Optimality conditions for continuous nonlinear optimization

Neculai Andrei
Research Institute for Informatics,
Center for Advanced Modeling and Optimization
8-10 Averescu Avenue, Bucharest 1, Romania
E-mail: nandrei@ici.ro

January 12, 2016

Abstract. The optimality conditions for the general nonlinear optimization problems are presented. The general concepts in nonlinear optimization, the optimality conditions for unconstrained optimization, the optimality conditions for problems with inequality constraints, the optimality conditions for problems with equality constraints, the optimality conditions for general problems, as well as some notes and references are detailed. The optimality conditions are introduced using the formalism of Lagrange. The sensitivity and interpretation of the Lagrange multipliers for nonlinear optimization problems are discussed. In Appendix we present some theoretical results which are used in proving some theorems concerning the optimality conditions for the general nonlinear optimization.

1. Preliminaries

The optimization problems considered in this book involve minimization or maximization of a function of several real variables subject to one or more constraints. The constraints may be non-negativity of variables, simple bounds on variables, equalities or inequalities as functions of these variables. These problems are known as continuous nonlinear constrained optimization or nonlinear programming.

The purpose of this chapter is to introduce the main concepts and the fundamental results in nonlinear optimization known as optimality conditions. Plenty of very good books dedicated to these problems are known in literature: [Luenberger, 1973], [Gill, Murray and Wright, 1981], [Peressini, Sullivan and Uhl, 1988], [Bazaraa, Sheraly and Shetty, 1993], [Bertsekas, 1999], [Boyd and Vandenberghe, 2006], [Nocedal and Wright, 2006], [Sun and Yuan, 2006], [Chachuat, 2007], [Andrei, 2009, 2015], etc.

The general continuous nonlinear optimization problem is expressed as:

$$\min f(x)$$

subject to: (1)

$$c_i(x) \le 0, \quad i = 1,\dots,m,$$
$$h_i(x) = 0, \quad i = 1,\dots,p,$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,m,$ and $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,p,$ are continuous differentiable functions.

Usually, the function f is called the *objective function*. Each of the constraints $c_i(x) \le 0,$ $i = 1,\dots,m,$ is *inequality constraints* and $h_i(x) = 0,$ $i = 1,\dots,p,$ is *equality constraints*.

Often (1) is called a *nonlinear program*. A vector x satisfying all the equality and inequality constraints is called a *feasible solution* (*point*) to the problem (1). Define

$$X = \{x : c_i(x) \le 0, i = 1,\dots,m, h_i(x) = 0, i = 1,\dots,p\}$$

as the *feasible region* (or *feasible domain*).

In this chapter we are interested to specify what is meant by optimality for the general nonlinear optimization problem and give conditions under which a solution for the problem (1) exists. Both necessary and sufficient conditions for optimality are presented, starting with unconstrained problems and continuing with problems with inequality constraints, equality constraints and finally for general nonlinear optimization problems with equality and inequality constraints. The key to understanding the nonlinear optimization is the Karush-Kuhn-Tucker (KKT) optimality conditions. This is a major result which identifies an algebraic system of equations and inequalities which corresponds to the solution to any nonlinear optimization problem. This system often can be used to develop algorithms for computing a solution for the problem or can also be considered to get some additional information about the sensitivity of the minimum value of the problem subject to changes in the constraints. In general, many optimization algorithms can be interpreted as methods for numerically solving the KKT nonlinear system of equations.

In mathematical optimization, the KKT conditions are first-order necessary conditions for a solution in nonlinear optimization to be optimal, provided that some regularity conditions are satisfied. For problems with inequality constraints, the KKT approach generalizes the method of Lagrange multipliers, which allows only equality constraints. For the development of the KKT optimality conditions three possible approaches can be used. One is based on the *separation and support theorems* from convex set theory. Another one uses *penalty functions* and the third one comes from the *theory of Lagrange multipliers*. Each of these approaches has its own virtues and provides its own insights on the KKT Theorem. In this text we consider the optimality conditions for continuously nonlinear optimization (mathematical programming) using the formalism of Lagrange.

2. General concepts in nonlinear optimization

In the following we shall present some definitions and results used in the context of nonlinear programming. At the same time we shall define a particular class of nonlinear programs that is convex programming. In this section $X \subset \mathbb{R}$ denotes a nonempty set of real numbers.

Definition 1 *(Upper bound, Lower bound). A real number* α *is called an upper bound for* X *if* $x \le \alpha$ *for all* $x \in X$. *The set* X *is said to be bounded above if it has an upper bound. Similarly, a real number* α *is called a lower bound for* X *if* $x \ge \alpha$ *for all* $x \in X$. *The set* X *is said to be bounded below if it has a lower bound.*

Definition 2 *(Least upper bound, Greatest lower bound). A real number* α *is called the least upper bound (or supremum, or* sup*) of* X, *if (i)* α *is an upper bound for* X; *and (ii) there does not exists an upper bound for* X *that is strictly smaller than* α. *The supremum, if it exists, is unique and is denoted by* $\sup X$. *A real number* α *is called the greatest lower bound (or infimum, or* inf*) of* X, *if (i)* α *is a lower bound for* X; *and (ii) there does not exists a lower bound for* X *that is strictly greater than* α. *The infimum, if it exists, is unique and is denoted by* $\inf X$.

It is worth saying that for sups and infs, the following equivalent definition is useful.

Definition 3 *(Supremum, Infimum). The supremum of* X, *provided it exists, is the least upper bound for* X, *i.e. a real number* α *satisfying: (i)* $z \le \alpha$ *for any* $z \in X$; *(ii) for any* $\bar{\alpha} < \alpha$, *there exists* $z \in X$ *such that* $z > \bar{\alpha}$. *Similarly, the infimum of* X, *provided it exists, is the greatest lower bound for* X, *i.e. a real number* α *satisfying: (i)* $z \ge \alpha$ *for any* $z \in X$; *(ii) for any* $\bar{\alpha} > \alpha$, *there exists* $z \in X$ *such that* $z < \bar{\alpha}$.

Definition 4 *(Maximum, Minimum). The maximum of a set* X *is its largest element if such an element exists. The minimum of a set* X *is its smallest element if such an element exists.*

The key differences between the *supremum* and *maximum* concepts are as follows. If a set has a maximum, then the maximum is also a supremum for this set, but the converse is not true. A finite set always has a maximum which is also its supremum, but an infinite set need not have a maximum. The supremum of a set X need not be an element of the set X itself, but the maximum of X must always be an element of X.

Concerning the existence of infima and suprema in $\mathbb{R}$, fundamental is the *axiom of completeness*. „*If a nonempty subset of real numbers has an upper bound, then it has a least upper bound. If a nonempty set of real numbers has a lower bound, it has a greatest lower bound*". In other words, the completeness axiom guarantees that, for any nonempty set of real numbers that is bounded above, a supremum exists (in contrast to the maximum, which may or may not exist).

Let us consider the minimization problem

$$\min\{f(x) : x \in X\}, \tag{2}$$

where $X \subset \mathbb{R}^n$ represents the *feasible set*. Any point $x \in X$ is a *feasible point* or an *admissible point*. Any point $x \in \mathbb{R}^n \setminus X$ is called to be *infeasible*.

Definition 5 *(Global Minimum, Strict Global Minimum). A point $x^* \in X$ is said to be a global minimum of f on X if $f(x) \geq f(x^*)$ for any $x \in X$. A point $x^* \in X$ is said to be a strict global minimum of f on X if $f(x) > f(x^*)$ for any $x \in X$ with $x \neq x^*$.*

Definition 6. *(Global Maximum, Strict Global Maximum). A point $x^* \in X$ is said to be a global maximum of f on X if $f(x) \leq f(x^*)$ for any $x \in X$. It is a strict global maximum of f on X if $f(x) < f(x^*)$ for any $x \in X$ with $x \neq x^*$.*

The point x^* is called an *optimal solution* of the optimization problem. The real number $f(x^*)$ is known as the *optimal value* of the objective function subject to the constraints $x \in X$.

Observe the distinction between the minimum/maximum and infimum/supremum. The value $\min\{f(x) : x \in X\}$ must be attained at one or more points $x \in X$. On the other hand, the value $\inf\{f(x) : x \in X\}$ does not necessarily have to be attained at any points $x \in X$. However, if a minimum (maximum) exists, then its optimal value is equal the imfimum (supremum).

It is worth saying that if a minimum exists, it is not necessarily unique. That is, there may be a finite number, or even an infinite number, of feasible points x^* that satisfy the inequality $f(x) \geq f(x^*)$ for any $x \in X$. The notation:

$$\arg\min\{f(x) : x \in X\} \triangleq \{x \in X : f(x) = \inf\{f(x) : x \in X\}\}$$

is reserved for the set of minima of function f on X, that is a set in $\mathbb{R}^n$.

Definition 7 *(Local Minimum, Strict Local Minimum). A point* $x^* \in X$ *is said to be a local minimum of* f *on* X *if there exists* $\varepsilon > 0$ *such that* $f(x) \geq f(x^*)$ *for any* $x \in B(x^*, \varepsilon) \cap X$, *where* $B(x^*, \varepsilon)$ *is the open ball centered at* x^* *of radius* ε. *Similarly, a point* $x^* \in X$ *is said to be a strict local minimum of* f *on* X *if there exists* $\varepsilon > 0$ *such that* $f(x) > f(x^*)$ *for any* $x \in B(x^*, \varepsilon) \setminus \{x^*\} \cap X$.

Definition 8. *(Local Maximum, Strict Local Maximum). A point* $x^* \in X$ *is said to be a local maximum of* f *on* X *if there exists* $\varepsilon > 0$ *such that* $f(x) \leq f(x^*)$ *for any* $x \in B(x^*, \varepsilon) \cap X$. *Similarly, it is said to be a strict local maximum of* f *on* X *if there exists* $\varepsilon > 0$ *such that* $f(x) < f(x^*)$ *for any* $x \in B(x^*, \varepsilon) \setminus \{x^*\} \cap X$.

A fundamental problem in optimizing a function on a given set is whether a minimum or a maximum point exists in the given set. This result is known as the theorem of Weierstrass. It shows that if X is nonempty, closed and bounded and f is continuous on X, then a minimum of f on X exists.

Theorem 1 *(Weierstrass). Let* X *be a nonempty and compact set. Assume that* $f : X \to \mathbb{R}$ *is continuous on* X. *Then, the problem* $\min\{f(x) : x \in X\}$ *attains its minimum.*

Proof. If f is continuous on X and X is both closed and bounded, it follows that f is bounded below on X. Now, since X is nonempty, from the axiom of completeness there exists a greatest lower bound $\alpha = \inf\{f(x) : x \in X\}$. Let $0 < \varepsilon < 1$ and consider the set $X_k = \{x \in X : \alpha \leq f(x) \leq \alpha + \varepsilon^k\}$ $k = 1, 2, \ldots$ By the definition of the infimum, for each k it follows that $X_k \neq \emptyset$. Therefore, a sequence of points $\{x_k\} \subset X$ can be constructed by selecting a point x_k for each $k = 1, 2, \ldots$ Since X is bounded, there exists a convergent subsequence $\{x_k\}_K \subset X$ indexed by the set $K \subset \mathbb{N}$ with x^* as its limit. Since X is closed it follows that $x^* \in X$. By continuity of f on X, since $\alpha \leq f(x_k) \leq \alpha + \varepsilon^k$, we have $\alpha = \lim_{k \to \infty, k \in K} f(x_k) = f(x^*)$. Therefore, there exists a solution $x^* \in X$ so that $f(x^*) = \alpha = \inf\{f(x) : x \in X\}$, i.e. x^* is a minimizing solution. ♦

All the hypotheses of this theorem are important. The feasible set must be *nonempty*, otherwise there are no feasible points at which the minimum is attained. The feasible set must be *closed*, i.e. it must contain its boundary points. The objective function must be *continuous* on the feasible set, otherwise the limit at a point may not exist or be different from the value of the function at that point. Finally, the feasible set must be *bounded*, otherwise even continuous functions can be unbounded on the feasible set.

Definition 9 *(Convex Program). Let C be a convex set in $\mathbb{R}^n$ and let $f : C \to \mathbb{R}^n$ be a convex function on C. Then, $\min\{f(x) : x \in C\}$ is called a convex optimization problem, or a convex program.*

The fundamental result in convex programming is the following theorem.

Theorem 2. *Let x^* be a local minimum of a convex program. Then, x^* is also a global minimum.*

Proof. If x^* is a local minimum, then there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for any $x \in B(x^*, \varepsilon).$ Now, suppose that x^* is not a global minimum. Then there exists $y \in C$ such that $f(y) < f(x^*).$ Let $\lambda \in (0,1)$ be chosen such that the point $z = \lambda y + (1-\lambda)x^* \in B(x^*, \varepsilon).$ By convexity of $C,$ $z \in C.$ Therefore,

$$f(z) \leq \lambda f(y) + (1-\lambda) f(x^*) < \lambda f(x^*) + (1-\lambda) f(x^*) = f(x^*),$$

which is a contradiction, since x^* is a local minimum. ♦

3. Optimality conditions for unconstrained optimization

Let us consider the problem of minimizing of a function $f(x)$ without constraints on the variables $x \in \mathbb{R}^n$:

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

For a given point $x \in \mathbb{R}^n$ the optimality conditions determine whether or not a point is a local or a global minimum of f. To formulate the optimality conditions it is necessary to introduce some concepts which characterize an improving direction along which the values of the function f decrease.

Definition 10 *(Descent Direction). Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at x^*. A vector $d \in \mathbb{R}^n$ is a descent direction for f at x^* if there exists $\delta > 0$ such that $f(x^* + \lambda d) < f(x^*)$ for any $\lambda \in (0,\delta).$ The cone of descent directions at x^*, denoted by $C_{dd}(x^*)$ is given by*

$$C_{dd}(x^*) = \{d : \text{there exists } \delta > 0 \text{ such that } f(x^* + \lambda d) < f(x^*), \text{ for any } \lambda \in (0,\delta)\}.$$

Assume that f is a differentiable function. To get an algebraic characterization for a descent direction for f at x^* let us define the set

$$C_0(x^*) = \{d : \nabla f(x^*)^{\mathrm{T}} d < 0\}.$$

The following result shows that every $d \in C_0(x^*)$ is a descent direction at x^*.

Proposition 1 *(Algebraic Characterization of a Descent Direction). Suppose that* $f:\mathbb{R}^n \to \mathbb{R}$ *is differentiable at* x^**. If there exists a vector* d *such that* $\nabla f(x^*)^{\mathrm{T}} d < 0,$ *then* d *is a descent direction for* f *at* x^**, i.e.* $C_0(x^*) \subseteq C_{dd}(x^*)$*.*

Proof. Since f is differentiable at x^*, it follows that

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)^{\mathrm{T}} d + \lambda \|d\| o(\lambda d),$$

where $\lim_{\lambda\to 0} o(\lambda d) = 0.$ Therefore,

$$\frac{f(x^* + \lambda d) - f(x^*)}{\lambda} = \nabla f(x^*)^{\mathrm{T}} d + \|d\| o(\lambda d).$$

Since $\nabla f(x^*)^{\mathrm{T}} d < 0$ and $\lim_{\lambda\to 0} o(\lambda d) = 0,$ it follows that there exists a $\delta > 0$ such that $\nabla f(x^*)^{\mathrm{T}} d + \|d\| o(\lambda d) < 0$ for all $\lambda \in (0,\delta).$ ♦

Theorem 3 *(First-Order Necessary Conditions for a Local Minimum). Suppose that* $f:\mathbb{R}^n \to \mathbb{R}$ *is differentiable at* x^**. If* x^* *is a local minimum, then* $\nabla f(x^*) = 0.$

Proof. Suppose that $\nabla f(x^*) \neq 0.$ If we consider $d = -\nabla f(x^*),$ then $\nabla f(x^*)^{\mathrm{T}} d = -\|\nabla f(x^*)\|^2 < 0.$ By Proposition 1 there exists a $\delta > 0$ such that for any $\lambda \in (0,\delta),$ $f(x^* + \lambda d) < f(x^*).$ But this is in contradiction with the assumption that x^* is a local minimum for f. ♦

Observe that the above necessary condition represents a system of n algebraic nonlinear equations. All the points x^* which solve the system $\nabla f(x) = 0$ are called *stationary points*. Clearly, the stationary points need not all be local minima. They could very well be local maxima or even saddle points. In order to characterize a local minimum we need more restrictive necessary conditions involving the Hessian matrix of the function f.

Theorem 4 *(Second-Order Necessary Conditions for a Local Minimum). Suppose that* $f:\mathbb{R}^n \to \mathbb{R}$ *is twice differentiable at point* x^**. If* x^* *is a local minimum, then* $\nabla f(x^*) = 0$ *and* $\nabla^2 f(x^*)$ *is positive semidefinite.*

Proof. Consider an arbitrary direction d. Then, using the differentiability of f at x^* we get:

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2}\lambda^2 d^{\mathrm{T}} \nabla^2 f(x^*) d + \lambda^2 \|d\|^2 o(\lambda d),$$

where $\lim_{\lambda\to 0} o(\lambda d) = 0.$ Since x^* is a local minimum, $\nabla f(x^*) = 0.$ Therefore,

$$\frac{f(x^*+\lambda d)-f(x^*)}{\lambda^2}=\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x^*)d+\|d\|^2 o(\lambda d).$$

Since x^* is a local minimum, for λ sufficiently small, $f(x^*+\lambda d)\geq f(x^*)$. For $\lambda\to 0$ it follows from the above equality that $d^{\mathrm{T}}\nabla^2 f(x^*)d\geq 0$. Since d is an arbitrary direction, it follows that $\nabla^2 f(x^*)$ is positive semidefinite. ♦

In the above theorems we have presented the *necessary* conditions for a point x^* to be a local minimum, i.e. these conditions *must be satisfied* at every local minimum solution. However, a point satisfying these necessary conditions need not be a local minimum. In the following theorems the *sufficient* conditions for a global minimum are given, provided that the objective function is convex on $\mathbb{R}^n$.

Theorem 5 *(First-Order Sufficient Conditions for a Strict Local Minimum). Suppose that* $f:\mathbb{R}^n\to\mathbb{R}$ *is differentiable at* x^* *and convex on* $\mathbb{R}^n$. *If* $\nabla f(x^*)=0$, *then* x^* *is a global minimum of* f *on* $\mathbb{R}^n$.

Proof. Since f is convex on $\mathbb{R}^n$ and differentiable at x^* then from the property of convex functions given by the Proposition A1 it follows that for any $x\in\mathbb{R}^n$ $f(x)\geq f(x^*)+\nabla f(x^*)^{\mathrm{T}}(x-x^*)$. But x^* is a stationary point, i.e. $f(x)\geq f(x^*)$ for any $x\in\mathbb{R}^n$. ♦

The following theorem gives second-order sufficient conditions characterizing a local minimum point for those functions which are strictly convex in a neighborhood of the minimum point.

Theorem 6 (*Second-Order Sufficient Conditions for a Strict Local Minimum*). *Suppose that* $f:\mathbb{R}^n\to\mathbb{R}$ *is twice differentiable at point* x^*. *If* $\nabla f(x^*)=0$ *and* $\nabla^2 f(x^*)$ *is positive definite, then* x^* *is a local minimum of* f.

Proof. Since f is twice differentiable, for any $d\in\mathbb{R}^n$, we can write:

$$f(x^*+d)=f(x^*)+\nabla f(x^*)^{\mathrm{T}}d+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x^*)d+\|d\|^2 o(d),$$

where $\lim_{d\to 0}o(d)=0$. Let λ be the smallest eigenvalue of $\nabla^2 f(x^*)$. Since $\nabla^2 f(x^*)$ is positive definite, it follows that $\lambda>0$ and $d^{\mathrm{T}}\nabla^2 f(x^*)d\geq\lambda\|d\|^2$. Therefore, since $\nabla f(x^*)=0$, we can write:

$$f(x^*+d)-f(x^*)\geq\left[\frac{\lambda}{2}+o(d)\right]\|d\|^2.$$

Since $\lim_{d\to 0} o(d) = 0,$ then there exists a $\eta > 0$ such that $|o(d)| < \lambda/4$ for any $d \in B(0,\eta).$ Hence

$$f(x^* + d) - f(x^*) \geq \frac{\lambda}{4}\|d\|^2 > 0$$

for any $d \in B(0,\eta) \setminus \{0\},$ i.e. x^* is a strict local minimum of function f. ♦

If we assume f to be twice continuously differentiable, we observe that, since $\nabla^2 f(x^*)$ is positive definite, $\nabla^2 f(x^*)$ is positive definite in a small neighborhood of x^* and so f is strictly convex in a small neighborhood of x^*. Therefore, x^* is a strict local minimum, that is, it is the unique global minimum over a small neighborhood of x^*.

4. Optimality conditions for problems with inequality constraints

In the following we shall consider the nonlinear optimization problems with inequality constraints:

$$\begin{aligned} &\min f(x) \\ &\text{subject to:} \qquad\qquad (3) \\ &x \in X, \end{aligned}$$

where X is a general set. Later on, we will be more specific and define the problem as to minimize $f(x)$ subject to $c(x) \leq 0,$ where $c(x)$ is the vector of constraint functions.

Definition 11 *(Feasible direction). Let X be a nonempty set in $\mathbb{R}^n$. A nonzero vector $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in cl(X)$ if there exists a $\delta > 0$ such that $x^* + \eta d \in X$ for any $\eta \in (0,\delta).$ Moreover, the cone of feasible directions at x^*, denoted by $C_{fd}(x^*),$ is given by*

$$C_{fd}(x^*) \triangleq \{d \neq 0, \text{there is } \delta > 0 \text{ such that } x^* + \eta d \in X, \text{for any } \eta \in (0,\delta)\}.$$

Clearly, a small movement from x^* along the direction $d \in C_{fd}(x^*)$ leads to feasible points. On the other hand, a similar movement along a direction $d \in C_0(x^*)$ (see Definition 10 of descent direction) leads to solutions which improve the value of the objective function. The following theorem, which gives a geometrical interpretation of local minima, shows that a necessary condition for local optimality is that: *every improving direction is not a feasible direction*.

Theorem 7 *(Geometric Necessary Condition for a Local Minimum). Let* X *be a nonempty set in* $\mathbb{R}^n$ *and let* $f:\mathbb{R}^n \to \mathbb{R}$ *be a differentiable function. Suppose that* x^* *is a local minimum of the problem (3). Then,* $C_0(x^*) \cap C_{fd}(x^*) = \varnothing$.

Proof. Suppose that there exists a nonzero vector $d \in C_0(x^*) \cap C_{fd}(x^*)$. By the Proposition 1 of algebraic characterization of a descent direction there exists $\delta_1 > 0$ such that $f(x^* + \eta d) < f(x^*)$ for any $\eta \in (0, \delta_1)$. On the other hand, by Definition 11 of feasible direction there exists $\delta_2 > 0$ such that $x^* + \eta d \in X$ for any $\eta \in (0, \delta_2)$. Therefore, there exists $x \in B(x^*, \eta) \cap X$ such that $f(x^* + \eta d) < f(x^*)$, for every $\eta \in (0, \min\{\delta_1, \delta_2\})$, which contradicts the assumption that x^* is a local minimum of f on X (see Definition 7). ♦

So far we have obtained a geometric characterization of the optimality condition for the problem (3) given by Theorem 7 where $C_{fd}(x^*)$ is the cone of feasible directions. To get practical optimality condition, implementable in computer programs, we need to convert this geometric condition into an algebraic one. For this we introduce the concept of *active constraints* at x^* and define a cone $C_{ac}(x^*) \subseteq C_{fd}(x^*)$ in terms of the gradients of these active constraints. Now we specify the feasible set X as:

$$X \triangleq \{x : c_i(x) \le 0, i = 1, \ldots, m\}, \tag{4}$$

where $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are continuous functions. Define the vector $c(x) = [c_1(x), \ldots, c_m(x)]$.

Definition 12 *(Active constraint, Active-Set). Let* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, *which define the feasible set* $X = \{x : c_i(x) \le 0, i = 1, \ldots, m\}$, *and consider* $x^* \in X$ *a feasible point. For each* $i = 1, \ldots, m$, *the constraint* c_i *is said to be active or binding at* x^* *if* $c_i(x^*) = 0$. *It is said to be inactive at* x^* *if* $c_i(x^*) < 0$. *The set*

$$A(x^*) \triangleq \{i : c_i(x^*) = 0\},$$

denotes the set of active constraints at x^*.

The following proposition gives an algebraic characterization of a feasible direction showing the relation between a cone $C_{ac}(x^*)$ expressed in terms of the gradients of the active constraints and the cone of the feasible directions.

Proposition 2 *(Algebraic Characterization of a Feasible Direction). Let* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,m$, *be differentiable functions and consider the feasible set* $X = \{x : c_i(x) \le 0, i = 1,\ldots,m\}$. *For any feasible point* $x^* \in X$, *we have:*

$$C_{ac}(x^*) \triangleq \{d : \nabla c_i(x^*)^{\mathrm{T}} d < 0, i \in A(x^*)\} \subseteq C_{fd}(x^*).$$

Proof. Suppose that $C_{ac}(x^*)$ is a nonempty set. Let $d \in C_{ac}(x^*)$. Observe that $\nabla c_i(x^*)^{\mathrm{T}} d < 0$ for each $i \in A(x^*)$. Therefore, by Proposition 1, the algebraic characterization of a descent direction; it follows that d is a descent direction for c_i at x^*, i.e. there exists $\delta_2 > 0$ such that $c_i(x^* + \eta d) < c_i(x^*) = 0$ for any $\eta \in (0, \delta_2)$ and for any $i \in A(x^*)$. On the other hand, since c_i is differentiable at x^*, it follows that it is continuous at x^*. Therefore, since $c_i(x^*) < 0$ and c_i is continuous at x^* for each $i \notin A(x^*)$, there exists $\delta_1 > 0$ such that $c_i(x^* + \eta d) < 0$ for any $\eta \in (0, \delta_1)$ and for any $i \notin A(x^*)$. Besides, for all $\eta \in (0, \min\{\delta_1, \delta_2\})$, the points $x^* + \eta d \in X$. Therefore, by Definition 11 of feasible direction, $d \in C_{fd}(x^*)$. ♦

Remark 1. From the Theorem 7 we know that $C_0(x^*) \cap C_{fd}(x^*) = \varnothing$. But, from Proposition 2 we have that $C_{ac}(x^*) \subseteq C_{fd}(x^*)$. Therefore $C_0(x^*) \cap C_{ac}(x^*) = \varnothing$, for any local optimal solution x^*. ♦

It is worth saying that the above geometric characterization of local optimal solution (see Theorem 7) holds either at interior points $\operatorname{int} X \triangleq \{x \in \mathbb{R}^n : c_i(x) < 0, i = 1,\ldots,m\}$, or boundary points. For interior points any direction is feasible and the necessary condition $C_0(x^*) \cap C_{ac}(x^*) = \varnothing$ reduces to the very well known condition $\nabla f(x^*) = 0$, which is identical to the necessary optimal condition for unconstrained optimization (see the Theorem 3).

It is important to notice that the condition $C_0(x^*) \cap C_{ac}(x^*) = \varnothing$ can be satisfied by non-optimal points, i.e. this condition is *necessary but not sufficient* for a point x^* to be a local minimum of function f on X. For example, at any point x^* for which $\nabla c_i(x^*) = 0$ for an arbitrary index $i \in A(x^*)$ the condition $C_0(x^*) \cap C_{ac}(x^*) = \varnothing$ is trivially satisfied.

In the following, in order to get an algebraic necessary optimality condition to be used in numerical computation, we want to transform the geometric necessary optimality condition $C_0(x^*) \cap C_{ac}(x^*) = \varnothing$ to a statement in terms of the gradient of the objective function and the gradient of the constraints. Thus, the first-order optimality conditions, known as the *Karush-Kuhn-Tucker (KKT) necessary* *conditions* are obtained. In order to formulate the KKT conditions we need to introduce the concepts of *regular point* and of *KKT point*.

Definition 13 *(Regular Point – Inequality Constraints). Let* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, *be differentiable functions and consider the feasible set* $X = \{x \in \mathbb{R}^n : c_i(x) \le 0, i = 1, \ldots, m\}$. *A point* $x^* \in X$ *is a regular point if the gradient vectors* $\nabla c_i(x^*)$, $i \in A(x^*)$, *are linear independent, i.e.*

$$rank[\nabla c_i(x^*), i \in A(x^*)] = card(A(x^*)).$$

Definition 14 *(KKT point). Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, *be differentiable functions. Consider the problem* $\min\{f(x) : c(x) \le 0\}$. *If a point* $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m$ *satisfies the algebraic conditions:*

$$\nabla f(x^*) + (\mu^*)^{\mathrm{T}} \nabla c(x^*) = 0, \tag{5}$$

$$\mu^* \ge 0, \tag{6}$$

$$c(x^*) \le 0, \tag{7}$$

$$(\mu^*)^{\mathrm{T}} c(x^*) = 0. \tag{8}$$

then (x^*, μ^*) *is called a KKT point.*

In the Definition 14, the scalars μ_i, $i = 1, \ldots, m$, are called the *Lagrange multipliers*. The first condition (5) is known as the *primal feasibility* condition. The conditions (6) and (7) are known as *dual feasibility* conditions. The last condition (8), expressed as $\mu_i^* c_i(x^*) = 0$, $i = 1, \ldots, m$, are the *complementarity slackness* (or *transversality*) condition.

Now we are in position to present the KKT necessary condition for optimality of the nonlinear optimization problem with inequality constraints. For this a very useful result is given by the Theorem of Gordan (see Theorem A2). This is extensively used in the derivation of the optimality conditions of linear and nonlinear programming problems.

Theorem 8 *(KKT Necessary Conditions). Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, *be differentiable functions. Consider the problem* $\min\{f(x) : c(x) \le 0\}$. *If* x^* *is a local minimum and a regular point of the constraints, then there exists a unique vector* μ^* *such that* (x^*, μ^*) *is a KKT point.*

Proof. We know that x^* is an optimal solution for the problem $\min\{f(x) : c(x) \le 0\}$. Therefore, using the Remark 1, no direction $d \in \mathbb{R}^n$ exists such that $\nabla f(x^*)^{\mathrm{T}} d < 0$ and $\nabla c_i(x^*)^{\mathrm{T}} d < 0$, for any $i \in A(x^*)$ are simultaneously satisfied. Now, let $A \in \mathbb{R}^{(card(A(x^*)+1)\times n}$ be the matrix whose rows are $\nabla f(x^*)^T$ and $\nabla c_i(x^*)^T$, $i \in A(x^*)$. By the Gordan Theorem (see Theorem A2) there exists a nonzero vector $p = [u_0, u_1, \dots, u_{card(A(x^*))}] \geq 0$ in $\mathbb{R}^{card(A(x^*))+1}$ such that $A^T p = 0$. Therefore,

$$u_0 \nabla f(x^*) + \sum_{i \in A(x^*)} u_i \nabla c_i(x^*) = 0,$$

where $u_0 \geq 0$ and $u_i \geq 0$ for $i \in A(x^*)$ and $[u_0, u_1, \dots, u_{card(A(x^*))}]$ is not the vector zero. Considering $u_i = 0$ for all $i \notin A(x^*)$, the following conditions are obtained:

$$\begin{aligned}
&u_0 \nabla f(x^*) + u^T \nabla c(x^*) = 0, \\
&u^T c(x^*) = 0, \\
&u_0 \geq 0, \;\; u \geq 0, \\
&(u_0, u) \neq (0,0),
\end{aligned}$$

where u is the vector with components u_i for $i = 1, \dots, m$, some of them being $u_0, u_1, \dots, u_{card(A(x^*))}$ and the others being zero. Observe that $u_0 \neq 0$, because otherwise the assumption that the gradient of the active constraints are linear independent at x^* is not satisfied. Now, considering the vector μ^* as the vector u whose components are divided by u_0, we get that (x^*, μ^*) is a KKT point. ♦

The above theorem shows the importance of the active constraints. A major difficulty in applying this result is that we do not know, in advance, which constraints are active and which are inactive at solution of the problem. In other words, we do not know the active-set. The majority of algorithms for solving this optimization problem with inequalities face this difficulty of identifying the active-set. Of course, the idea of investigating all possible active-sets of a problem in order to get the points satisfying the KKT conditions is usually impractical.

Remark 2 *(Constrained qualification)*. Observe that *not every local minimum is a KKT point*. For a local minimum x^* to be a KKT point an additional condition must be introduced on the behavior of the constraints. Such a condition is known as *constraint qualification*. In Theorem 8 such a constraint qualification is that x^* be a regular point, which also is known as the *linear independence constraint qualification* (LICQ). The Lagrange multipliers are guaranteed to be unique in the Theorem 8 if LICQ holds. Another weaker constraint qualification is the *Mangasarian-Fromovitz constraint qualification* (MFCQ). The Mangasarian-Fromovitz constraint qualification requires that there exists (at least) one direction $d \in C_{ac}(x^*)$, i.e. such that $\nabla c_i(x^*)^T d < 0$, for each $i \in A(x^*)$. The MFCQ is weaker than LICQ, i.e. the Lagrange multipliers are guaranteed to be unique if LICQ holds, while this uniqueness property may be lost under MFCQ. ♦

In the following theorem we present a sufficient condition which guarantees that any KKT point of an inequality constrained nonlinear optimization problem is a global minimum of the problem. Of course, this result is obtained under the convexity hypothesis.

Theorem 9 *(KKT Sufficient Conditions). Let* $f:\mathbb{R}^n \to \mathbb{R}$ *and* $c_i:\mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,m$, *be convex and differentiable functions. Consider the problem* $\min\{f(x):c(x)\le 0\}$. *If* (x^*,μ^*) *is a KKT point, then* x^* *is a global minimum of the problem.*

Proof. Let us define the function $L(x)\triangleq f(x)+\sum_{i=1}^m \mu_i^* c_i(x)$. Since f and c_i, $i=1,\dots,m$, are convex functions and $\mu_i^*\ge 0$, $i=1,\dots,m$, it follows that L is also convex. Now, the dual feasibility conditions determine that $\nabla L(x^*)=0$. Therefore, by Theorem 5, x^* is a global minimum for L on $\mathbb{R}^n$, i.e. $L(x)\ge L(x^*)$ for any $x\in\mathbb{R}^n$. Therefore, for any x such that $c_i(x)\le c_i(x^*)=0$, $i\in A(x^*)$, it follows that

$$f(x)-f(x^*)\ge -\sum_{i\in A(x^*)} \mu_i^*[c_i(x)-c_i(x^*)]\ge 0.$$

On the other hand, the set $\{x\in\mathbb{R}^n : c_i(x)\le 0, i\in A(x^*)\}$ contains the feasible set $\{x\in\mathbb{R}^n : c_i(x)\le 0, i=1,\dots,m\}$. Therefore, x^* is a global minimum for the problem with inequality constraints. ♦

5. Optimality conditions for problems with equality constraints

In this section, the nonlinear optimization problem with equality constraints is considered:

$$\min f(x)$$

subject to: (9)

$$h_i(x)=0,\ \ i=1,\dots,p,$$

where $f:\mathbb{R}^n \to \mathbb{R}$ and $h_i:\mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,p$, are continuously differentiable functions. The functions $h_i(x)=0$, $i=1,\dots,p$, of the above problem define the vector $h(x)=[h_1(x),\dots h_p(x)]$. If x^* satisfies the constraints from (9), i.e. $h_i(x^*)=0$, $i=1,\dots,p$, it is said to be *feasible*. Otherwise it is called *infeasible*.

The optimality of x^* can be seen as a balance between function minimization and constraint satisfaction. A move away from x^* cannot be made without either violating a constraint or increasing the value of the objective function. Formally, this can be stated as the following proposition.

Proposition 3 *(Balance Between Function and Constraints). If x^* is a solution of (9) and $x^* + \delta x$ is a nearby point, then:*

1) *If $f(x^* + \delta x) < f(x^*)$ then $h_i(x^* + \delta x) \neq 0$ for some i.*
2) *If $h_1(x^* + \delta x) = \ldots = h_p(x^* + \delta x) = 0$ then $f(x^* + \delta x) \geq f(x^*)$.* ♦

In order to establish the optimality conditions for the nonlinear optimization problems with equality constraints we need to introduce some relevant concepts. An equality constraint $h(x) = 0$ defines in $\mathbb{R}^n$ a set which can be viewed as a *hypersurface*. When there are p equality constraints $h_i(x) = 0,\ i = 1,\ldots,p,$ then their intersection defines a (possible empty) set:

$$X \triangleq \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1,\ldots,p\}.$$

If the functions defining the equality constraints are differentiable, then the set X is said to be a *differentiable manifold*, or a *smooth manifold*.

Now, in any point on a differentiable manifold the *tangent set* can be defined as follows. A *curve* η on a manifold X is a continuously application $\eta : I \subset \mathbb{R} \to X$, i.e. a family of points $\eta(t) \in X$ continuously parameterized by t in the interval $I \subset \mathbb{R}.$ Clearly, a curve passes through the point x^* if $x^* = \eta(t^*)$ for some $t^* \in I$. The *derivative* of a curve at t^*, if it exist, is defined in a classical manner as

$$\dot{\eta}(t^*) \triangleq \lim_{\xi \to 0} \frac{\eta(t^* + \xi) - \eta(t^*)}{\xi}.$$

A curve is *differentiable*, or smooth, if a derivative exists for each $t \in I$.

Definition 15 *(Tangent Set). Let X be a differentiable manifold in $\mathbb{R}^n$ and a point $x^* \in X$. Consider the collection of all the continuously differentiable curves on X passing through x^*. Then, the collection of all the vectors tangent to these curves at x^* is the tangent set to X at x^*, denoted by $T_X(x^*)$.*

Definition 16 *(Regular point – Equality Constraints). Let $h_i : \mathbb{R}^n \to \mathbb{R},\ i = 1,\ldots,p,$ be differentiable functions on $\mathbb{R}^n$ and consider the set $X \triangleq \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1,\ldots,p\}$. A point $x^* \in X$ is a regular point if the gradient vectors $\nabla h_i(x^*),\ i = 1,\ldots,p,$ are linearly independent, i.e.*

$$rank[\nabla h_1(x^*),\ldots,\nabla h_p(x^*)] = p. \tag{10}$$

If the constraints are regular, in the sense of the above definition, then X is a subspace of dimension $n - p$. In this case $T_X(x^*)$ is a subspace of dimension

$n-p$, called *tangent space*. At regular points the tangent space can be characterized in terms of the gradients of the constraints [Luenberger, 1973].

Proposition 4 *(Algebraic Characterization of a Tangent Space). Let* $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,p$, *be differentiable functions on* $\mathbb{R}^n$ *and consider the set* $X \triangleq \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1,\ldots,p\}$. *At a regular point* $x^* \in X$ *the tangent space is such that*

$$T_X(x^*) = \{d : \nabla h(x^*)^{\mathrm{T}} d = 0\}. \tag{11}$$

Proof. Let $T_X(x^*)$ be the tangent space at x^* and $M(x^*) = \{d : \nabla h(x^*)^{\mathrm{T}} d = 0\}$. Consider any curve $\eta(t)$ passing through x^* at $t = t^*$, having derivative $\dot{\eta}(t^*)$ such that $\nabla h(x^*)^{\mathrm{T}} \dot{\eta}(t^*) \neq 0$. Since such a curve would not lie on X, it follows that $T_X(x^*) \subset M(x^*)$. Now to prove that $T_X(x^*) \supset M(x^*)$ we must show that if $d \in M(x^*)$ then there is a curve on X passing through x^* with derivative d. In order to construct such a curve we consider the equations

$$h(x^* + td + \nabla h(x^*)^{\mathrm{T}} u(t)) = 0,$$

where for fixed t the vector $u(t) \in \mathbb{R}^p$ is unknown. Observe that we have a nonlinear system of p equations with p unknowns, continuously parameterized by t. At $t = 0$ there is a solution $u(0) = 0$. The Jacobian matrix of the above system with respect to u at $t = 0$ is the matrix $\nabla h(x^*)\nabla h(x^*)^{\mathrm{T}}$ which is nonsingular, since $\nabla h(x^*)$ is of full rank if x^* is a regular point. Thus, by the Implicit Function Theorem (Theorem A1) there is a continuous solution $u(t)$ for $-a \le t \le a$. The curve $\eta(t) = x^* + td + \nabla h(x^*)^{\mathrm{T}} u(t)$ by construction is a curve on X. By differentiating the above nonlinear system with respect to t at $t = 0$ we get:

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} h(\eta(t))\bigg|_{t=0} = \nabla h(x^*)^{\mathrm{T}} d + \nabla h(x^*)\nabla h(x^*)^{\mathrm{T}} \dot{u}(0).$$

By definition of d we have $\nabla h(x^*)^{\mathrm{T}} d = 0$. Therefore, since $\nabla h(x^*)\nabla h(x^*)^{\mathrm{T}}$ is nonsingular, it follows that $\dot{u}(0) = 0$. Therefore, $\dot{\eta}(0) = d + \nabla h(x^*)^{\mathrm{T}} \dot{u}(0) = d$ and the constructed curve has derivative d at x^*. ♦

The Method of Lagrange Multipliers.

In the following we present the optimality conditions for the nonlinear optimization problems with equality constraints using the method of Lagrange multipliers. The idea is to restrict the search of a minimum of (9) to the manifold

$X \triangleq \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1,\ldots,p\}$. The following theorem gives the geometric necessary condition for a local minimum of a nonlinear optimization problem with equality constraints. It is shown that the tangent space $T_X(x^*)$ at a regular local minimum point x^* is *orthogonal* to the gradient of the objective function at x^*.

Theorem 10 *(Geometric Necessary Condition for a Local Minimum). Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,p$, *be continuously differentiable functions. Suppose the* x^* *is a local minimum point of the problem* $\min\{f(x) : h(x) = 0\}$. *Then,* $\nabla f(x^*)$ *is orthogonal to the tangent space* $T_X(x^*)$, *i.e.*

$$C_0(x^*) \cap T_X(x^*) = \varnothing.$$

Proof. Assume that there exists a $d \in T_X(x^*)$ such that $\nabla f(x^*)^{\mathrm{T}} d \neq 0$. Let $\eta : I = [-a,a] \to X$, $a > 0$, be any smooth curve passing through x^*, with $\eta(0) = x^*$ and $\dot{\eta}(0) = d$. Also let φ be the function defined as $\varphi(t) \triangleq f(\eta(t))$ for any $t \in I$. Since x^* is a local minimum of f on $X \triangleq \{x \in \mathbb{R}^n : h(x) = 0\}$, by Definition 7, it follows that there exists $\delta > 0$ such that $\varphi(t) = f(\eta(t)) \geq f(x^*) = \varphi(0)$ for any $t \in B(0,\delta) \cap I$. Therefore, $t^* = 0$ is an unconstrained local minimum point for φ, and

$$0 = \nabla\varphi(0) = \nabla f(x^*)^{\mathrm{T}} \dot{\eta}(0) = \nabla f(x^*)^{\mathrm{T}} d.$$

But, this is in contradiction with the assumption that $\nabla f(x^*)^{\mathrm{T}} d \neq 0$. ♦

The conclusion of this theorem is that if x^* is a regular point of the constraints $h(x) = 0$ and a local minimum point of f subject to these constraints, then all $d \in \mathbb{R}^n$ satisfying $\nabla h(x^*)^{\mathrm{T}} d = 0$ must also satisfy $\nabla f(x^*)^{\mathrm{T}} d = 0$.

The following theorem shows that this property that $\nabla f(x^*)$ is orthogonal to the tangent space implies that $\nabla f(x^*)$ is a linear combination of the gradients of $h_i(x^*)$, $i = 1,\ldots,p$, at x^*. This relation leads to the introduction of Lagrange multipliers and Lagrange function.

Theorem 11 *(First-Order Necessary Optimality Conditions). Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,p$, *be continuously differentiable functions. Consider the problem* $\min\{f(x) : h(x) = 0\}$. *If* x^* *is a local minimum and is a regular point of the constraints, then there exists a unique vector* $\lambda^* \in \mathbb{R}^p$ *such that*

$$\nabla f(x^*) + \nabla h(x^*)^{\mathrm{T}} \lambda^* = 0. \tag{12}$$

Proof. Since x^* is a local minimum of f on $X = \{x \in \mathbb{R}^n : h(x) = 0\}$, by Theorem 10, it follows that $C_0(x^*) \cap T_X(x^*) = \varnothing$, i.e. the system

$$\nabla f(x^*)^{\mathrm{T}} d < 0, \quad \nabla h(x^*)^{\mathrm{T}} d = 0,$$

is inconsistent. Now, consider the following two sets:

$$C_1 \triangleq \{(z_1, z_2) \in \mathbb{R}^{p+1} : z_1 = \nabla f(x^*)^{\mathrm{T}} d,\, z_2 = \nabla h(x^*)^{\mathrm{T}} d\},$$
$$C_2 \triangleq \{(z_1, z_2) \in \mathbb{R}^{p+1} : z_1 < 0,\, z_2 = 0\}.$$

Observe that C_1 and C_2 are convex sets and $C_1 \cap C_2 = \varnothing$. Therefore, by the separation of two convex sets, given by Proposition A2, there exists a nonzero vector $(\mu, \lambda) \in \mathbb{R}^{p+1}$ $(\mu \in \mathbb{R},\ \lambda \in \mathbb{R}^p)$ such that for any $d \in \mathbb{R}^n$ and for any $(z_1, z_2) \in C_2$

$$\mu \nabla f(x^*)^{\mathrm{T}} d + \lambda^{\mathrm{T}} [\nabla h(x^*)^{\mathrm{T}} d] \geq \mu z_1 + \lambda^{\mathrm{T}} z_2.$$

Now, considering $z_2 = 0$ and having in view that z_1 can be made an arbitrary large negative number, it follows that $\mu \geq 0.$ Additionally, considering $(z_1, z_2) = (0,0),$ we must have $[\mu \nabla f(x^*) + \lambda^{\mathrm{T}} \nabla h(x^*)]^{\mathrm{T}} d \geq 0,$ for any $d \in \mathbb{R}^n.$ In particular, letting $d = -[\mu \nabla f(x^*) + \lambda^{\mathrm{T}} \nabla h(x^*)],$ it follows that $-\left\| \mu \nabla f(x^*) + \lambda^{\mathrm{T}} \nabla h(x^*) \right\|^2 \geq 0,$ and thus,

$$\mu \nabla f(x^*) + \lambda^{\mathrm{T}} \nabla h(x^*) = 0, \text{ with } (\mu, \lambda) \neq (0,0).$$

Observe that $\mu > 0,$ for otherwise the above relation would contradict the assumption that $\nabla h_i(x^*),$ $i = 1, \ldots, p,$ are linear independent. The conclusion of the theorem follows letting $\lambda^* = \lambda / \mu$ and noting that the linear independence assumption implies the uniqueness of the $\lambda^*.$ ♦

Remark 3. The first-order necessary conditions given by the theorem 11 together with the constraints of the problem (9):

$$\nabla f(x^*) + \nabla h(x^*)^{\mathrm{T}} \lambda^* = 0, \tag{13a}$$

$$h(x^*) = 0, \tag{13b}$$

represent a total of $n + p$ nonlinear equations in the variables $(x^*, \lambda^*).$ These conditions determine, at least locally, a unique solution $(x^*, \lambda^*).$ However, as in the unconstrained case, a solution to the first-order necessary optimality conditions need not be a local minimum of the problem (9). ♦

Definition 17 *(Lagrange multipliers). The scalars* $\lambda_1^*, \ldots, \lambda_p^*$ *in (12) are called the Lagrange multipliers.*

Definition 18 *(Constraint normals). The vectors* $\nabla h_1(x), \ldots, \nabla h_p(x)$ *are called the constraint normals.*

The condition (12) shows that $\nabla f(x^*)$ is linearly dependent on the constraint normals. Therefore, a constrained minimum occurs when the gradients of the objective function and the constraints interact in such a way that any reduction in f can only be obtained by violating the constraints.

Definition 19 *(Lagrange function - Lagrangian). The function* $L:\mathbb{R}^n\times\mathbb{R}^p\to\mathbb{R}$ *associated to the nonlinear optimization problem (9) is defined as:*

$$L(x,\lambda)\triangleq f(x)+\lambda^{\mathrm{T}}h(x). \tag{14}$$

Remark 4 *(Regularity Assumption)*. It is worth saying that for a local minimum to satisfy the above first-order necessary conditions (13) and, in particular, for unique Lagrange multipliers to exist, it is necessary that the equality constraints $h_i(x)=0,$ $i=1,\dots,p,$ satisfy a regularity condition. As we have already seen, for a local minimum of an inequality constrained nonlinear optimization problem to be a KKT point a constrained qualification is needed. For the equality constrained nonlinear optimization problems the condition that *the minimum point is a regular point* corresponds *to linear independence constrained qualification*. ♦

If x^* is a local minimum of (9) which is regular, then the first-order necessary optimality conditions (13) can be rewritten as:

$$\nabla_x L(x^*,\lambda^*)=0, \tag{15a}$$

$$\nabla_\lambda L(x^*,\lambda^*)=0. \tag{15b}$$

Observe that the second condition (15b) is a restatement of the constraints. The solution of the optimization problem (9) corresponds to a saddle point of the Lagrangian.

In the following the second-order necessary optimality conditions for a point to be a local minimum for (9) are presented.

Theorem 12 *(Second-Order Necessary Optimality Conditions). Let* $f:\mathbb{R}^n\to\mathbb{R}$ *and* $h_i:\mathbb{R}^n\to\mathbb{R},$ $i=1,\dots,p,$ *be continuously differentiable functions. Consider the problem* $\min\{f(x):h(x)=0\}$. *If* x^* *is a local minimum and it is a regular point of the constraints, then there exists a unique vector* $\lambda^*\in\mathbb{R}^p$ *such that*

$$\nabla f(x^*)+\nabla h(x^*)^{\mathrm{T}}\lambda^*=0, \tag{16}$$

and

$$d^{\mathrm{T}}\left(\nabla^2 f(x^*)+\nabla^2 h(x^*)^{\mathrm{T}}\lambda^*\right)d\ge 0 \tag{17}$$

for any $d\in\mathbb{R}^n$ *such that* $\nabla h(x^*)^{\mathrm{T}}d=0.$

Proof. The first condition $\nabla f(x^*)+\nabla h(x^*)^{\mathrm{T}}\lambda^*=0$ follows from the Theorem 11. Now we concentrate to the second condition. Let x^* be a regular point and consider d an arbitrary direction from $T_X(x^*)$, i.e. $\nabla h(x^*)^{\mathrm{T}} d = 0$. Let $\eta : I = [-a, a] \to X$, $a > 0$, be an arbitrary twice-differential curve passing through x^* with $\eta(0) = x^*$ and $\dot{\eta}(0) = d$. Consider φ a function defined as $\varphi(t) \triangleq f(\eta(t))$, for any $t \in I$. Since x^* is a local minimum of f on $X \triangleq \{x \in \mathbb{R}^n : h(x) = 0\}$, it follows that $t^* = 0$ is an unconstrained local minimum point for φ. Therefore, by Theorem 4, it follows that

$$\nabla^2 \varphi(0) = \dot{\eta}(0)^{\mathrm{T}} \nabla^2 f(x^*) \dot{\eta}(0) + \nabla f(x^*)^{\mathrm{T}} \ddot{\eta}(0) \geq 0.$$

On the other hand, differentiating the relation $h(\eta(t))^{\mathrm{T}} \lambda = 0$ twice, we get:

$$\dot{\eta}(0)^{\mathrm{T}} \left(\nabla^2 h(x^*)^{\mathrm{T}} \lambda \right) \dot{\eta}(0) + \left(\nabla h(x^*)^{\mathrm{T}} \lambda \right)^{\mathrm{T}} \ddot{\eta}(0) = 0.$$

Now, adding the last two relations we obtain:

$$d^{\mathrm{T}} \left(\nabla^2 f(x^*) + \nabla^2 h(x^*)^{\mathrm{T}} \lambda^* \right) d \geq 0,$$

which must hold for every d such that $\nabla h(x^*)^{\mathrm{T}} d = 0$. ♦

The above theorem says that if $T_X(x^*)$ is the tangent space to X at x^*, then the matrix $\nabla^2_{xx} L(x^*, \lambda^*) = \nabla^2 f(x^*) + \nabla^2 h(x^*)^{\mathrm{T}} \lambda^*$ is positive semidefinite on $T_X(x^*)$. In other words, the matrix $\nabla^2_{xx} L(x^*, \lambda^*)$, (which is the Hessian of the Lagrange function) restricted to the subspace $T_X(x^*)$ is positive semidefinite.

Remark 5 *(Feasible Directions and Second-Order Conditions)* An n-vector d is said to be a *feasible direction* at x^* if $\nabla h(x^*) d = 0$, where $\nabla h(x^*)$ is the Jacobian of the constraints at x^*. Let us assume d is a feasible direction normalized so that $\|d\| = 1$. Considering the Taylor's expansion:

$$h(x^* + \varepsilon d) = h(x^*) + \varepsilon \nabla h(x^*) d + O(\|\varepsilon d\|^2),$$

then $h(x^* + \varepsilon d) = O(\varepsilon^2)$. Therefore, a move away from x^* along d keeps the constraints satisfied to first-order accuracy. In particular, if all the constraints in (9) are linear, then $x^* + \varepsilon d$ is a feasible point for all $\varepsilon > 0$. On the other hand, if any of the $h_i(x)$ in (9) are nonlinear, then d is a direction *tangential* to the constraints at x^*. It is easy to see that condition (12) implies that, for any feasible direction d,

$$d^{\mathrm{T}} \nabla f(x^*) = 0.$$

To distinguish a *minimum* from a *maximum* or a *saddle-point*, the second-order optimality condition must be used. These conditions can be stated as:

1) If the constraint functions h_i are all linear, the second-order condition that guarantees x^* is a minimum of problem (9) is

$$d^{\mathrm{T}} \nabla^2 f(x^*) d > 0$$

for any feasible direction d.

2) If the constraint functions h_i are nonlinear, the second-order condition that guarantees x^* is a minimum of problem (9) is

$$d^{\mathrm{T}}\nabla^2 L(x^*,\lambda^*)d > 0$$

for any feasible direction d. ♦

Remark 6 *(Eigenvalues in Tangent Space)*. Geometrically, the restriction of the matrix $\nabla^2_{xx}L(x^*,\lambda^*)$ to $T_X(x^*)$ corresponds to the projection $P_{T_X(x^*)}[\nabla^2_{xx}L(x^*,\lambda^*)]$. A vector $y \in T_X(x^*)$ is an *eigenvector* of the projection $P_{T_X(x^*)}[\nabla^2_{xx}L(x^*,\lambda^*)]$ if there is a real number ν such that

$$P_{T_X(x^*)}[\nabla^2_{xx}L(x^*,\lambda^*)]y = \nu y.$$

The real number ν is called the *eigenvalue* of $P_{T_X(x^*)}[\nabla^2_{xx}L(x^*,\lambda^*)]$. To obtain a matrix representation for $P_{T_X(x^*)}[\nabla^2_{xx}L(x^*,\lambda^*)]$ it is necessary to introduce a basis of the tangent subspace $T_X(x^*)$. It is best to introduce an orthonormal basis, say $E = [e_1,\dots,e_{n-p}]$. Any vector $y \in T_X(x^*)$ can be written as $y = Ez$, where $z \in \mathbb{R}^{n-p}$. Now, $\nabla^2_{xx}L(x^*,\lambda^*)Ez$ represents the action of $\nabla^2_{xx}L(x^*,\lambda^*)$ on such a vector. To project this result back into $T_X(x^*)$ and express the result in terms of the basis $E = [e_1,\dots,e_{n-p}]$ it is necessary to multiply by E^{T}. Therefore, $E^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)Ez$ is the vector whose components give the representation in terms of the basis E. The $(n-p)\times(n-p)$ matrix $E^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)E$ is the matrix representation of $\nabla^2_{xx}L(x^*,\lambda^*)$ restricted to $T_X(x^*)$. *The eigenvalues of $\nabla^2_{xx}L(x^*,\lambda^*)$ restricted to $T_X(x^*)$ can be determined by computing the eigenvalues of $E^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)E$.* These eigenvalues are independent of the particular choice of the basis E. ♦

Recall that the conditions given in Theorems 11 and 12 are necessary conditions. These must hold at each local minimum point. However, a point satisfying these conditions may not be a local minimum. As in the unconstrained case it is possible to derive second-order conditions for constrained optimization problems. The following theorem provides sufficient conditions for a stationary point of the Lagrange function to be a local minimum.

Theorem 13 *(Second-Order Sufficient Conditions). Let $f:\mathbb{R}^n \to \mathbb{R}$ and $h_i:\mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,p$, be twice continuously differentiable functions. Consider the problem $\min\{f(x): h(x) = 0\}$. If x^* and λ^* satisfy:*

$$\nabla_x L(x^*,\lambda^*) = 0, \tag{18a}$$

$$\nabla_\lambda L(x^*,\lambda^*)=0 \tag{18b}$$

and

$$y^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)y>0 \tag{19}$$

for any $y\neq 0$ *such that* $\nabla h(x^*)^{\mathrm{T}}y=0,$ *then* x^* *is a strict local minimum.*

Proof. Consider the augmented Lagrange function:

$$\bar{L}(x,\lambda)=f(x)+\lambda^{\mathrm{T}}h(x)+\frac{c}{2}\|h(x)\|^2,$$

where c is a scalar. Clearly,

$$\nabla_x\bar{L}(x,\lambda)=\nabla_x L(x,\bar{\lambda}),$$

$$\nabla^2_{xx}\bar{L}(x,\lambda)=\nabla^2_{xx}L(x,\bar{\lambda})+c\nabla h(x)^{\mathrm{T}}\nabla h(x),$$

where $\bar{\lambda}=\lambda+ch(x).$ Since (x^*,λ^*) satisfy the sufficient conditions, by the Theorem A3, we obtain that $\nabla_x\bar{L}(x^*,\lambda^*)=0$ and $\nabla^2_{xx}\bar{L}(x^*,\lambda^*)>0,$ for sufficiently large $c.$ $\bar{L}$ being positive definite at $(x^*,\lambda^*),$ it follows that there exist $\rho>0$ and $\delta>0$ such that

$$\bar{L}(x,\lambda^*)\geq \bar{L}(x^*,\lambda^*)+\frac{\rho}{2}\left\|x-x^*\right\|^2$$

for $\left\|x-x^*\right\|<\delta.$ Besides, since $\bar{L}(x,\lambda^*)=f(x)$ when $h(x)=0,$ we get

$$f(x)\geq f(x^*)+\frac{\rho}{2}\left\|x-x^*\right\|^2$$

if $h(x)=0,$ $\left\|x-x^*\right\|<\delta,$ i.e. x^* is a strict local minimum. ♦

Sensitivity – Interpretation of the Lagrange Multipliers

The ith Lagrange multiplier can be viewed as measuring the sensitivity of the objective function with respect to the ith constraint, i.e. how much the optimal value of the objective function would change if that constraint were perturbed.

For the very beginning let us consider $p=1,$ i.e. the problem (9) has one constraint $h_1(x)=0.$ Now, suppose that x^* is a local solution of the problem

$$\min\{f(x):h_1(x)=0\},$$

and consider the *perturbed* problem

$$\min\{f(x):h_1(x)=\delta\},$$

where δ is a known scalar. If the solution of the perturbed problem is $x^*+u,$ then using the Taylor's expansion a first-order estimate of the optimum function value is $f(x^*+u)\approx f(x^*)+u^{\mathrm{T}}\nabla f(x^*).$ But, the optimality condition for the original problem given by (12) states that $\nabla f(x^*)=-\lambda_1^*\nabla h_1(x^*),$ where λ_1^* is the Lagrange multiplier. Hence

$$f(x^*+u)\approx f(x^*)-\lambda_1^*u^{\mathrm{T}}\nabla h_1(x^*).$$

Since $x^* + u$ solve the perturbed problem, it follows that $h_1(x^* + u) = \delta,$ hence $h_1(x^*) + u^T \nabla h_1(x^*) \approx \delta.$ But, $h_1(x^*) = 0.$ Therefore, $u^T \nabla h_1(x^*) \approx \delta,$ that is

$$f(x^* + u) - f(x^*) \approx -\delta\lambda_1^*.$$

In other words, *the Lagrange multiplier is an approximate measure of the change in the objective function that will occur if a unit amount is added to the right-hand side of the constraint.* In general we have the following theorem.

Theorem 14 *(Interpretation of the Lagrange Multipliers). Consider the family of problems* $\min\{f(x) : h(x) = w\},$ *where* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $h : \mathbb{R}^n \to \mathbb{R}^p$ *are twice continuously differentiable. Suppose for* $w = 0$ *there is a local solution* x^* *that is a regular point and that, together with its associated Lagrange multiplier vector* $\lambda,$ *satisfies the second-order sufficient conditions for a strict local minimum. Then, for every* $w \in \mathbb{R}^p$ *in a region containing* 0 *there is a* $x(w),$ *depending continuously on* $w,$ *such that* $x(0) = x^*$ *and such that* $x(w)$ *is a local minimum of the problem. Furthermore,*

$$\nabla_w f(x(w))\big|_{w=0} = -\lambda.$$

Proof. Consider the system of equations:

$$\nabla f(x) + \nabla h(x)^T \lambda = 0,$$
$$h(x) = w.$$

By hypothesis, when $w = 0,$ there is a solution x^*, λ^* to this system. The Jacobian matrix of this system, at this solution is

$$\begin{bmatrix} L(x^*) & \nabla h(x^*)^T \\ \nabla h(x^*) & 0 \end{bmatrix},$$

where $L(x^*) = \nabla^2 f(x^*) + \nabla^2 h(x^*)^T \lambda^*.$

Since x^* is a regular point and $L(x^*)$ is positive definite on $\{y : \nabla h(x^*)^T y = 0\},$ it follows that this matrix is nonsingular. Thus, by the Implicit Function Theorem (Theorem A1), there is a solution $x(w), \lambda(w)$ to the system which is twice continuously differentiable. Therefore,

$$\nabla_w f(x(w))\big|_{w=0} = \nabla f(x^*)^T \nabla_w x(0),$$
$$\nabla_w h(x(w))\big|_{w=0} = \nabla h(x^*) \nabla_w x(0).$$

But, since $h(x^*) = w$ it follows that $\nabla h(x^*) \nabla_w x(0) = I.$ On the other hand, from $\nabla f(x) + \nabla h(x)^T \lambda = 0$ it follows that $\nabla_w f(x(c))\big|_{w=0} = -\lambda.$ ♦

6. Optimality conditions for general problems

In the following we present a generalization of the Theorems 8, 11, 12 and 13 to nonlinear optimization problems with equality and inequality constraints:

$$\min f(x)$$

subject to: (20)

$$c_i(x) \le 0, \quad i = 1, \dots, m,$$
$$h_j(x) = 0, \quad j = 1, \dots, p,$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p$, are continuous differentiable functions. Define the vectors: $c(x) = [c_1(x), \dots, c_m(x)]$ and $h(x) = [h_1(x), \dots, h_p(x)]$.

Remark 7 *(Discarding the Inactive Constraints)*. Let us consider the nonlinear optimization problem $\min\{f(x) : c_i(x) \le 0, i = 1, \dots, m\}$. Suppose that x^* is a local minimum point for this problem. Clearly, x^* is also a local minimum of the above problem where the inactive constraints $c_i(x) \le 0$, $i \notin A(x^*)$ have been *discarded*. Therefore, *the inactive constraints at x^* can be ignored in the statement of the optimality conditions*. On the other hand, the active constraints can be treated as equality constraints at a local minimum point. Hence, x^* is also a local minimum point to the equality constrained problem:

$$\min\{f(x) : c_i(x) = 0, i \in A(x^*)\}$$

The difficulty is that we do not know the set of the active constraints at x^*.

From Theorem 11 it follows that if x^* is a regular point, there exists a unique Lagrange multiplier vector $\mu^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) + \sum_{i \in A(x^*)} \mu_i^* \nabla c_i(x^*) = 0.$$

Now, assigning zero Lagrange multipliers to the inactive constraints, we get:

$$\nabla f(x^*) + \nabla c(x^*)^{\mathrm{T}} \mu^* = 0,$$
$$\mu_i = 0, \ i \notin A(x^*).$$

Clearly, the last condition can be rewritten as $\mu_i^* c_i(x^*) = 0$, $i = 1, \dots, m$.

It remains to show that $\mu > 0$. For this assume that $\mu_q < 0$ for some $q \in A(x^*)$. Now, let $\mathrm{A} \in \mathbb{R}^{(m+1)\times n}$ be the matrix whose rows are $\nabla f(x^*)$ and $\nabla c_i(x^*)$, $i = 1, \dots, m$. Since x^* is a regular point, it follows that the Lagrange multiplier vector μ^* is unique. Therefore the condition $\mathrm{A}^{\mathrm{T}} y = 0$, can only be satisfied by $y^* = \gamma\begin{pmatrix}1 & \mu^*\end{pmatrix}^{\mathrm{T}}$ with $\gamma \in \mathbb{R}$. But, $\mu_q < 0$. Therefore, by Gordan's Theorem A2 there exists a direction $\bar{d} \in \mathbb{R}^n$ such that $\mathrm{A}\bar{d} < 0$. In other words, $\bar{d} \in C_0(x^*) \cap C_{ac}(x^*) \ne \emptyset$, which contradicts the hypothesis that x^* is a local minimum of the problem. All these results represent the KKT optimality conditions as stated by Theorem 8. Although this development is straightforward, it is somewhat limited by the regularity-type assumption at the optimal solution. ♦

Definition 20 *(Regular Point – General Case). Let* $c_i:\mathbb{R}^n\to\mathbb{R},\ i=1,\ldots,m,$ *and* $h_j:\mathbb{R}^n\to\mathbb{R},\ \ j=1,\ldots,p,$ *be continuously differentiable functions. Consider the set* $X=\{x\in\mathbb{R}^n: c_i(x)\le 0, i=1,\ldots,m, h_j(x)=0, j=1,\ldots,p\}$. *A point* $x^*\in X$ *is a regular point of the constraints from (20) if the gradients* $\nabla c_i(x^*),\ i\in A(x^*)$ *and* $\nabla h_j(x^*),\ \ j=1,\ldots,p,$ *are linearly independent.* ♦

The Definition 20 introduces the Linear Independence Constraint Qualification (LICQ) for general nonlinear optimization problems, i.e. the gradients of the *active* inequality constraints and the gradients of the equality constraints are linearly independent at x^*. Another constraint qualification is the *Linear Constraint Qualification* (LCQ): i.e. $c_i(x), i=1,\ldots,m,$ and $h_j:\mathbb{R}^n\to\mathbb{R},\ \ j=1,\ldots,p,$ are affine functions. Another one is the *Slater condition*, for a convex problem, i.e. there exists a point $\bar{x}$ such that $c_i(\bar{x})<0,\ i=1,\ldots,m,$ and $h(\bar{x})=0.$

We emphasize that the constraint qualification ensures that the linearized approximation to the feasible set X captures the essential shape of X in a neighborhood of x^*.

Theorem 15 *(First- and Second-Order Necessary Conditions). Let* $f:\mathbb{R}^n\to\mathbb{R},$ $c_i:\mathbb{R}^n\to\mathbb{R},\ \ i=1,\ldots,m,$ *and* $h_i:\mathbb{R}^n\to\mathbb{R},\ \ i=1,\ldots p,$ *be twice continuously differentiable functions. Consider the problem* $\min\{f(x): c(x)\le 0, h(x)=0\}$. *If* x^* *is a local minimum for this problem and it is a regular point of the constraints, then there exist unique vectors* $\mu^*\in\mathbb{R}^m$ *and* $\lambda^*\in\mathbb{R}^p$ *such that:*

$$\nabla f(x^*)+\nabla c(x^*)^{\mathrm{T}}\mu^*+\nabla h(x^*)^{\mathrm{T}}\lambda^*=0, \tag{21a}$$

$$\mu^*\ge 0, \tag{21b}$$

$$c(x^*)\le 0, \tag{21c}$$

$$h(x^*)=0, \tag{21d}$$

$$(\mu^*)^T c(x^*)=0, \tag{21e}$$

and

$$y^{\mathrm{T}}\left(\nabla^2 f(x^*)+\nabla^2 c(x^*)^{\mathrm{T}}\mu^*+\nabla^2 h(x^*)^{\mathrm{T}}\lambda^*\right)y\ge 0, \tag{22}$$

for all $y\in\mathbb{R}^n$ *such that* $\nabla c_i(x^*)^{\mathrm{T}}y=0,\ i\in A(x^*)$ *and* $\nabla h(x^*)^{\mathrm{T}}y=0.$ ♦

Proof. Observe that since $\mu^*\ge 0$ and $c(x^*)\le 0,$ (21e) is equivalent to the statement that a component of μ^* is nonzero only if the corresponding constraint is active. Since x^* is a minimum point over the constraint set, it is also a minimum over the subset of that set defined by setting the active constraints to zero. Therefore, for the resulting equality constrained problem defined in a neighborhood of x^*, there are Lagrange multipliers. Hence, (21a) holds with $\mu_i^* = 0$ if $c_i(x^*) \neq 0$.

It remains to show that $\mu^* \geq 0$. This is a little more elaborate. Suppose that for some $k \in A(x^*)$, $\mu_k^* < 0$. Let $\bar{X}$ and $\bar{T}$ be the surface and the tangent space, respectively, defined by all the other active constraints at x^*. By the regularity assumptions, there is a d such that $d \in \bar{T}$ and $\nabla c_k(x^*)^{\mathrm{T}} d < 0$. Let $\eta(t)$ be a curve on $\bar{X}$ passing through x^* at $t = 0$ with $\dot{\eta}(0) = d$. Then, for small $t \geq 0$, it follows that $\eta(t)$ is feasible and

$$\left.\frac{\mathrm{d}f}{\mathrm{d}t}(\eta(t))\right|_{t=0} = \nabla f(x^*)^{\mathrm{T}} d < 0$$

by (21a), which contradicts the fact that x^* is a minimum point. ♦

The conditions (21) are known as the *Karush-Kuhn-Tucker conditions*, or *KKT conditions*. The conditions (21e) written as $\mu_i^* c_i(x^*) = 0$, $i = 1, \ldots, m$, are the *complementary conditions*. They show that either constraint i is active or the corresponding Lagrange multiplier $\mu_i^* = 0$, or possibly both. For a given nonlinear optimization problem (20) and a solution point x^*, there may be many Lagrange multipliers (μ^*, λ^*) for which the conditions (21) and (22) are satisfied. However, when x^* is a regular point (the LICQ is satisfied), the optimal (μ^*, λ^*) are unique.

The KKT conditions motivate the following definition which *classifies* constraints according to whether or not their corresponding Lagrange multiplier is zero.

Definition 21 *(Strongly Active (Binding) – Weakly Active Constraints). Let x^* be a local solution to the problem (20) and the Lagrange multipliers (μ^*, λ^*) which satisfy the KKT conditions (21). We say that an inequality constraint $c_i(x)$ is strongly active or binding if $i \in A(x^*)$ and the corresponding Lagrange multiplier $\mu_i^* > 0$. We say that $c_i(x)$ is weakly active if $i \in A(x^*)$ and the corresponding Lagrange multiplier $\mu_i^* = 0$.* ♦

The nonnegativity condition (21b) on the Lagrange multiplier for the inequality constraints ensures that the function $f(x)$ will not be reduced by a move off any of the binding constraints at x^* to the interior of the feasible region.

A special case of complementarity is important because it introduces the concept of *degeneracy* in optimization.

Definition 22 *(Strict Complementarity). Let x^* be a local solution to the problem (20) and the Lagrange multipliers (μ^*,λ^*) which satisfy the KKT conditions (21). We say that the strict complementarity holds if exactly one of μ_i^* and $c_i(x^*)$ is zero for each index $i=1,\dots,m.$ In other words, we have $\mu_i^*>0$ for each $i\in A(x^*)$.* ♦

Usually, satisfaction of strict complementarity is beneficial for algorithms and makes it easier to determine the active set $A(x^*)$ so that convergence is more rapid.

Remark 8 *(Degeneracy).* A property that causes difficulties for some optimization algorithms is degeneracy. This concept refers to the following two situations:
- The gradient of the active constraints $\nabla c_i(x^*),\ i\in A(x^*),$ are linearly dependent at the solution point $x^*.$ Linear dependence of the gradient of the active constraints can cause difficulties during the computation of the step direction because certain matrices that must be factorized become rank deficient.
- Strict complementarity fails to hold, that is, there is some index $i\in A(x^*)$ such that all the Lagrange multipliers satisfying the KKT conditions (21) have $\mu_i^*=0.$ In the case when the problem contains weakly active constraints it is difficult for an algorithm to determine whether these constraints are active at the solution. For some optimization algorithms (active-set algorithms and gradient projection algorithms) the presence of weakly active constraints can cause the algorithm to zigzag as the iterates move on and off the weakly constraints along the successive iterations. ♦

Theorem 16 *(Second-Order Sufficient Conditions). Let $f:\mathbb{R}^n\to\mathbb{R},\ c_i:\mathbb{R}^n\to\mathbb{R},\ i=1,\dots,m,$ and $h_j:\mathbb{R}^n\to\mathbb{R},\ \ j=1,\dots p,$ be twice continuously differentiable functions. Consider the problem $\min\{f(x):c(x)\le 0, h(x)=0\}$. If there exist x^*, μ^* and λ^* satisfying the KKT conditions (21a)-(21e) and*

$$y^{\mathrm{T}}\nabla_{xx}^2 L(x^*,\mu^*,\lambda^*)y>0,$$

for all $y\neq 0$ such that

$$\nabla c_i(x^*)^{\mathrm{T}}y=0,\quad i\in A(x^*)\ \text{with}\ \mu_i^*>0, \tag{23a}$$

$$\nabla c_i(x^*)^{\mathrm{T}}y\le 0,\quad i\in A(x^*)\ \text{with}\ \mu_i^*=0, \tag{23b}$$

$$\nabla h(x^*)^{\mathrm{T}}y=0, \tag{23c}$$

where $L(x,\mu,\lambda)=f(x)+\mu^{\mathrm{T}}c(x)+\lambda^{\mathrm{T}}h(x),$ then x^* is a strict local minimum of the problem. ♦

Proof. The theorem says that, the Hessian of the Lagrangian is positive definite on the critical cone $C(x^*,\mu^*,\lambda^*)$ defined by (23) for x^*, μ^* and λ^* satisfying the KKT conditions (21a)-(21e).
Assume that x^* is not a strict local minimum and let $\{y_k\}$ be a sequence of feasible points converging to x^* such that $f(y_k) \le f(x^*)$. Consider y_k of the form $y_k = x^* + \delta_k s_k$ with $\delta_k > 0$ and $\|s_k\| = 1$. Assume that $\delta_k \to 0$ and $s_k \to s^*$. Clearly, $\nabla f(x^*)^{\mathrm{T}} s^* \le 0$ and $\nabla h_j(x^*) s^* = 0$ for $j = 1,\dots,p$.
On the other hand, for each active constraint c_i we have $c_i(y_k) - c_i(x^*) \le 0$. Therefore, $\nabla c_i(x^*)^{\mathrm{T}} s^* \le 0$.
If $\nabla c_i(x^*)^{\mathrm{T}} s^* = 0$, for all $i \in \{l : c_l(x^*) = 0, \mu_l^* > 0\}$, then the proof is similar to that in Theorem 13. If $\nabla c_i(x^*)^{\mathrm{T}} s^* < 0$ for at least one $i \in \{l : c_l(x^*) = 0, \mu_l^* > 0\}$, then

$$0 \ge \nabla f(x^*)^{\mathrm{T}} s^* = -\lambda^{\mathrm{T}} \nabla h(x^*) s^* - \mu^{\mathrm{T}} \nabla c(x^*) s^* > 0,$$

which represents a contradiction. ♦

The KKT sufficient conditions for convex programming with inequality constraints given in Theorem 9 can immediately be generalized to nonlinear optimization problems with convex inequalities and affine equalities.

Theorem 17 *(KKT Sufficient Conditions for General Problems). Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,m$, *be convex and differentiable functions. Also let* $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,p$, *be affine functions. Consider the problem* $\min f(x)$ *subject to* $x \in X \triangleq \{x \in \mathbb{R}^n : c(x) \le 0, h(x) = 0\}$. *If* (x^*,μ^*,λ^*) *satisfies the KKT conditions (21a)-(21e), then* x^* *is a global minimum for* f *on* X. ♦

Sensitivity – Interpretation of the Lagrange Multipliers for General Problems
As we have already seen Theorem 14 presents an interpretation of the Lagrange multipliers for nonlinear optimization problems with equality constraints. Each Lagrange multiplier tells us something about the *sensitivity* of the optimal objective function value $f(x^*)$ with respect to the corresponding constraint. Clearly, for an inactive constraint $i \notin A(x^*)$ the solution x^* and the function value $f(x^*)$ are independent of whether this constraint is present or not. If we slightly perturb c_i by a tiny amount, it will still be inactive and therefore x^* will still be a local solution of the optimization problem. Since $\mu_i^* = 0$ from (21e), the Lagrange multiplier shows that constraint i has no importance in the system of the constraints. Otherwise, as in Theorem 14 the following theorem can be presented.

Theorem 18 *(Interpretation of the Lagrange Multipliers for General Problems). Consider the family of problems* $\min\{f(x): c(x) \le v, h(x) = w\},$ *where* $f:\mathbb{R}^n \to \mathbb{R}, c:\mathbb{R}^n \to \mathbb{R}^m$ *and* $h:\mathbb{R}^n \to \mathbb{R}^p$ *are twice continuously differentiable. Suppose for* $v = 0, w = 0$ *there is a local solution* x^* *that is a regular point and that, together with its associated Lagrange multiplier* $\mu^* \ge 0, \lambda^*,$ *satisfies the second-order sufficient conditions for a strict local minimum. Then, for every* $(v,w) \in \mathbb{R}^{m+p},$ *in a region containing* $(0,0) \in \mathbb{R}^{m+p},$ *there is a solution* $x(v,w)$ *continuously depending on* $(v,w),$ *such that* $x(0,0) = x^*$ *and such that* $x(v,w)$ *is a local minimum of the problem. Furthermore,*

$$\nabla_v f(x(v,w))\big|_{0,0} = -\mu^*.$$
$$\nabla_w f(x(v,w))\big|_{0,0} = -\lambda^*.$$
♦

Appendix

Proposition A1 *(First-Order Condition of Convexity). Let* C *be a convex set in* $\mathbb{R}^n$ *with a nonempty interior. Consider the function* $f:C \to \mathbb{R}$ *which is continuous on* C *and differentiable on* $\mathrm{int}(C).$ *Then,* f *is convex on* $\mathrm{int}(C)$ *if and only if* $f(y) \ge f(x) + \nabla f(x)^{\mathrm{T}}(y-x)$ *for any points* $x, y \in C.$ ♦

Proposition A2 *(Separation of Two Convex Sets). Let* C_1 *and* C_2 *be two nonempty and convex set in* $\mathbb{R}^n.$ *Suppose that* $C_1 \cap C_2 = \varnothing.$ *Then, there exists a hyperplane that separates* C_1 *and* $C_2,$ *i.e. there exists a nonzero vector* $p \in \mathbb{R}^n$ *such that* $p^{\mathrm{T}} x_1 \ge p^{\mathrm{T}} x_2$ *for any* $x_1 \in cl(C_1)$ *and for any* $x_2 \in cl(C_2).$ ♦

Theorem A1 *(Implicit Function Theorem). Let* $h:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ *be a function such that:*

1) $h(z^*,0) = 0$ *for some* $z^* \in \mathbb{R}^n.$

2) The function $h(.,.)$ *is continuously differential in some neighborhood of* $(z^*,0).$

3) $\nabla_z h(z,t)$ *is nonsingular at the point* $(z,t) = (z^*,0).$

Then, there exists an open sets $N_z \subset \mathbb{R}^n$ *and* $N_t \subset \mathbb{R}^m$ *containing* z^* *and* $0,$ *respectively, and a continuous function* $z: N_t \to N_z$ *such that* $z^* = z(0)$ *and* $h(z(t),t) = 0$ *for all* $t \in N_t.$ $z(t)$ *is uniquely defined. If* h *is* q *times continuously differentiable with respect to both its arguments for some* $q > 0,$ *then* $z(t)$ *is also* q *times continuously differentiable with respect to* t *and*

$$\nabla z(t) = -\nabla_t h(z(t),t)\left[\nabla_z h(z(t),t)\right]^{-1},$$

for all $t \in N_t$.

The implicit function theorem is applied to parameterized system of linear equations in which z is obtained as the solution of $M(t)z = g(t)$, where $M(.) \in \mathbb{R}^{n\times n}$ has $M(0)$ nonsingular and $g(t) \in \mathbb{R}^n$ (see the algebraic characterization of a tangent space). To apply the theorem, define $h(z,t) = M(t)z - g(t)$. If $M(.)$ and $g(.)$ are continuously differential in some neighborhood of 0, the theorem implies that $z(t) = M(t)^{-1} g(t)$ is a continuous function of t in some neighborhood of 0.

Theorem A2 *(Gordan's Theorem). Let A be an $m \times n$ matrix. Then exactly one of the following two statements holds:*

System 1. *There exists $x \in \mathbb{R}^n$ such that $Ax < 0$.*

System 2. *There exists $y \in \mathbb{R}^m$, $y \neq 0$ such that $A^{\mathrm{T}} y = 0$ and $y \geq 0$.*

Proof. System 1 can be equivalently written as $Ax + es \leq 0$ for some $x \in \mathbb{R}^n$ and $s > 0$, $s \in \mathbb{R}$, where e is a vector of m ones. Now, rewriting this system in the form of System 1 of Theorem A4.1, we obtain $\begin{bmatrix} A & e \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} \leq 0$ and $(0,\ldots,0,1)\begin{pmatrix} x \\ s \end{pmatrix} > 0$ for some $\begin{pmatrix} x \\ s \end{pmatrix} \in \mathbb{R}^{n+1}$. By Theorem A4.1, the associated System 2 states that $\begin{bmatrix} A^{\mathrm{T}} \\ e^{\mathrm{T}} \end{bmatrix} y = (0,\ldots,0,1)^{\mathrm{T}}$ and $y \geq 0$ for some $y \in \mathbb{R}^m$, that is, $A^{\mathrm{T}} y = 0$, $e^{\mathrm{T}} y = 1$ and $y \geq 0$ for some $y \in \mathbb{R}^m$. But, this is equivalent to System 2. ♦

Theorem A3. *Let P and Q be two symmetric matrices, such that $P \geq 0$ and $P > 0$ on the null space of Q (i.e. $y^{\mathrm{T}} P y > 0$ for any $y \neq 0$ with $Qy = 0$). Then, there exists $\overline{c} > 0$ such that $P + cQ > 0$ for any $c > \overline{c}$.*

Proof. Assume the contrary. Then, for any $k > 0$ there exists x^k, $\|x^k\| = 1$ such that $(x^k)^{\mathrm{T}} P(x^k) + k(x^k)^{\mathrm{T}} Q(x^k) \leq 0$. Consider a subsequence $\{x^k\}_K$ convergent to some $\overline{x}$ with $\|\overline{x}\| = 1$. Dividing the above inequality by k and taking the limit as $k \in K \to \infty$, we get $\overline{x}^{\mathrm{T}} Q \overline{x} \leq 0$. On the other hand, Q being semipositive definite, we must have $\overline{x}^{\mathrm{T}} Q \overline{x} \geq 0$, hence $\overline{x}^{\mathrm{T}} Q \overline{x} = 0$. Therefore, using the hypothesis, it follows that $\overline{x}^{\mathrm{T}} P \overline{x} \geq 0$. But, this contradicts the fact that $\overline{x}^{\mathrm{T}} P \overline{x} + \limsup_{k\to\infty, k\in K} k(x^k)^{\mathrm{T}} Q(x^k) \leq 0$. ♦

Notes

Plenty of books and papers are dedicated to the theoretical developments of optimality conditions for continuous nonlinear optimization. Many details and properties of theoretical aspects of optimality conditions can be found in Bertsekas [1999], Boyd and Vandenberghe [2006], Nocedal and Wright [2006], Sun and Yuan [2006], Bartholomew-Biggs [2008], etc. The content of this chapter is taken from the books by Chachuat [2007], Bazaraa, Sherali and Shetty [1993], Luenberger [1973, 1984], [Luenberger and Ye, 2008] and [Andrei, 2015]. Concerning the optimality conditions for problems with inequality constraints the material is inspired by Bazaraa, Sherali and Shetty [1993]. The derivation of the necessary and sufficient optimality conditions for problems with equality constraints follows the developments presented by Luenberger [1973]. The sensitivity analysis and interpretation of the Lagrange multipliers for nonlinear optimization is taken from Luenberger [1973]. We did not treat here the duality and the saddle point optimality condition of the Lagrangian, but these can be found, for example, in [Bazaraa, Sherali and Shetty, 1993] or [Nocedal and Wright, 2006].

The KKT conditions were originally named after Harold W. Kuhn (1925-2014), and Albert W. Tucker (1905-1995), who first published these conditions in 1951 [Kuhn and Tucker, 1951]. Later on scholars discovered that the necessary conditions for this problem had been stated by William Karush (1917-1997) in his master's thesis in 1939 [Karush, 1939]. Another approach of the optimality conditions for nonlinear optimization problem was given in 1948 by Fritz John (1910-1994) [John 1948]. (See also [Cottle, 2012].)

References

Andrei, N., (2009) *Critica Raţiunii Algoritmilor de Optimizare fără Restricţii.* [*Criticism of the Unconstrained Optimization Algorithms Reasoning*]. Editura Academiei Române, Bucureşti, 2009.

Andrei, N., (2015) *Critica Raţiunii Algoritmilor de Optimizare cu Restricţii*. [*Criticism of the Constrained Optimization Algorithms Reasoning*]. Editura Academiei Române, Bucureşti, 2015.

Bartholomew-Biggs, M., (2008) *Nonlinear Optimization with Engineering Applications.* Springer Science +Business Media, New York, 2008.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., (1993) *Nonlinear Programming Theory and Algorithms*. John Wiley, New York, 2nd edition, 1993.

Bertsekas, D.P., (1999). *Nonlinear Programming.* Athena Scientific, Belmont, MA, second ed., 1999.

Boyd, S., Vandenberghe, L., (2006) *Convex Optimization.* Cambridge University Press, Cambridge (UK), 2nd edition, 2006.

Chachuat, B.C., (2007) *Nonlinear and Dynamic Optimization – From Theory to Practice.* IC-31: Winter Semester 2006/2007. École Politechnique Fédérale de Lausanne.

Cottle, R.W., (2012) William Karush and the KKT Theorem. In: Grötschel, M. (Ed.) *Optimization stories.* Documenta Mathematica, Journal der Deutschen

Methematiker-Vereinigung, Extra volume, 21st International Symposium on Mathematical Programming, Berlin, August 19-24, 2012, pp. 255-269.

Gill, P.E., Murray, W., Wright, M.H., (1981) *Practical Optimization.* Academic Press, 1981.

John, F., (1948) Extremum problems with inequalities as subsidiary conditions. In: *Studies and Essays, presented to R. Courant on his 60th birthday,* January 8, 1948. Interscience, New York, 1948, pp.187-204.

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 1939.

Kuhn, H. W., Tucker, A. W. (1951). Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*. Berkeley: University of California Press, pp. 481–492. 1951.

Luenberger, D.G., (1973) *Introduction to linear and nonlinear programming.* Addison-Wesley Publishing Company, Reading, 1973.

Luenberger, D. G.**,** Ye, Y.**,** (2008) *Linear and nonlinear programming*. International Series in Operations Research & Management Science 116 (Third ed.). New York, Springer. 2008.

Nocedal, J., Wright, S.J., (2006) *Numerical optimization.* Springer Series in Operations Research. Springer Science+Business Media, New York, Second ed., 2006.

Peressini, A.L., Sullivan, F.E., Uhl, J.J.Jr., (1988) *The Mathematics of Nonlinear Programming.* Springer-Verlag, New York, 1988.

Sun, W., Yuan, Y.X., (2006) *Optimization Theory and Methods. Nonlinear Programming*. Springer Science + Business Media, New York, 2006.

-----oooooOooooo-----