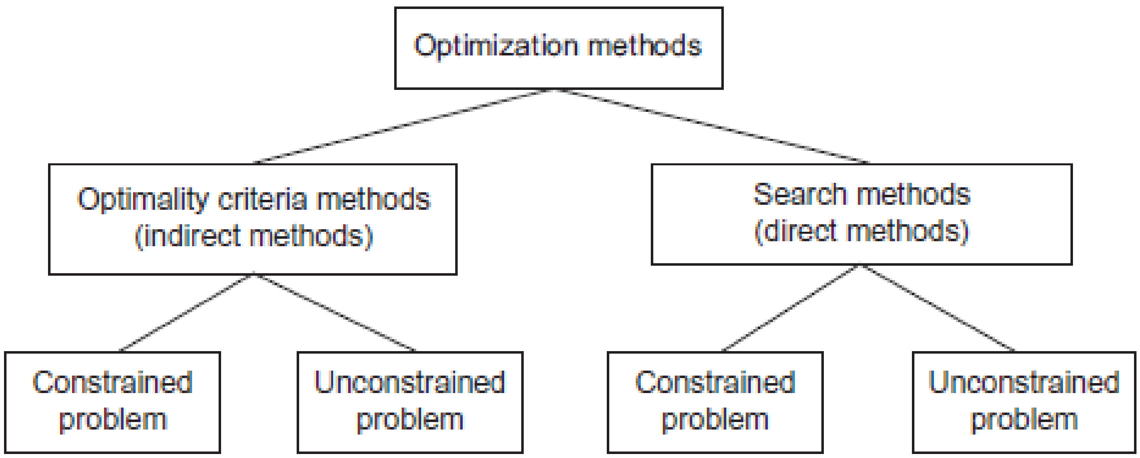
Module No. 2: Classical Optimization Techniques

Following Figure shows a broad classification of the optimization approaches for continuous variable constrained and unconstrained optimization problems.

The following two philosophically different viewpoints are shown:

- ❖ Optimality criteria (or indirect) methods and
- ❖ Search (or direct) methods.



Classification of Optimization Methods

- ❖ **Optimality Criteria Methods**—Optimality criteria are the conditions a function must satisfy at its minimum point. Optimization methods seeking solutions (perhaps using numerical methods) to the optimality conditions are often called optimality criteria or indirect methods.
- ❖ **Search Methods**—Search (direct) methods are based on a different philosophy. Here we start with an estimate of the optimum design for the problem. Usually the starting design will not satisfy the optimality criteria; therefore, it is improved iteratively until they are satisfied. Thus, in the direct approach we search the design space for optimum points.
- ❖ **Fundamental Concepts**: Optimality conditions for a minimum point are derived from fundamental calculus. Therefore, we should know the basic concepts of calculus using the vector and matrix notations.
- ❖ The concepts of gradient and Hessian of a function and Taylor's expansion of a function plays a key role in optimization methods and theory.
- ❖ **Minimum:** we defined the feasible set S (also called the constraint set, feasible region, or feasible design space) for a design problem as a collection of feasible designs:
  - $S = \{x \mid h_i(x) = 0;\ i = 1 \text{ to } p; \text{ and } g_j(x) \leq 0;\ j = 1 \text{ to } m\}$
  - Since there are no constraints in unconstrained problems, the entire design space is feasible for them.

The optimization problem is to find a point in the feasible design space that gives a minimum value to the cost function. Methods to locate optimum designs are discussed. We must first carefully define what is meant by an optimum.

In general, x^* is used to designate a particular point of the feasible set.

Global (Absolute) Minimum: A function $f(x)$ of n variables has a global (absolute) minimum at x^* if the value of the function at x^* is less than or equal to the value of the function at any other point x in the feasible set S. That is,

$f(x^*) \leq f(x)$

for all x in the feasible set S. If strict inequality holds for all x other than x^* in above Equation, then x^* is called a strong (strict) global minimum; otherwise, it is called a weak global minimum.

Local (Relative) Minimum: A function $f(x)$ of n variables has a local (relative) minimum at x^* if Inequality $f(x^*) \leq f(x)$ holds for all x in a small neighborhood N (vicinity) of x^* in the feasible set S.

If strict inequality holds, then x^* is called a strong (strict) local minimum; otherwise, it is called a weak local minimum.

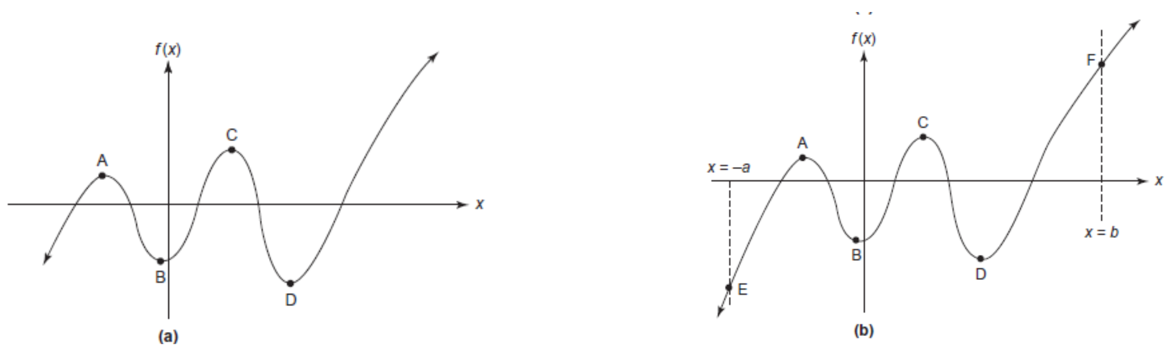
Neighborhood N of point x* is defined as the set of points –

$$N = \{\, X | X \in S \; with \; ||X \; - X * || < \; \delta \}$$

For some small $\delta > 0$. geometrically, it is a small feasible region containing the point X*.

Note that, A function $f(x)$ can have a strict local minimum at only one point in the neighborhood N (vicinity) of x^*. It may, however, have a local minimum at several points in N if the function value is the same at each of those points.

Graphical significance of global and local minima



Representation of optimum points. (a) The unbounded domain and function (no global optimum). (b) The bounded domain and function (global minimum and maximum exist).

In Part (a) of the figure, where x is between -∞ and +∞ ($-\infty \leq x \leq +\infty$), points B and D are local minima since the function has its smallest value in their neighborhood. Similarly, both A and C are points of local maxima for the function. There is, however, no global minimum or maximum for the function since the domain and the function $f(x)$ are unbounded; that is, x and $f(x)$ are allowed to have any value between - ∞ and + ∞.

If we restrict x to lie between - a and b, as in Part (b) of Figure, then point E gives the global minimum and F the global maximum for the function. Both of these points have active constraints, while points A, B, C, and D are unconstrained.

Gradient Vector: Partial Derivatives of a Function: Consider a function $f(x)$ of n variables $x_1, x_2, \ldots, x_n$. The partial derivative of the function with respect to x_1 at a given point x^* is defined as $\partial f(x^*)/\partial x_1$, with respect to x_2 as $\partial f(x^*)/\partial x_2$, and so on. Let c_i represent the partial derivative of $f(x)$ with respect to xi at the point x*. Then, using the index notation, we can represent all partial derivatives of $f(x)$ as follows:

$$c_i = \frac{\partial f(x^*)}{\partial x_i};\ i = 1 \text{ to } n$$

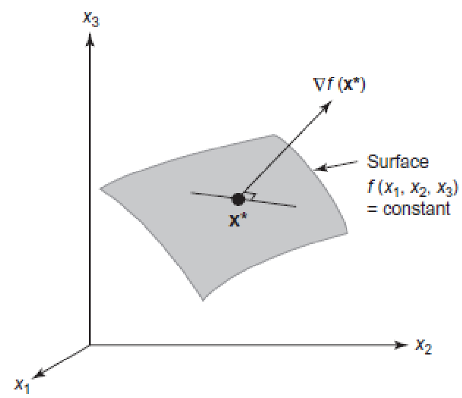
For convenience and compactness of notation, we arrange the partial derivatives $\partial f(x^*)/\partial x_1, \partial f(x^*)/\partial x_2, \ldots, \partial f(x^*)/\partial x_n$ into a column vector called the gradient vector and represent it by any of the following symbols: c, ∇f, $\partial f/\partial x$, or grad f, as –

$$c = \left[\frac{\partial f(x^*)}{\partial x_1}\frac{\partial f(x^*)}{\partial x_2} \quad \ldots\ldots\ldots.. \frac{\partial f(x^*)}{\partial x_n}\right]^T$$ where superscript T denotes transpose of a vector or a matrix. Note that all partial derivatives

are calculated at the given point x^*. That is, each component of the gradient vector is a function in itself which must be evaluated at the given point x^*.

Geometrical Representation of Gradient Vector

Geometrically, the gradient vector is normal to the tangent plane at the point x^*, as shown in Figure for a function of three variables. Also, it points in the direction of maximum increase in the function.



These properties are quite important, and used in developing optimality conditions and numerical methods for optimum design.

Hessian Matrix: Second-Order Partial Derivatives

Differentiating the gradient vector once again, we obtain a matrix of second partial derivatives for the function $f(x)$ called the Hessian matrix or, simply, the Hessian. That is, differentiating each component of the gradient vector given in Eq.

$$c = \left[\frac{\partial f(x^*)}{\partial x_1}\frac{\partial f(x^*)}{\partial x_2} \quad \ldots\ldots\ldots.. \frac{\partial f(x^*)}{\partial x_n}\right]^T$$ with respect to $x_1, x_2, \ldots, x_n$, we obtain –

$$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Where all derivatives are calculated at the given point x*. The Hessian is an n3n matrix, also denoted as *H* or $\nabla^2 f$. It is important to note that each element of the Hessian is a function in itself that is evaluated at the given point *x**.

Therefore, the Hessian is always a symmetric matrix. It plays a prominent role in the sufficiency conditions for optimality as discussed later in this chapter. It will be written as -

$$\mathbf{H} = \left[\frac{\partial^2 f}{\partial x_j \partial x_i}\right]; \quad i = 1 \text{ to } n,\ j = 1 \text{ to } n$$

Taylor's Expansion

The idea of Taylor's expansion is fundamental to the development of optimum design concepts and numerical methods, so it is explained here. A function can be approximated by polynomials in a neighborhood of any point in terms of its value and derivatives using Taylor's expansion.

Consider first a function f(x) of one variable. Taylor's expansion for *f*(*x*) about the point *x** is –

$$f(x) = f(x^*) + \frac{df(x^*)}{dx}(x - x^*) + \frac{1}{2}\frac{d^2 f(x^*)}{dx^2}(x - x^*)^2 + R$$

Where *R* is the remainder term that is smaller in magnitude than the previous terms if *x* is sufficiently close to *x**. If we let *x* - *x** = *d* (a small change in the point *x**), Taylor's expansion of above Equation becomes a quadratic polynomial in ***d***:

$$f(x^* + d) = f(x^*) + \frac{df(x^*)}{dx}d + \frac{1}{2}\frac{d^2 f(x^*)}{dx^2}d^2 + R$$

Taylor's expansion can also be written in matrix notation as –

$$f(x^* + \mathbf{d}) = f(x^*) + \nabla f^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T H \mathbf{d} + R$$

Where $x = (x_1, x_2)$, $x^* = (x_1^*, x_2^*)$, $x - x^* = d$, and *H* is the 2×2 Hessian matrix.

Quadratic Forms and Definite Matrices:

Quadratic Form: The quadratic form is a special nonlinear function having only second-order terms (either the square of a variable or the product of two variables). For example, the function –

$$F(x) = x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_2\,x_3 + 2x_3x_1$$

Quadratic forms play a prominent role in optimization theory and methods.

Generalizing the quadratic form of three variables in above Equation to n variables and writing it in the double summation notation, we obtain –

$$F(x) = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}x_ix_j$$

where p_{ij} are constants related to the coefficients of various terms in above Equation.

The Matrix of the Quadratic Form

The quadratic form can be written in the matrix notation. Let $\mathbf{P} = [p_{ij}]$ be an $n \times n$ matrix and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be an n-dimensional vector. Then the quadratic form of Eq. (4.17) is given as

$$F(\mathbf{x}) = \mathbf{x}^{\mathbf{T}}\mathbf{P}\mathbf{x} \tag{4.18}$$

$\mathbf{P}$ is called the *matrix of the quadratic form* $F(\mathbf{x})$. Elements of $\mathbf{P}$ are obtained from the coefficients of the terms in the function $F(\mathbf{x})$.

There are many matrices associated with the given quadratic form; in fact, there are infinite such matrices. All of the matrices are asymmetric except one. The symmetric matrix $\mathbf{A}$ associated with the quadratic form can be obtained from any asymmetric matrix $\mathbf{P}$ as

$$\mathbf{A} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^T) \quad \text{or} \quad a_{ij} = \frac{1}{2}(p_{ij} + p_{ji}), \;\; i, j = 1 \text{ to } n \tag{4.19}$$

Using this definition, the matrix $\mathbf{P}$ can be replaced with the symmetric matrix $\mathbf{A}$ and the quadratic form of Eq. (4.18) becomes

$$F(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \tag{4.20}$$

CONCEPT OF NECESSARY AND SUFFICIENT CONDITIONS

It is important to understand the meaning of the terms necessary and sufficient. These terms have general meaning in mathematical analyses. However we will discuss them for the optimization problem only.

Necessary Conditions: The optimality conditions are derived by assuming that we are at an optimum point, and then studying the behavior of the functions and their derivatives at that point.

The conditions that must be satisfied at the optimum point are called necessary. Stated differently, if a point does not satisfy the necessary conditions, it cannot be optimum.

Note, however, that the satisfaction of necessary conditions does not guarantee optimality of the point; that is, there can be nonoptimum points that satisfy the same conditions. This indicates that the number of points satisfying necessary conditions can be more than the number of optima.

Points satisfying the necessary conditions are called candidate optimum points. We must, therefore, perform further tests to distinguish between optimum and nonoptimum points, both satisfying the necessary conditions.

Sufficient Condition: If a candidate optimum point satisfies the sufficient condition, then it is indeed an optimum point. If the sufficient condition is not satisfied, however, or cannot be used, we may not be able to conclude that the candidate design is not optimum. Our conclusion will depend on the assumptions and restrictions used in deriving the sufficient condition. Further analyses of the problem or other higher-order conditions are needed to make a definite statement about optimality of the candidate point.

OPTIMALITY CONDITIONS: UNCONSTRAINED PROBLEM

We will discuss necessary and sufficient conditions for unconstrained optimization problems defined as "Minimize $f(x)$ without any constraints on x." Such problems arise infrequently in practical engineering applications. However, we consider them here because optimality conditions for constrained problems are a logical extension of these conditions. In addition, one numerical strategy for solving a constrained problem is to convert it into a sequence of unconstrained problems. Thus, it is important to completely understand unconstrained optimization concepts.

The optimality conditions for unconstrained or constrained problems can be used in two ways:

1. They can be used to check whether a given point is a local optimum for the problem.

2. They can be solved for local optimum points.

Concepts Related to Optimality Conditions

The basic concept for obtaining local optimality conditions is to assume that we are at a minimum point x^* and then examine its neighborhood to study properties of the function and its derivatives. Basically, we use the definition of a local minimum, given in Inequality

$f(x^*) \leq f(x)$, to derive the optimality conditions. Since we examine only a small neighborhood, the conditions we obtain are called local optimality conditions.

Let x^* be a local minimum point for $f(x)$. To investigate its neighborhood, let x be any point near x*.

Define increments d and Δf in x^* and $f(x^*)$ as $\mathbf{d} = x - x^*$ and $\Delta f = f(x) - f(x^*)$.

Since $f(x)$ has a local minimum at x^*, it will not reduce any further if we move a small distance away from x^*. Therefore, a change in the function for any move in a small neighborhood of x^* must be non-negative; that is, the function value must either remain constant or increase. This condition, also obtained directly from the definition of local minimum given in Eq. $f(x^*) \leq f(x)$, can be expressed as the following inequality:

$\Delta f = f(x) - f(x^*) \geq 0$ for all small changes $\mathbf{d}$.

The inequality in above Equation can be used to derive necessary and

sufficient conditions for a local minimum point. Since $\mathbf{d}$ is small, we can approximate Δf by Taylor's expansion at x^* and derive optimality conditions using it.

Optimality Conditions for Functions of a Single Variable:

- **First-Order Necessary Condition:** Let us first consider a function of only one variable. Taylor's expansion of f(x) at the point x* gives –

$$\boldsymbol{f(x) = f(x^*) + f'(x^*)d + \frac{1}{2} f''(x^*)d^2 + R}$$

- Where R is the remainder containing higher-order terms in d and "primes" indicate the order of the derivatives. From this equation, the change in the function at x^* (i.e., $\Delta f=f(x)-f(x^*)$) is given as –

$$\Delta \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x}^*)\boldsymbol{d} + \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{f}''(\boldsymbol{x}^*)\boldsymbol{d}^{\mathbf{2}} + \boldsymbol{R}$$

- Inequality $\Delta f = f(x) - f(x^*) \geq 0$, shows that the expression for Δf must be non-negative (≥ 0) because $x^*$is a local minimum for $f(x)$. Since d is small, the first-order term $f'(x^*)d$ dominates other terms, and therefore Δf can be approximated as –

  $\Delta f=f'(x^*)d$.

- Thus, the quantity $f'(x^*)d$ can have a negative value regardless of the sign of $f'(x^*)$, unless it is zero. The only way it can be non-negative for all d in a neighborhood of x^* is when

  $f'(x^*) = 0$

  This is a first order necessary condition for local minimum of $f(x)$ at x^*. It is called first order because it only involves the first derivative of the function.

  - **Stationary Points:** The points satisfying Eq. $f'(x^*) = 0$ can be local minima or maxima, or neither minimum nor maximum (inflection points), they are called stationary points.

  - **Sufficient Condition:** Now we need a sufficient condition to determine which of the stationary points are actually minimum for the function. Since stationary points satisfy the necessary condition

$f'(x^*)=0$, the change in function Δf of Eq.

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^*) + \boldsymbol{f}'(\boldsymbol{x}^*)\boldsymbol{d} + \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{f}''(\boldsymbol{x}^*)\boldsymbol{d}^{\mathbf{2}} + \boldsymbol{R}$$

Becomes –

$$\Delta \boldsymbol{f}(\boldsymbol{x}) = \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{f}''(\boldsymbol{x}^*)\boldsymbol{d}^{\mathbf{2}} + \boldsymbol{R}$$

Since the second-order term dominates all other higher-order terms, we need to focus on it. Note that the term can be positive for all $d \neq 0$ if

$f''(x^*) > 0$ is the sufficient condition for x^* to be a local minimum. If both conditions are satisfied by x^* then, $f(x^*)$ has the smallest value in a small neighborhood of point x^*.

Optimality Conditions for Functions of Several Variables

For the general case of a function of several variables $f(x)$ where x is an n-vector, we can repeat the derivation of the necessary and sufficient conditions using the multidimensional form of Taylor's expansion –

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^*) + \boldsymbol{\nabla f}(\boldsymbol{x}^*)^{\boldsymbol{T}}\boldsymbol{d} + \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{H}(\boldsymbol{x}^*)\boldsymbol{d} + \boldsymbol{R}$$

Alternatively, a change in the function $\Delta f = f(x) - f(x^*)$ is given as –

$$\boldsymbol{\nabla f(x)} = \boldsymbol{\nabla f(x^*)^T d} + \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{d^T H(x^*) d} + \boldsymbol{R}$$

If we assume a local minimum at x^* then Δf must be non-negative due to the definition of a local minimum given in Inequality, that is, $\Delta f > 0$. Concentrating only on the first-order term in above Equation, we observe (as before) that Δf can be non-negative for all possible d unless -

$\nabla f(x^*) = 0$

In other words, the gradient of the function at x^* must be zero. In the component form, this necessary condition becomes –

$$\frac{\partial f(x^*)}{\partial x_i} = 0; i = 1 \; to \; n$$

Points satisfying above Equation are called stationary points.

Considering the second term in Equation –

$$\boldsymbol{\nabla f(x)} = \boldsymbol{\nabla f(x^*)^T d} + \frac{\mathbf{1}}{\mathbf{2}}\, \boldsymbol{d^T H(x^*) d} + \boldsymbol{R}$$

evaluated at a stationary point, the positivity of Δf is assured if –

$\boldsymbol{d^T H(x^*) d} > 0$ for all $d \neq 0$.

This is true if the Hessian $H(x^*)$ is a positive definite matrix, which is then the sufficient condition for a local minimum of $f(x)$ at x^*.

THEOREM 4.4

Necessary and Sufficient Conditions for Local Minimum

Necessary condition. If f(x) has a local minimum at $\mathbf{x}^*$ then

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0; \quad i = 1 \text{ to } n \qquad \text{(a)}$$

Second-order necessary condition. If $f(\mathbf{x})$ has a local minimum at $\mathbf{x}^*$, then the Hessian matrix of Eq. (4.8)

$$\mathbf{H}(\mathbf{x}^*) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{(n \times n)} \qquad \text{(b)}$$

is positive semidefinite or positive definite at the point $\mathbf{x}^*$.

Second-order sufficiency condition. If the matrix $\mathbf{H}(\mathbf{x}^*)$ is positive definite at the stationary point $\mathbf{x}^*$, then $\mathbf{x}^*$ is a local minimum point for the function $f(\mathbf{x})$.

Constrained Optimum Design Problem

The general design optimization model is to find a design variable vector –

$x = (x_1, x_2, \ldots, x_n)$

To minimize a cost function –

$f(x) = f(x_1; x_2; \ldots; x_n)$

subject to the equality constraints –

$h_i(x) = 0$; $i = 1$ to p

and the inequality constraints –

$g_i(x) \leq 0$; $i = 1$ to m

NECESSARY CONDITIONS: EQUALITY-CONSTRAINED PROBLEM

The optimality conditions for the constrained problem by including only the equality constraints in the formulation are now considered; that is, inequalities in Equation ($g_i(x) \leq 0$); are ignored. The reason is that the nature of equality constraints is quite different from that of inequality constraints. Equality constraints are always active for any feasible design, whereas an inequality constraint may not be active at a feasible point.

This changes the nature of the necessary conditions for the problem when inequalities are included.

Lagrange Multipliers

It turns out that a scalar multiplier is associated with each constraint, called the Lagrange multiplier. These multipliers play a prominent role in optimization theory as well as in numerical methods. Their values depend on the form of the cost and constraint functions. If these functions change, the values of the Lagrange multipliers also change. Here we will introduce the idea of Lagrange multipliers by considering a simple example problem. The example outlines the development of the Lagrange Multiplier theorem.

Before presenting the example problem, however, we consider an important concept of a regular point of the feasible set.

REGULAR POINT

Consider the constrained optimization problem of minimizing $f(x)$ subject to the constraints $h_i(x) = 0, i = 1$ to p.

A point x^* satisfying the constraints $h(x^*) = 0$ is said to be a regular point of the feasible set if $f(x^*)$ is differentiable and gradient vectors of all constraints at the point x^* are linearly independent.

Linear independence means that no two gradients are parallel to each other, and no gradient can be expressed as a linear combination of the others.

When inequality constraints are also included in the problem definition, then for a point to be regular, gradients of all of the active constraints must also be linearly independent.

Lagrange Multiplier Theorem

Consider the optimization problem –

Minimize $f(x)$

Subject to equality constraints –

$h_i(x) = 0$, $i = 1$ to p

Let x^* be a regular point that is a local minimum for the problem. Then there exist unique Lagrange multiplier v_j^*, $j = 1$ to p such that –

$$\frac{\partial f(x^*)}{\partial x_i} + \sum_{j=1}^{p} v_j^* \frac{\partial h_j(x^*)}{\partial x_i} = 0; \quad i = 1\ to\ n \qquad \text{(A)}$$

Hj(x*) = 0; j = 1 to p

It is convenient to write these conditions in terms of a Lagrange function, defined as –

$$L(x, v) = f(x) + \sum_{j=1}^{p} v_j h_j(x) \qquad \text{(B)}$$

$= f(x) + v^T h(x)$

Then Equation A becomes –

$$\nabla L(x^*,\ v^*) = 0,\ or\ \frac{\partial L(x^*, v^*)}{\partial x_i} = 0; i = 1\ to\ n \qquad \text{(C)}$$

Differentiating L(x, v) with respect to vj, we recover the equality constraints as –

$$\frac{\partial L(x^*, v^*)}{\partial v_j} = 0 \ \rightarrow\ h_j(x^*) = 0; j = 1\ to\ p \qquad \text{(D)}$$

The gradient conditions of Eqs. (C) and (D) show that the Lagrange function is stationary with respect to both x and v. Therefore, it may be treated as an unconstrained function in the variables x and v to determine the stationary points.

Karush-Kuhn-Tucker Necessary Conditions

The optimality design problem included inequality constraints of the form –

$g_i(x)$ 0; $i = 1$ to m

We can transform an inequality constraint to an equality by adding a new variable to it, called the *slack variable*.

Since the constraint is of the form "≤", its value is either negative or zero at a feasible point. Thus, the slack variable must always be non-negative (i.e., positive or zero) to make the inequality an equality.

An inequality constraint $g_i(x) \leq 0$ is equivalent to the equality constraint $g_i(x) + s_i = 0$, where $s_i \geq 0$ is a slack variable. The variables s_i are treated as unknowns of the design problem along with the original variables.

Their values are determined as a part of the solution. When the variable si has zero value, the corresponding inequality constraint is satisfied at equality. Such inequality is called an active (tight) constraint; that is, there is no "slack" in the constraint.

For any $s_i > 0$, the corresponding constraint is a strict inequality.

It is called an inactive constraint, and has slack given by s_i.

Thus, the status of an inequality constraint is determined as a part of the solution to the problem.

Note that with the preceding procedure, we must introduce one additional variable s_i and an additional constraint $s_i \geq 0$ to treat each inequality constraint. This increases the dimension of the design problem. The constraint $s_i \geq 0$ can be avoided if we use s_i^2 as the slack

variable instead of just si. Therefore, the inequality $g_i \leq 0$ is converted to an equality as –

$g_i + s_i^2 = 0$

Where s_i can have any real value. This form can be used in the Lagrange Multiplier

Theorem to treat inequality constraints and to derive the corresponding necessary conditions.

Karush-Kuhn-Tucker Optimality Conditions

Let x^* be a regular point of the feasible set that is a local minimum for $f(x)$, subject to $h_i(x) = 0$; $i = 1$ to p; $g_j(x) \leq 0$; $j = 1$ to m.

Then there exist Lagrange multipliers v^* (a p-vector) and u^* (an m-vector) such that the Lagrangian function is stationary with respect to x_j, v_i, u_j, and s_j at the point x^*.

1. Lagrangian Function for the Problem Written in the Standard Form:

$$L(x,\ v,\ u,\ s) = f(x) + \sum_{i=1}^{p} v_i h_i(x) + \sum_{j=1}^{m} u_j\,(g_j(x) + s_j^2)$$

$$= f(x) + \mathbf{v}^{\mathrm{T}}\mathbf{h}(\boldsymbol{x}) + \mathbf{u}^{\mathrm{T}}(\mathbf{g}(\boldsymbol{x}) + \boldsymbol{s}^2)$$

2. Gradient Conditions:

$$\frac{\partial L}{\partial x_k} = \frac{\partial f}{\partial x_k} + \sum_{i=1}^{p} v_i^* \frac{\partial h_i}{\partial x_i} + \sum_{j=1}^{m} u_j^* \frac{\partial g_j}{\partial x_k} = 0; k = 1\ to\ n$$

$$\frac{\partial L}{\partial v_i} = 0\ \rightarrow\ h_i\ (x^*) = 0; i = 1\ to\ \ p$$

$$\frac{\partial L}{\partial u_j} = 0\ \rightarrow \left(g_i(x^*) +\ s_j^2\right) = 0; j = 1\ to\ m$$

3. Feasibility Check for Inequalities:

$$s_j^2\ \geq 0; or\ equivalently\ g_j\ \leq 0; j = 1\ to\ m$$

4. Switching Conditions:

$$\frac{\partial L}{\partial s_j} = 0\ \rightarrow 2u_j^*\ s_j = 0; j = 1\ to\ m$$

5. Non-negativity of Lagrange Multipliers for Inequalities:

$$u_j^*\ \geq 0; j = 1\ to\ m$$

6. Regularity Check: Gradients of the active constraints must be linearly independent. In such a case the Lagrange multipliers for the constraints are unique.

Important Observations about KKT Conditions

The following important points should be noted relative to the KKT first-order necessary conditions for the problem written in the standard form.

1. The KKT conditions are not applicable at the points that are not regular. In those cases their use may yield candidate minimum points; however, the Lagrange multipliers may not be unique.

2. Any point that does not satisfy the KKT conditions cannot be a local minimum point unless it is an irregular point (in that case the KKT conditions are not applicable). Points satisfying the conditions are called KKT points.

3. The points satisfying the KKT conditions can be constrained or unconstrained. They are unconstrained when there are no equalities and all inequalities are inactive. If the candidate point is unconstrained, it can be a local minimum, maximum, or inflection point depending on the form of the Hessian matrix of the cost function

4. If there are equality constraints and no inequalities are active (i.e., u50), then the points satisfying the KKT conditions are only stationary. They can be minimum, maximum, or inflection points.

5. If some inequality constraints are active and their multipliers are positive, then the points satisfying the KKT conditions cannot be local maxima for the cost function (they may be local maximum points if active inequalities have zero multipliers). They may not be local minima either; this will depend on the second-order necessary and sufficient conditions.

6. It is important to note that the value of the Lagrange multiplier for each constraint depends on the functional form for the constraint.

Exercise:

1. Problems on Taylor's Series:

Write the Taylor's expansion for the following functions up to quadratic terms.

4.2 $\cos x$ about the point $x^* = \pi/4$
4.3 $\cos x$ about the point $x^* = \pi/3$
4.4 $\sin x$ about the point $x^* = \pi/6$
4.5 $\sin x$ about the point $x^* = \pi/4$
4.6 e^x about the point $x^* = 0$
4.7 e^x about the point $x^* = 2$

2. Nature of Quadratic Form:

Determine the nature of the following quadratic forms.

4.9 $F(\mathbf{x}) = x_1^2 + 4x_1x_2 + 2x_1x_3 - 7x_2^2 - 6x_2x_3 + 5x_3^2$
4.10 $F(\mathbf{x}) = 2x_1^2 + 2x_2^2 - 5x_1x_2$
4.11 $F(\mathbf{x}) = x_1^2 + x_2^2 + 3x_1x_2$
4.12 $F(\mathbf{x}) = 3x_1^2 + x_2^2 - x_1x_2$

3. Stationary Points:

Find stationary points for the following functions (use a numerical method or a software package like Excel, MATLAB, and Mathematica, if needed). Also determine the local minimum, local maximum, and inflection points for the functions (inflection points are those stationary points that are neither minimum nor maximum).

4.22 $f(x_1, x_2) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + 7$
4.23 $f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + 3$
4.24 $f(x_1, x_2) = x_1^3 + 12x_1x_2^2 + 2x_2^2 + 5x_1^2 + 3x_2$

4. Necessary Condition: Equality Constraints

Section 4.5 Necessary Conditions: Equality Constrained Problem

Find points satisfying the necessary conditions for the following problems; check if they are optimum points using the graphical method (if possible).

4.43 Minimize $f(x_1, x_2) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1$
subject to $x_1 + x_2 = 4$

4.44 Maximizte $f(x_1, x_2) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1$
subject to $x_1 + x_2 = 4$

4.45 Minimize $f(x_1, x_2) = (x_1 - 2)^2 + (x_2 + 1)^2$
subject to $2x_1 + 3x_2 - 4 = 0$

4.46 Minimize $f(x_1, x_2) = 4x_1^2 + 9x_2^2 + 6x_2 - 4x_1 + 13$
subject to $x_1 - 3x_2 + 3 = 0$

5. KKT Necessary Condition:

Find points satisfying KKT necessary conditions for the following problems; check if they are optimum points using the graphical method (if possible).

4.54 Maximize $F(x_1, x_2) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8$
subject to $x_1 + x_2 \le 4$

4.55 Minimize $f(x_1, x_2) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8$
subject to $x_1 + x_2 \le 4$

4.56 Maximize $F(x_1, x_2) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1$
subject to $x_1 + x_2 \le 4$

4.57 Minimize $f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$
subject to $x_1 + x_2 \ge 4$
$x_1 - x_2 - 2 = 0$