangent cone $T(\bar{x})$

Lemma – 7:

For any $\bar{x} \in \Omega$, and $T(\bar{x}) \subset D(\bar{x}), where\ T(\bar{x})\ and\ D(\bar{x})$, they are the tangent cone and the first order feasible variations cone at a point $\bar{x}$

Proof:

Consider $d \in T(\bar{x})\ and\ d \neq 0$, and then there exists a sequence $(x^k) \subset \Omega\ with\ x^k \longrightarrow \bar{x}$ such that $\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \longrightarrow \frac{d}{\|d\|}$, from the smoothness of $g\ and\ h$, it follows that

$h(x^k) = h(\bar{x}) + \nabla h_i(\bar{x})^T(x^k - \bar{x}) + O(\|x^k - \bar{x}\|)\ and$

$g(x^k) = g(\bar{x}) + \nabla g_j(\bar{x})^T(x^k - \bar{x}) + O(\|x^k - \bar{x}\|)$

Since $x^k, \bar{x} \in \Omega$ and we have for all $j \in A(\bar{x})$

$\nabla h_i(\bar{x})^T \left[\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right] + \frac{O(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} = 0$, and $\nabla g_j(\bar{x})^T \left[\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right] + \frac{O(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \le 0$

Taking the limit $k \longrightarrow \infty$, we get $\nabla h_i(\bar{x})^T \frac{d}{\|d\|} = 0\ and\ \nabla g_j(\bar{x})^T \frac{d}{\|d\|} \le 0, for\ all\ j \in A(\bar{x})$

$\Longrightarrow d \in D(\bar{x})$, and hence we get $T(\bar{x}) \subset D(\bar{x})$

Remark:

The converse of the lemma – 7 need not be true, see the following counter example

Counter Example:

Consider the functions $h(x): \mathbb{R}^2 \longrightarrow \mathbb{R};\ g(x): \mathbb{R}^2 \longrightarrow \mathbb{R}^2\ for\ j = 1, 2$ defined by

$h(x) = x_1 x_2$, and

$g(x) = -x_1 - x_2$, and the point $\bar{x} = (0, 0)^T$

$\Longrightarrow T(\bar{x}) = \{(d_1, d_2) \in \mathbb{R}^2 | d_1 \ge 0, d_2 \ge 0\ \ and\ d_1 d_2 = 0\}$, and also

$D(\bar{x}) = \{(d_1, d_2) \in \mathbb{R}^2 | -d_1 - d_2 \le 0\ \ and\ T(\bar{x}) \neq D(\bar{x})\}$, clearly $D(\bar{x}) \not\subset T(\bar{x})$

Note:

If $T(\bar{x}) = D(\bar{x})$, it is a constraint qualification known as "Quasi regularity" [2], [12]

2. **Optimality Conditions and Constraint Qualifications:**

In this topic we prove the KKT theorem assuming the weakest qualification condition and discuss other ones easier to be verified, supposes that the objective function increases along tangent direction we have the following lemma:

Lemma – 9:

If $x^* \in \Omega$, it is a local minimizer of the problem (P), and then $\nabla f(\overline{x})^T d \geq 0, for\ all\ d \in T(x^*)$

Proof:

This follows directly from the relation

$0 \leq f(x^k) - f(x^*) = \nabla f(\overline{x})^T (x^k - \overline{x}) + O(\|x^k - \overline{x}\|)$, it is valid for $(x^k) \subset \Omega$

Classical Karush – Kuhn – Tucker Theorem – 10:

Let $x^* \in \Omega$, it is a local minimizer of the problem (P) and if $P\big(T(x^*)\big) = P\big(D(x^*)\big)$, and then there exists $\lambda^* \in \mathbb{R}^m\ and\ \mu^* \in \mathbb{R}^p$ such that

$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0$

With $\mu_j^* \geq 0\ \ for\ j = 1, \dots, p\ \ and\ \mu_j^* \nabla g_j(x^*) = 0\ for\ j = 1, \dots, p$

Proof:

Consider $x^* \in \Omega$, it is a local minimizer of the problem (P), by the lemma – 9, we have $\nabla f(x^*)^T d \leq 0, for\ all\ d \in T(x^*)$, again using the lemma – 5 and given hypothesis, we get $-\nabla\ f(x^*) \in P\big(T(x^*)\big) = P\big(D(x^*)\big) = G(x^*)$

Define $\lambda^* = \lambda\ \ and\ \mu_j^* = \begin{cases} \mu_j, & for\ j \in A(x^*) \\ 0, & Otherwise \end{cases}$

This means that there exists $\lambda \in \mathbb{R}^m\ \ and\ \mu_j \geq 0\ \ for\ j \in A(x^*)$ such that

$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0$

Hence complete the proof

Note:

$P\big(T(x^*)\big)\ and\ P\big(D(x^*)\big)$, they are the Polar of the tangent cone and the first order feasible variations cone respectively

Constraint Qualifications:

The Kuhn – Tucker conditions are only if some regularity conditions are satisfied. These conditions are called the constraint qualification which imposes a certain restriction on constraint functions of a Nonlinear Programming problem [1], [11], for the specific purpose of ruling out certain irregularities on the boundary of the feasible set that would available KKT conditions should be the optimal solution occurs there.

Quasi – regularity constraint Qualification:

We say that the quasi – regularity constraint qualification is satisfied at $\overline{x}\, when\, T(\overline{x}) = D(\overline{x})$

Where $T(\overline{x})\ and\ D(\overline{x})$, they are the tangent cone and the set of first order feasible variations cone at $\overline{x}$

Note:

These conditions are not equivalent [4], for example

Consider the functions $h(x): \mathbb{R}^2 \longrightarrow \mathbb{R};\ g(x): \mathbb{R}^2 \longrightarrow \mathbb{R}\ for\ j = 1, 2$ defined by

$h(x) = x_1 x_2$, and

$g(x) = -x_1 - x_2$, and the feasible point $\overline{x} = (0, 0)^T$

It is easy to see that

$T(\overline{x}) = \{(d_1, d_2) \in \mathbb{R}^2 | d_1 \geq 0, d_2 \geq 0\ \ and\ d_1 d_2 = 0\}$, and also

$D(\overline{x}) = \{(d_1, d_2) \in \mathbb{R}^2 | d_1 \geq 0, d_2 \geq 0\}$,

And also $P\big(D(\overline{x})\big) = P\big(T(\overline{x})\big) = \{(d_1, d_2) \in \mathbb{R}^2 | d_1 \leq 0, d_2 \leq 0\}$

Slater Constraint Qualification:

Regarding the problem (P), we say that the Slater constraint qualification holds if h, it is linear and g, it is convex and then there exists $\tilde{x} \in \Omega$, such that $h(\tilde{x}) = 0\ \ and\ g(\tilde{x}) < 0$

The Slater constraint is, in fact, a constraint qualification [5]

Theorem – 11:

If the Slater conditions hold and then $T(\tilde{x}) = D(\tilde{x}), for\ all\ \tilde{x} \in \Omega$

Proof:

Using lemma – 7, it is enough to prove that $D(\tilde{x}) \subset T(\tilde{x})$

Consider an arbitrary direction $d \in D(\overline{x})\ and\ \ \tilde{x} \in \Omega$ it is given by Slater condition

Define: $\overline{d} = \tilde{x} - \overline{x}$ by the convexity of g_j, we have

$0 > g_j(\tilde{x}) \geq g_j(\overline{x}) + \nabla g_j(\overline{x})^T \overline{d} \Longrightarrow For\ j \in A(\overline{x})$, we have $\nabla g_j(\overline{x})^T \overline{d} < 0$, given $\lambda \in (0, 1]$

Define: $\hat{d} = (1 - \lambda)d + \lambda \overline{d}$

To prove that $\hat{d} \in T(\tilde{x}), for\ all\ \lambda \in (0, 1]$

For $j \in A(\overline{x})$, we have$\nabla g_j(\overline{x})^T \overline{d} < 0\ and\ \nabla g_j(\overline{x})^T d < 0$, consequently we have

$\nabla g_j(\overline{x})^T \hat{d} < 0 \Longrightarrow$There exists $\hat{x} = \overline{x} + t\hat{d}\ \ with\ t > 0$ such that $g_j(\hat{x}) < g_j(\overline{x})$

Taking a sequence (t_k) and $t_k \longrightarrow 0$

Define: $x^k = (1-t_k)\overline{x} + t_k\hat{x} = \overline{x} + t_k t\hat{d}$

$\Longrightarrow \frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} = \frac{t_k t\hat{d}}{\|x^k - \overline{x}\|} = \frac{\hat{d}}{\|\hat{d}\|}$ $for\ t \notin A(\overline{x})$ $and\ g_j(\overline{x}) < 0$ $(because\ it\ is\ slater)$

By the continuity of g, we have $g(x^k) < 0, for\ all\ k$, it is sufficiently large.

To conclude that $\hat{d} \in T(\tilde{x})$, it is enough to show that $h(x^k) = 0\ for\ all\ k$, it is sufficiently large

Since $d \in D(\overline{x}), and\ Md = \nabla h(x)^T d = 0$

Furthermore, $M\overline{d} = M(\tilde{x} - \overline{x}) = 0$ consequently $M\hat{d} = 0$

$\Longrightarrow h(x^k) = Mx^k - c = M\overline{x} - c + t_k tM\hat{d} = 0, since\ \overline{x} \in \Omega \Longrightarrow \hat{d} \in T(\overline{x}) \Longrightarrow d \in T(\overline{x})$, since $T(\overline{x})$, it is closed and hence completes the proof

Linear Independence Constraint Qualification – LICQ:

This is the most known constraint qualification and states that the equality constraints gradients

$\nabla h_i(\overline{x})\ for\ i = 1, \dots, m$, and the active inequality constraint gradients $\nabla g_j(\overline{x}), for\ j \in A(\overline{x})$, they are linearly independent. Although easy to check, this condition is a very strong assumption

For example, consider

Min $\qquad f(x) = (x_1-3)^2 + (x_2-2)^2$

Subject to $\qquad g_1(x) = 2x_1 + x_2 - 6 \leq 0$

$\qquad g_2(x) = x_1 + 2x_2 - 6 \leq 0$

In this problem we have $\nabla f(x) = [2(x_1-3), 2(x_2-2)]$; $\nabla g_1(x) = [2,1]$, $\nabla g_1(x) = [1,2]$

$\Longrightarrow$The gradients of g, they are linearly independent so all points are regular.

We consider the following cases:

Case – I:

For $A(\overline{x}) = \emptyset$

From KKT conditions we get both $\lambda_1 = 0\ and\ \lambda_2 = 0\ then$

$\nabla f(x) = [2(x_1-3), 2(x_2-2)] = 0 \Longrightarrow x_1 = 3\ \ and\ x_2 = 2$

But $2x_1 + x_2 - 6 = 2 \nleq 0, i.e.\ x$, it is not feasible it cannot be a local minimum

Case – II:

For $A(\overline{x}) = \{1\}$

From KKT conditions we get $\lambda_2 = 0\ then$

$\nabla f(x) + \lambda_1 \nabla g_1(x) = [2(x_1 - 3), 2(x_2 - 2)] + \lambda_1[2, 1] = 0$

$\Longrightarrow x_1 = 3 - \lambda_1$ and $x_2 = 2 - \frac{\lambda_1}{2}$, and the assumption

$g_1(x) = 0$, it gives with theses x_1 and x_2, as follows

$2(3 - \lambda_1) + \left(2 - \frac{\lambda_1}{2}\right) - 6 = 0 \Longrightarrow \lambda_1 = \frac{4}{5}$

$\Longrightarrow \lambda_1$ and λ_2 Satisfies KKT condition and then we get

$x_1 = 3 - \frac{4}{5} = \frac{11}{5}$ and $x_2 = 2 - \frac{1}{2}\left(\frac{4}{5}\right) = \frac{8}{5}$

Finally $g_2(x) = \frac{11}{5} + 2\left(\frac{8}{5}\right) - 6 = -\frac{3}{5} \le 0$, and hence KKT conditions are satisfied

Case – III:

For $A(\overline{x}) = \{2\}$

From KKT conditions we get $\lambda_2 = 0$ $then$

$\nabla f(x) + \lambda_2 \nabla g_2(x) = [2(x_1 - 3), 2(x_2 - 2)] + \lambda_2[1, 2] = 0$

$\Longrightarrow x_1 = 3 - \frac{\lambda_2}{2}$ and $x_2 = 2 - \lambda_2$, and the assumptions

$g_2(x) = 0$, it gives with theses x_1 and x_2, as follows

$\left(3 - \frac{\lambda_2}{2}\right) + 2(2 - \lambda_2) - 6 = 0 \Longrightarrow \lambda_2 = \frac{2}{5}$

$\Longrightarrow \lambda_1$ and λ_2 Satisfies KKT condition and then we get

$x_1 = 3 - \frac{1}{5} = \frac{14}{5}$ and $x_2 = 2 - \left(\frac{2}{5}\right) = \frac{8}{5}$

But $g_1(x) = 2\left(\frac{14}{5}\right) + \frac{8}{5} - 6 = \frac{1}{5} > 0$, this is violated the condition of $g_1(x) \le 0$

Case – IV:

For $A(\overline{x}) = \{1, 2\}$

Finally $g_1(x) = 2x_1 + x_2 - 6$ and $g_2(x) = x_1 + 2x_2 - 6$, gives that $x_1 = 2 = x_2$

But KKT condition $\nabla f(x) + \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) = 0$

$\Longrightarrow [2(x_1 - 3), 2(x_2 - 2)] + \lambda_1[2, 1] + \lambda_2[1, 2] = 0$, gives with theses x_1 and x_2, as follows

$-2 + 2\lambda_1 + \lambda_2 = 0$ and $\lambda_1 + 2\lambda_2 = 0 \Longrightarrow \lambda_1 = \frac{4}{3}$ and $\lambda_2 = -\frac{2}{3}$, and since $\lambda_1 > 0$ and $\lambda_2 < 0$

It does not satisfies KKT condition

Note:

Many problems satisfy KKT conditions without LICQ, for example with $x^* = 0$

Consider

Min $f(x) = x_2$

Subject to $g_1(x) = x_1^2 + x_2 \leq 0$

$g_2(x) = -x_2 \leq 0$

Clearly it satisfies KKT conditions without LICQ

Mangasarian – Fromovitz's Constraint Qualification – MFCQ:

Another well known condition which ensures KKT is due to Mangasarian – Fromovitz's. We say that MFCQ holds at $\bar{x}$, when the equality constraint gradients are linearly independent and there exists a vector $d \in \mathbb{R}^n$ such that

$\nabla h_i(\bar{x})^T d = 0 \; for \; i = 1, \dots n \;\; and \; \nabla g_j(\bar{x})^T d \leq 0 \; for \; all \; j \in A(\bar{x})$

The best known necessary optimality criterion for a mathematical programming problem is the KKT optimality conditions [4], [10]; however, the above MFCQ condition is in a sense more general. In order for the KKT conditions to hold, one must impose a constraint qualification on the constraints of the problem. On the other hand, no such qualification need be imposed on the constraints in order that the MFCQ to hold. Moreover, the MFCQ itself can be used to drive a form of the constraint qualification for KKT conditions

Relation among Constraint Qualification:

Relation between LICQ and MFCQ:

Theorem – 12:

If $\bar{x} \in \Omega$ satisfies LICQ and then it satisfies MFCQ

Proof:

Supposes that without loss of generality that $A(\bar{x}) = \{1, \dots, q\}$, and consider the matrix

$M = [\nabla h_1(\bar{x}) \quad \cdots \quad \nabla h_m(\bar{x}) \quad \nabla g_1(\bar{x}) \quad \cdots \quad \nabla g_q(\bar{x})]^T \;\; and \; b \in \mathbb{R}^{m+q} \; given \; by$

For all $b_i = 0 \; for \; all \; i = 1, \dots, m \;\; and \; b_j = -1 \; for \; all \; j \in \{m+1, \dots, m+q\}$

Since the rows of M , they are linearly independent, the system $Md = b$ it has a solution

Let $\bar{d}$, it is a solution and then we have

$\nabla h(\bar{x})^T \bar{d} = 0 \;\; and \; \nabla g_j(\bar{x})^T \bar{d} = -1 < 0, \;\; for \; all \; j \; \in A(\bar{x})$

Hence, if $\bar{x} \in \Omega$ satisfies LICQ and then it satisfies MFCQ

Note:

The converse of theorem – 12 is need not be true, as shown a counter example below

Counter Example:

Consider the function $g_j(x): \mathbb{R}^2 \longrightarrow \mathbb{R} \;\; for \; j = 1, 2, 3$, defined by

$g_1(x) = (x_1 - 1)^2 + (x_2 - 1)^2 - 2$

$g_2(x) = (x_1 - 1)^2 + (x_2 + 1)^2 - 2$

$g_3(x) = -x_1$, with the feasible point $\overline{x} = (0,0)^T$

Note that $g_j(x): \mathbb{R}^2 \longrightarrow \mathbb{R}\ \ for\ j = 1,2,3$, it is linearly dependent

On the other hand, taking $d = (1,0)^T$, we have $\nabla g_j(\overline{x})^T d < 0\ for\ j = 1,2,3$

It means that MFCQ holds but LICQ does not hold

Relation between MFCQ and Quasi – Regularity:

Lemma – 13:

Let $\gamma: (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}^n$, it is a differentiable curve such that $h\big(\gamma(t)\big) = 0, for\ all\ t \in (-\varepsilon, \varepsilon)$

If $\gamma(0) = \overline{x}, and\ \gamma'(0) = d \neq 0$, and then there exists a sequence $(x^k)\ with$

$h(x^k) = 0\ and\ x^k \longrightarrow \overline{x}$, and also

$\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} \longrightarrow \frac{d}{\|d\|}$

Proof:

We have $\lim\limits_{t \to 0} \frac{\gamma(t) - \overline{x}}{t} = \lim\limits_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t} = \gamma'(0) = d \neq 0 \Longrightarrow \gamma(t) \neq \overline{x}\ \ for\ all\ t \neq 0$ sufficiently small, and taking a sequence $(t_k)\ and\ t_k \longrightarrow 0$

Define: $x^k = \gamma(t_k)$ thus $\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} = \frac{x^k - \overline{x}}{t_k} . \frac{t_k}{\|x^k - \overline{x}\|} \longrightarrow \frac{d}{\|d\|}$

Theorem – 14:

If $\overline{x} \in \Omega$ satisfies MFCQ and then $T(\overline{x}) = D(\overline{x})$

Proof:

Consider an arbitrary diction $d \in D(\overline{x})\ \ and\ \overline{d}$, it is given by MFCQ and $\lambda \in (0,1]$

Define: $\hat{d} = (1 - \lambda)d + \lambda\overline{d}$

To prove that $\hat{d} \in T(\overline{x}), for\ all\ \lambda \in (0,1]$

Denote $M = \nabla h(\overline{x})^T$, By MFCQ we have $rank\ (M) = m$

Consider the matrix $Z = [v^1 \quad \cdots \quad v^{n-m}] \in \mathbb{R}^{n \times (n-m)}$, whose columns are bases of $\mathcal{N}(M)$

Since $\{\nabla h_1(\overline{x}), \dots, \nabla h_m(\overline{x})\}$, it is a basis of $Im(M^T)$, and the matrix $\begin{bmatrix} M \\ Z^T \end{bmatrix}$, it is non singular

Define: $\varphi: \mathbb{R}^{(n+1)} \longrightarrow \mathbb{R}^n\ \ by\ \varphi \begin{bmatrix} x \\ t \end{bmatrix} = \begin{bmatrix} h(x) \\ Z^T\big(x - \overline{x} - t\hat{d}\big) \end{bmatrix}$

Since $\nabla_x \varphi^T = \begin{bmatrix} M \\ Z^T \end{bmatrix}$, it is non singular, by the implicit function theorem, there exists a differentiable curve $\gamma: (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}^n\ such\ that\ \varphi \begin{bmatrix} \gamma(t) \\ t \end{bmatrix} = 0, for\ all\ t \in (-\varepsilon, \varepsilon)$

$\Longrightarrow h\big(\gamma(t)\big) = 0 \;\; and \; Z^T\big(\gamma(t) - \overline{x} - t\hat{d}\big) = 0 \longrightarrow$ (7)

Since $\varphi \begin{bmatrix} \overline{x} \\ 0 \end{bmatrix} = 0$, by the unicity of γ, and then we have $\gamma(0) = \overline{x}$

Taking the derivative at $t = 0$, in the equation (7), we get

$M\gamma^{'}(0) = 0 \longrightarrow$ (8)

Using (7), for $t \neq 0$, we get

$Z^T\left(\frac{\gamma(t)-\overline{x}}{t} - \hat{d}\right) = 0 \longrightarrow$ (8)

Taking the limit as $t \longrightarrow 0$, we get

$\lim_{t \longrightarrow 0} Z^T\left(\frac{\gamma(t)-\overline{x}}{t} - \hat{d}\right) = 0 \Longrightarrow Z^T\gamma^{'}(0) = Z^T\hat{d} \longrightarrow$ (9)

As $d, \overline{d} \in D(\overline{x}) \;\; and \; then \; M\hat{d} = 0$, using (8) and (9), we get

$\begin{bmatrix} M \\ Z^T \end{bmatrix}\gamma^{'}(0) = \begin{bmatrix} M \\ Z^T \end{bmatrix}\hat{d} \Longrightarrow \hat{d} = \gamma^{'}(0)$

By lemma – 13, there exists a sequence $(x^k) \; with \; h(x^k) = 0 \;\; and \; x^k \longrightarrow \overline{x}$, and also

$\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} \longrightarrow \frac{\hat{d}}{\|\hat{d}\|}$

To conclude that $\hat{d} \in T(\overline{x})$, it is enough to show that $g(x^k) \leq 0 \;\; for \; all \; k$ it is sufficiently large

For $j \notin A(\overline{x})$, we have $g(\overline{x}) < 0$, by the continuity of g, we have

$g(x^k) \leq 0 \;\; for \; all \; k$, it is sufficiently large

For $j \in A(\overline{x})$, we have $\nabla g_j(\overline{x})^T d \leq 0 \;\; and \;\; \nabla g_j(\overline{x})^T \overline{d} < 0$

Consequently $\nabla g_j(\overline{x})^T \hat{d} < 0$

From the smooth of g_j, it follows that

$g_j(x^k) = g_j(\overline{x}) + \nabla g_j(\overline{x})^T(x^k - \overline{x}) + O(\|x^k - \overline{x}\|)$

$\Longrightarrow \frac{g_j(x^k)}{\|x^k - \overline{x}\|} = \nabla g_j(\overline{x})^T\left(\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|}\right) + \frac{O(\|x^k - \overline{x}\|)}{\|x^k - \overline{x}\|} \longrightarrow \nabla g_j(\overline{x})^T \frac{\hat{d}}{\|\hat{d}\|} < 0$

$\Longrightarrow g_j(x^k) < 0, for \; all \; k$, it is sufficiently large

$\Longrightarrow \hat{d} \in T(\overline{x}) \Longrightarrow d \in T(\overline{x})$, since $T(\overline{x})$, it is a closed set

$\Longrightarrow$If $\overline{x} \in \Omega$ satisfies MFCQ and then $T(\overline{x}) = D(\overline{x})$, and hence completes the proof

Note:

The Quasi – Regularity it does not MFCQ and also not LICQ, see the following counter example

Counter Example:

Consider the functions $g_j(x): \mathbb{R}^2 \longrightarrow \mathbb{R} \;\; for\ j = 1, 2$ defined by

$g_1(x) = -x_1^2 + x_2$

$g_2(x) = -x_1^2 - x_2$, with the feasible point $\overline{x} = (0, 0)^T$

In this case $D(\overline{x}) = \{(d_1, 0) | d_1 \in \mathbb{R}\}$, for obtaining $T(\overline{x})$, consider the sequence

$(x^k) = (t_k, 0)\ with\ t_k \longrightarrow 0\ and\ t_k > 0 \Longrightarrow x^k \longrightarrow \overline{x}\ \ and\ also\ \frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} = \frac{(t_k, 0)}{t_k} = (1, 0)$

$\Longrightarrow d = (1, 0)$, it is a tangent direction

In the same way for $t_k < 0$, we get $d = (-1, 0)$, it is also a tangent direction

Since $T(\overline{x})$it is a cone, we have $(d_1, 0) \in T(\overline{x})\ for\ all\ d_1 \in \mathbb{R}$

$\Longrightarrow D(\overline{x}) \subset T(\overline{x})$, by the lemma – 7 we conclude that $D(\overline{x}) = T(\overline{x})$

Note that there is no $d \in \mathbb{R}^2$such that $\nabla g_j(\overline{x})^T d < 0\ for\ j = 1, 2$

Furthermore $\{\nabla g_1(\overline{x}), \nabla g_2(\overline{x})\}$, it is linearly dependent

$\Longrightarrow \overline{x}$, it does not satisfy either MFCQ or LICQ, and hence completes the result

Applications for Optimization of NLP:

A typical non convex problem is that of optimizing transportation costs by selection from a set of transportation methods, one or more of which exhibit economics of scale, with various connectivities and capacity constraints. An example would be petroleum product transport given a selection or combination of pipeline, rail tanker, road tanker, river barge or coastal tank ship. Owing to economic batch size the cost functions may have discontinuities in addition to smooth changes.

Modern engineering practice involves much numerical optimization. Except in certain narrow but important cases such as passive electronic circuits, engineering problems are non – linear and they are usually very complicated.

In experimental science, some simple data analysis (Such as fitting a spectrum with a sum of peaks of know location and shape but unknown magnitude) can be done with linear methods, but in general these problems, also, are non – linear. Typically, one has a theoretical model of the system under study with variable parameters in it and a model the experiment or experiments which may also have unknown parameters. One tries to find a best fit numerically. In this case one often wants a measure of the precision of the result, as well as the best fit itself. Often in mathematical economics the KKT approach is used in theoretical models in order to obtain qualitative results.

For example in Marginal Cost – Marginal Revenue:

Selective Example:

Consider a firm that maximizes its sales revenue subject to a minimum profit constraint. Letting$Q$, it is the quantity of output product (To be chosen); $R(Q)$ it is sales revenue with a positive first derivative and with a zero value at zero output. $C(Q)$, it is production costs with a positive first derivative and with a nonnegative value at zero output and G_{min}, it is the positive minimal acceptable level of profit, then the problem is a meaningful one if the revenue function levels off so it eventually is less steep than the cost function. The problem expressed in the previously given minimization form is

Minimize $R(Q)$

Subject to $G_{min} \leq R(Q) - C(Q)$, and

$Q \geq 0$

And the KKT conditions are:

- $\left(\frac{dR}{dQ}\right)(1+\mu) - \mu\left(\frac{dC}{dQ}\right) = 0$
- $Q \geq 0$
- $Q\left[\left(\frac{dR}{dQ}\right)(1+\mu) - \mu\left(\frac{dC}{dQ}\right)\right] = 0$
- $\mu \geq 0$
- $\mu[R(Q) - C(Q) - G_{min}] = 0$

Since $Q = 0$, it would violate the minimum profit constraint, we have $Q > 0$, and hence the third condition implies that the first condition holds with equality. Solving that equality gives

$$\frac{dR}{dQ} = \frac{\mu}{1+\mu}\left(\frac{dC}{dQ}\right)$$

Because, it was given that $\frac{dR}{dQ}$ *and* $\frac{dC}{dQ}$, they are strictly positive, this inequality along with the Non negativity condition on μ guarantees that μ it is positive and so the revenue – maximizing firm operates at a level of output at with marginal revenue $\frac{dR}{dQ}$, it is less than marginal cost $\frac{dC}{dQ}$, a result that is of interest because it contrasts with the behavior of a profit maximizing firm, which operates at a level at which they are equal

CONCLUSION:

In this paper we observe that, KKT optimality condition for NLP has been proved assuming the equality of the polar of the tangent cone and the polar of the first order feasible variations cone. Despite the difficulty of this property, it needs to have more readily verifiable

conditions for the admittance of Lagrange Multipliers. Such conditions called constraint qualifications have been investigated extensively. Some of them were discussed as: Slater, Linear Independence Gradients (LICQ), Mangasarian – Fromovitz's (MFCQ), and Quasi – Regularity condition through which their relations are analyzed and can be summarized as LICQ implies MFCQ but the converse is not necessarily true. MFCQ implies Quasi – Regularity but the converse do not hold and that Slater condition satisfies Quasi – Regularity.
There are other constraint qualification is not discussed in this paper; such as Quasi – Normality condition which implies quasi – Regularity, the constant positive linear dependence (CPLD) which is weaker than MFCQ and implies quasi – Normality and constant rank constraint qualification (CRCQ) which shows the constraint positive linear dependence (CPLD) these are all our future aim to complete.

REFERENCES:

1. Abadie, J., 1967. On the Kuhn – Tucker Theorem
   "*Non Linear Programming pp. 19 – 36*" North Holland, Amsterdam
2. Andereani, R., J.M. Martinez and M.L. Schuverdt, 2005. On the relation between constant positive linear dependence condition quasi – normality constraint qualification.
   "*Journal of Optimization theory and Applications, 125: 473 – 485*"
3. Bertsekas, D.P., 1995. Nonlinear Athena Scientific, Belmont, Massachusetts. Boyd, S. and V. Levin, 2004. "*Convex Optimization*", Cambridge: Cambridge University Press , pp 244
4. Flegel, M.L., 2005. "*Constraint Qualification and Stationary Concepts for Mathematical Programs with Equilibrium Constraints*". PhD thesis, Institute of Applied Mathematics and Statistics, University of Wurzburg, Alemanha
5. Guignard, M., 1967. "Generalized Kuhn – Tucker Conditions for mathematical Programming Problems in a Banach Space" *SIAM Journal Optimal Control, 7: pp 232 – 241*
6. Hestenes, M.R., 1975. "Optimization Theory. The finite Dimensional Case". John Wiley, New York
7. Karush, W., 1939. "Minima of Functions of Several Variables with Inequalities as Side Constraints". M.Sc. Dissertation. Dept. of Mathematics, University of Chicago, Chicago, Illinois

8. Kjeldsen, T.H., 2000. "A Contextualized Historical Analysis of the Kuhn – Tucker Theorem in Nonlinear Programming" *The impact of World War II Historia Math, 4: pp 331 – 361*.
9. Kuhn, H. W. and A.W. Tucker, 1951. "Nonlinear programming" *Proceedings of 2nd Berkeley Symposium. Berkeley:* University of California Press. Pp. 481 – 492
10. Qi, Land Z., We, 2000. "On the constant Positive Linear Dependence Condition and its application to *sqp* methods *SIAM Journal On Optimization, 10: pp 963 – 981*
11. Rodrigo, G.E., W.K. Elizabeth and A.R., Admir. Karush – Kuhn – Tucker Optimality conditions and constraint qualifications. Department of Mathematics, Federal University of Parana
12. Tolle, J.W. and F.J. Gould, 1971. "A necessary and sufficient qualification for constrained Optimization. *SIAM Journal On Applied Mathematics, 20: pp 164 – 172*