OBJECTIVE FUNCTION, MAXIMA AND MINIMA AND SADDLE POINTS, CONVEXITY AND CONCAVITY

INTRODUCTION

In this lecture we will study the optimization problem, its various components and its formulation as a mathematical programming problem. The stationary points of a function, the necessary and sufficient conditions for the relative maximum of a function of single or two variables are also discussed. The determination of the convexity or concavity of functions is also discussed.

BASIC COMPONENTS OF AN OPTIMIZATION PROBLEM

An **objective function** expresses the main aim of the model which is either to be minimized or maximized. For example, in a manufacturing process, the aim may be to *maximize the profit* or *minimize the cost.* In comparing the data predicted by a user-defined model with the observed data, the aim is *minimizing the total deviation* of the predictions based on the model from the observed data. In designing a bridge pier, the goal is to *maximize the strength* and *minimize size.*

A set of **unknowns** or **variables** control the value of the objective function. In the manufacturing problem, the variables may include the *amounts of different resources used* or the *time spent on each activity.* In fitting-the-data problem, the unknowns are the *parameters* of the model. In the pier design problem, the variables are the *shape and dimensions* of the pier.

A set of **constraints** are those which allow the unknowns to take on certain values but exclude others. In the manufacturing problem, one cannot spend negative amount of time on any activity, so one constraint is that the "time" variables are to be non-negative. In the pier design problem, one would probably want to limit the breadth of the base and to constrain its size.

The optimization problem is then to find values of the variables that minimize or maximize the objective function while satisfying the constraints.

Objective Function

As already stated, the objective function is the mathematical function the planner wants to maximize or minimize, subject to certain constraints. Many optimization problems have a single objective function. (When they don't they can often be reformulated so that they do) The two exceptions are:

- *No objective function.* In some cases (for example, design of integrated circuit layouts), the goal is to find a set of variables that satisfies the constraints of the model. The user does not particularly want to optimize anything and so there is no reason to define an objective function. This type of problem is usually called a *feasibility problem.*
- *Multiple objective functions.* In some cases, the user may like to optimize a number of different objectives concurrently. For instance, in the optimal design of panel of a door or window, it would be required to *minimize weight* and *maximize strength* simultaneously. Usually, the different objectives are not compatible; the variables that optimize one objective may be far from optimal for the other objectives. In practice, problems with multiple objectives are reformulated as single-objective problems by either forming a weighted combination of the different objectives or by treating some of the objectives as constraints.

Statement of an optimization problem

An optimization or a mathematical programming problem can be stated as follows:

$$\text{To find } \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} \text{ which minimizes } f(\mathbf{X}) \qquad (1)$$

Subject to the constraints

$$g_i(\mathbf{X}) \le 0, \qquad i = 1, 2, \ldots, m$$

$$l_j(\mathbf{X}) = 0, \qquad j = 1, 2, \ldots, p$$

where $\mathbf{X}$ is an n-dimensional vector called the design vector, $f(\mathbf{X})$ is called the *objective function*, and $g_i(\mathbf{X})$ and $l_j(\mathbf{X})$ are known as inequality and equality constraints, respectively. The number of variables n and the number of constraints m and p need not be related in any way. This type of problem is called a *constrained optimization problem*.

If the locus of all points satisfying $f(\mathbf{X})$ = a constant **c,** is considered, it can form a family of surfaces in the design space called the *objective function surfaces.* When drawn with the constraint surfaces as shown in Fig 1 we can identify the optimum point (maxima). This is possible graphically only when the number of design variables is two. When we have three or more design variables because of complexity in the objective function surface, we have to solve the problem as a mathematical problem and this visualization is not possible.

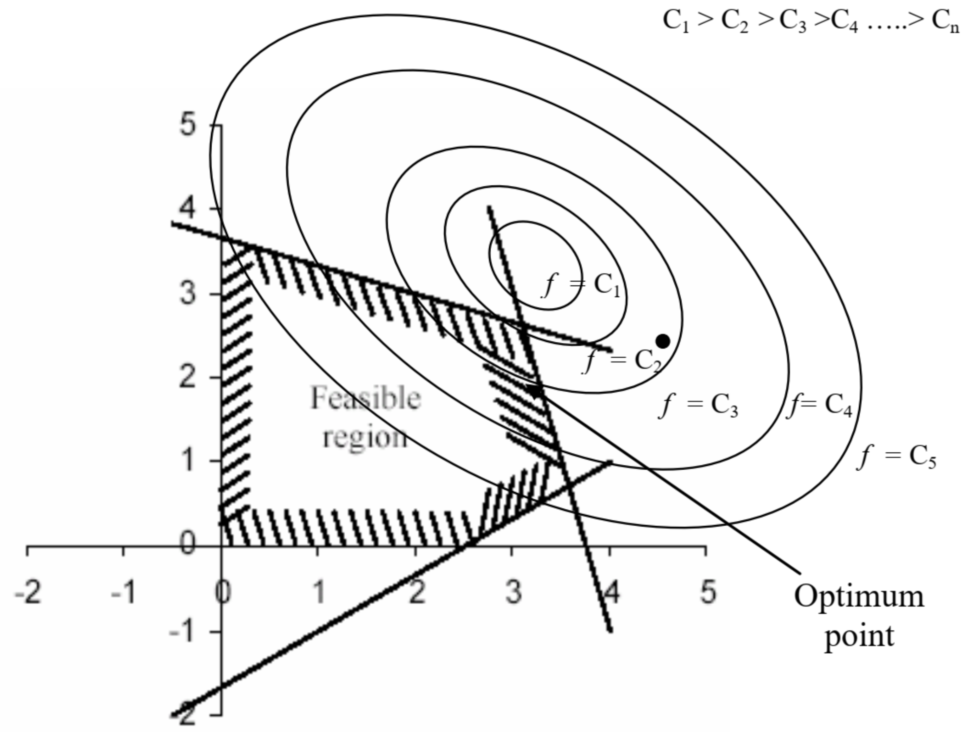

Fig 1. Objective function and constraint surfaces

Optimization problems can be defined without any constraints as well.

$$\text{To find } \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} \text{ which minimizes } \boldsymbol{f}(\mathbf{X}) \qquad (2)$$

Such problems are called *unconstrained optimization problems.* The field of unconstrained optimization is quite a large and prominent one, for which many algorithms and software are available.

Variables

These are essential. If there are no variables, we cannot define the objective function and the problem constraints. In many practical problems, one cannot choose the design variable arbitrarily. They have to satisfy certain specified functional and other requirements.

Constraints

Constraints are not essential. It's been argued that almost all problems really do have constraints. For example, any variable denoting the "number of objects" in a system can only be useful if it is less than the number of elementary particles in the known universe! In practice though, answers that make good sense in terms of the underlying physical or economic criteria can often be obtained without putting any constraints on the variables.

Design constraints are restrictions that must be satisfied to produce an acceptable design.

Constraints can be broadly classified as:

1) Behavioral or Functional constraints: These represent limitations on the behavior or performance of the system.
2) Geometric or Side constraints: These represent physical limitations on design variables such as availability, fabricability, and transportability.

Constraint Surfaces

Consider the optimization problem presented in eq. 1 with only the inequality constraint $g_i(\mathbf{X}) \leq 0$. The set of values of X that satisfy the equation $g_i(\mathbf{X}) \leq 0$ forms a boundary surface in the design space called a *constraint surface.* This will be a (n-1) dimensional subspace where n is the number of design variables. The constraint surface divides the design space into two regions: one with $g_i(\mathbf{X}) < 0$ (feasible region) and the other in which $g_i(\mathbf{X}) > 0$ (infeasible region). The points lying on the hyper surface will satisfy $g_i(\mathbf{X}) = 0$. The collection of all the constraint surfaces $g_i(\mathbf{X}) = 0$, $i = 1, 2, \ldots, m$, which separates the acceptable region is called the *composite constraint surface.*

Fig 2 shows a hypothetical two-dimensional design space where the feasible region is denoted by hatched lines. The two-dimensional design space is bounded by straight lines as shown in the figure. This is the case when the constraints are linear. However, constraints may be nonlinear as well and the design space will be bounded by curves in that case. A design point that lies on more than one constraint surface is called a *bound point,* and the associated constraint is called an active constraint. *Free points* are those that do not lie on any constraint surface. The design points that lie in the acceptable or unacceptable regions can be classified as following:

1. Free and acceptable point
2. Free and unacceptable point
3. Bound and acceptable point
4. Bound and unacceptable point.

Examples of each case are shown in Fig. 2.

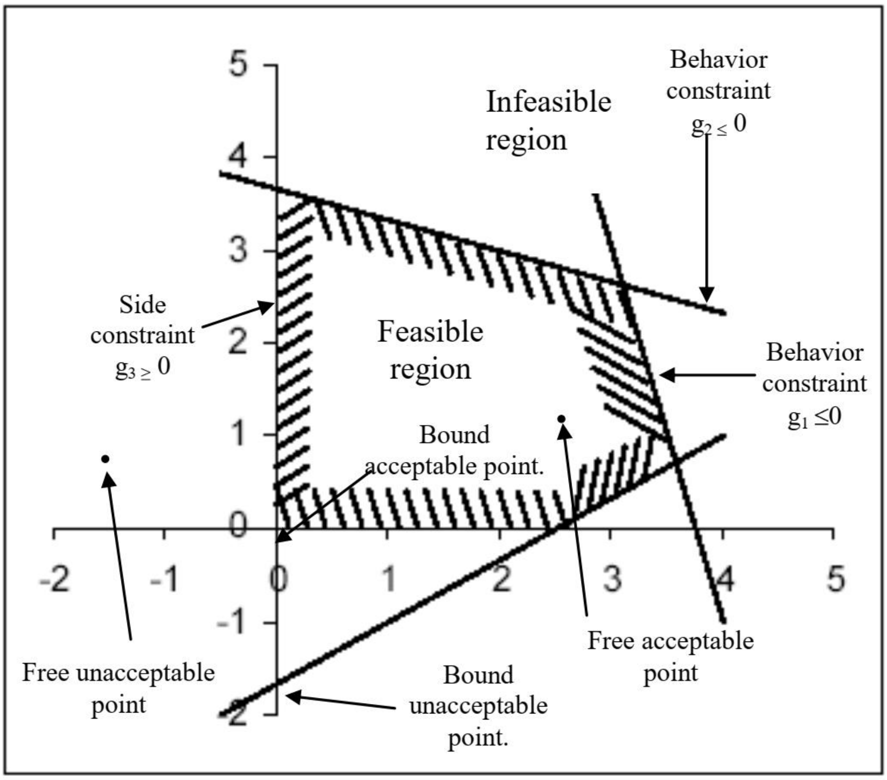

Fig. 2 Design space

<u>STATIONARY POINTS: FUNCTIONS OF SINGLE AND TWO VARIABLES</u>

For a continuous and differentiable function $f(x)$ a *stationary* point x^* is a point at which the slope of the function vanishes, i.e. $f'(x) = 0$ at $x = x^*$, where x^* belongs to its domain of definition.

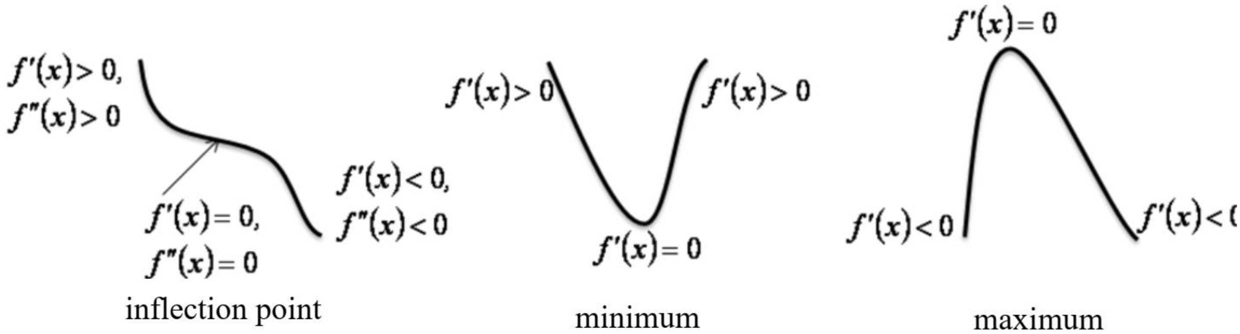

Fig. 3 Stationary points

A stationary point may be a minimum, maximum or an inflection point (Fig. 3).

<u>Relative and Global Optimum</u>

A function is said to have a *relative* or *local* minimum at $x = x^*$ if $f(x^*) \leq f(x^* + h)$ for all sufficiently small positive and negative values of *h,* i.e. in the near vicinity of the point x^*. Similarly a point x^* is called a *relative* or *local* maximum if $f(x^*) \geq f(x^* + h)$ for all values of h sufficiently close to zero. A function is said to have a *global* or *absolute* minimum at $x =$

x^* if $f(x^*) \le f(x)$ for all x in the domain over which $f(x)$ is defined. Similarly, a function is said to have a *global* or *absolute* maximum at $x = x^*$ if $f(x^*) \ge f(x)$ for all x in the domain over which $f(x)$ is defined.

Figure 4 shows the global and local optimum points.

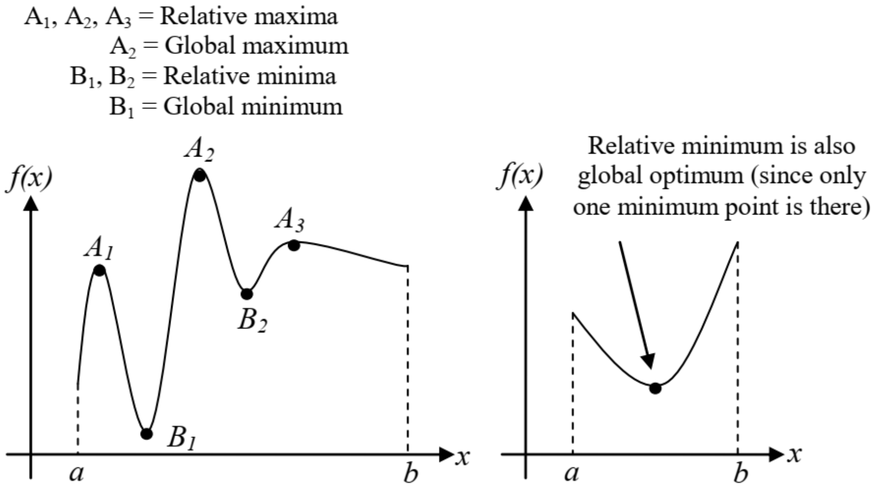

Fig. 4 Global and local optimum points

Functions of a single variable

Consider the function $f(x)$ defined for $a \le x \le b$. To find the value of $x^* \in [a,b]$ such that $x = x^*$ maximizes $f(x)$ we need to solve a *single-variable optimization* problem. We have the following theorems to understand the necessary and sufficient conditions for the relative maximum of a function of a single variable.

Necessary condition: For a single variable function $f(x)$ defined for $x \in [a,b]$ which has a relative maximum at $x = x^*$, $x^* \in [a,b]$ if the derivative $f'(X) = df(x)/dx$ exists as a finite number at $x = x^*$ then $f'(x^*) = 0$. This can be understood from the following.

Proof.

Since $f'(x^*)$ is stated to exist, we have

$$f'(x^*) = \lim_{h\to 0} \frac{f(x^*)+h - f(x^*)}{h} \tag{3}$$

From our earlier discussion on relative maxima we have $f(x^*) \ge f(x^*+h)$ for $h \to 0$. Hence

$$\frac{f(x^*+h) - f(x^*)}{h} \ge 0 \qquad h < 0 \tag{4}$$

$$\frac{f(x^*+h) - f(x^*)}{h} \le 0 \qquad h > 0 \tag{5}$$

which implies that for substantially small negative values of h we have $f(x^*) \geq 0$ and for substantially small positive values of h we have $f(x^*) \leq 0$. In order to satisfy both (4) and (5), $f(x^*)$ = 0. Hence this gives the necessary condition for a relative maxima at $x = x^*$ for $f(x)$.

It has to be kept in mind that the above theorem holds good for relative minimum as well. The theorem only considers a domain where the function is continuous and differentiable. It cannot indicate whether a maxima or minima exists at a point where the derivative fails to exist. This scenario is shown in Fig 5, where the slopes m_1 and m_2 at the point of a maxima are unequal, hence cannot be found as depicted by the theorem by failing for continuity. The theorem also does not consider if the maxima or minima occurs at the end point of the interval of definition, owing to the same reason that the function is not continuous, therefore not differentiable at the boundaries. The theorem does not say whether the function will have a maximum or minimum at every point where $f'(x) = 0$, since this condition $f'(x) = 0$ is for stationary points which include inflection points which do not mean a maxima or a minima. A point of inflection is shown already in Fig.3

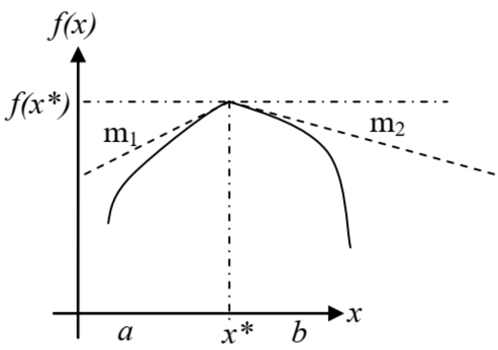

Fig. 5 Point of inflection

Sufficient condition: For the same function stated above let $f'(x^*) = f''(x^*) = \ldots = f^{(n-1)}(x^*) = 0$, but $f^{(n)}(x^*) \neq 0$, then it can be said that $f(x^*)$ is (a) a minimum value of $f(x)$ if $f^{(n)}(x^*) > 0$ and n is even; (b) a maximum value of $f(x)$ if $f^{(n)}(x^*) < 0$ and n is even; (c) neither a maximum nor a minimum if n is odd.

Proof

Applying the Taylor's theorem with remainder after n terms, we have

$$f(x^*+h) = f(x^*) + hf'(x^*) + \frac{h^2}{2!}f''(x^*) + \ldots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x^*) + \frac{h^n}{n!}f^n(x^*+\theta h) \qquad (6)$$

for $0<\theta<1$

since $f'(x^*) = f(x^*) = \ldots = f^{n-1}(x^*) = 0$, (6) becomes

$$f(x^*+h)-f(x^*)=\frac{h^n}{n!}f^n(x^*+\theta h) \qquad (7)$$

As $f^{(n)}(x^*) \neq 0$, there exists an interval around x^* for every point x of which the *n*th derivative $f^{(n)}(x)$ has the same sign, viz., that of $f^{(n)}(x^*)$. Thus for every point $(x^* + h)$ of this interval, $f^{(n)}(x^* + h)$ has the sign of $f^{(n)}(x^*)$. When *n* is even $\frac{h^n}{n!}$ is positive irrespective of the sign of *h,* and hence $f(x^*+h)-f(x^*)$ will have the same sign as that of $f^{(n)}(x^*)$. Thus x^* will be a relative minimum if $f^{(n)}(x^*)$ is positive, with $f(x)$ convex around x^*, and a relative maximum if $f^{(n)}(x^*)$ is negative, with $f(x)$ concave around x^*. When *n* is odd, $\frac{h^n}{n!}$ changes sign with the change in the sign of *h* and hence the point x^* is neither a maximum nor a minimum. In this case the point x^* is called a *point of inflection*.

Example 1.

Find the optimum value of the function $f(x)=x^2+3x-5$ and also state if the function attains a maximum or a minimum.

Solution

$f'(x)=2x+3=0$ for maxima or minima.

or $x^* = -3/2$

$f''(x^*)=2$ which is positive hence the point $x^* = -3/2$ is a point of minima and the function attains a minimum value of -29/4 at this point.

Example 2.

Find the optimum value of the function $f(x)=(x-2)^4$ and also state if the function attains a maximum or a minimum.

Solution

$f'(x)=4(x-2)^3=0$ for maxima or minima.

or $x=x^* = 2$ for maxima or minima.

$f''(x^*)=12(x^*-2)^2=0$ at $x^* = 2$

$f'''(x^*)=24(x^*-2)=0$ at $x^* = 2$

$f''''(x^*) = 24$ at $x^* = 2$

Hence $f^n(x)$ is positive and n is even hence the point $x = x^* = 2$ is a point of minimum and the function attains a minimum value of 0 at this point.

Example 3.

Analyze the function $f(x)=12x^5-45x^4+40x^3+5$ and classify the stationary points as maxima, minima and points of inflection.

Solution

$$f'(x)=60x^4-180x^3+120x^2=0$$
$$=> x^4-3x^3+2x^2=0$$
$$\text{or} \quad x=0,1,2$$

Consider the point $x = x^* = 0$

$f''(x^*)=240(x^*)^3-540(x^*)^2+240x^*=0$ at $x^*=0$

$f'''(x^*)=720(x^*)^2-1080x^*+240=240$ at $x^*=0$

Since the third derivative is non-zero, $x = x^* = 0$ is neither a point of maximum or minimum but it is a point of inflection.

Consider $x = x^* = 1$

$f''(x^*)=240(x^*)^3-540(x^*)^2+240x^*=-60$ at $x^*=1$

Since the second derivative is negative the point $x = x^* = 1$ is a point of local maxima with a maximum value of $f(x) = 12 – 45 + 40 + 5 = 12$

Consider $x = x^* = 2$

$f''(x^*)=240(x^*)^3-540(x^*)^2+240x^*=240$ at $x^*=2$

Since the second derivative is positive, the point $x = x^* = 2$ is a point of local minima with a minimum value of $f(x) = -11$

Example 4.

The horse power generated by a Pelton wheel is proportional to *u(v-u)* where *u* is the velocity of the wheel, which is variable and *v* is the velocity of the jet which is fixed. Show that the efficiency of the Pelton wheel will be maximum at $u = v/2$.

Solution

$$f=\mathrm{K}.u(v-u)$$
$$\frac{\partial f}{\partial u}=0 => \mathrm{K}v-2\mathrm{K}u=0 \quad \text{or} \quad u=\frac{v}{2}$$

where K is a proportionality constant (assumed positive).

$$\left.\frac{\partial^2 f}{\partial u^2}\right|_{u=\frac{v}{2}}=-2\mathrm{K} \text{ which is negative.}$$

Hence, f is maximum at $u=\frac{v}{2}$

Functions of two variables

This concept may be easily extended to functions of multiple variables. Functions of two variables are best illustrated by contour maps, analogous to geographical maps. A contour is a line representing a constant value of $f(x)$ as shown in Fig.4. From this we can identify *maxima, minima* and *points of inflection.*

Necessary conditions

As can be seen in Fig. 6 and 7, perturbations from points of local minima in any direction result in an increase in the response function $f(x)$, i.e. the slope of the function is zero at this point of local minima. Similarly, at *maxima* and *points of inflection* as the slope is zero, the first derivatives of the function with respect to the variables are zero.

Which gives us $\frac{\partial f}{\partial x_1} = 0; \frac{\partial f}{\partial x_2} = 0$ at the stationary points, i.e., the gradient vector of $f(\mathbf{X})$, $\Delta_x f$ at $\mathbf{X} = \mathbf{X}^* = [x_1, x_2]$ defined as follows, must equal zero:

$$\Delta_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(X^*) \\ \frac{\partial f}{\partial x_2}(X^*) \end{bmatrix} = 0 \tag{8}$$

This is the necessary condition.

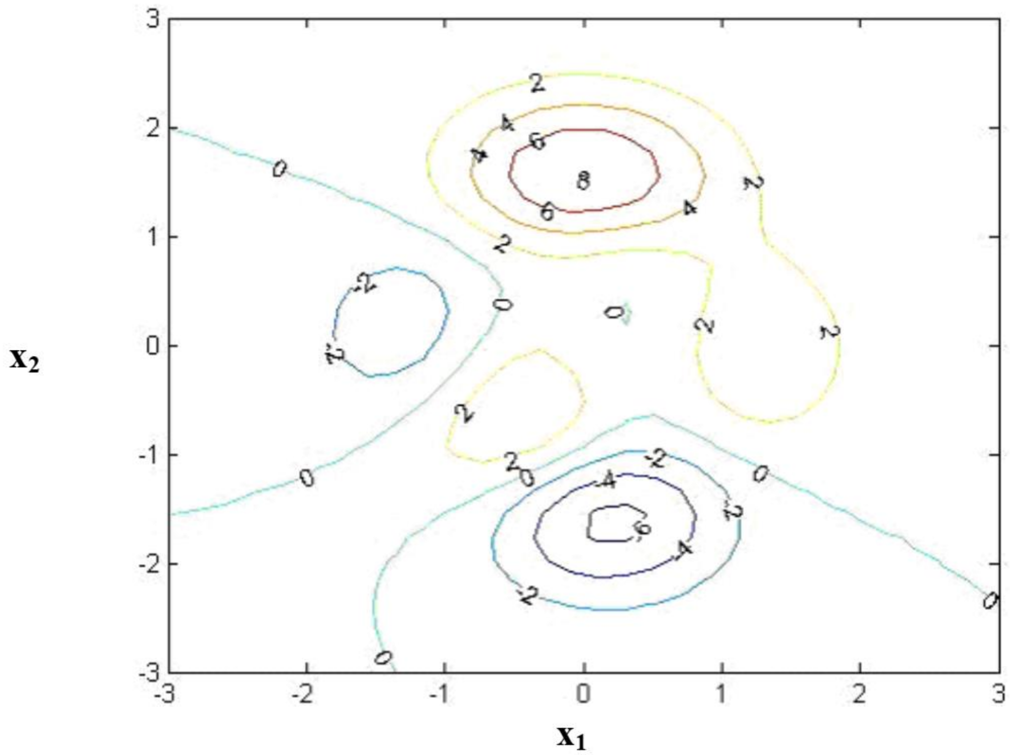

Fig. 6 Maxima, minima and point of inflection

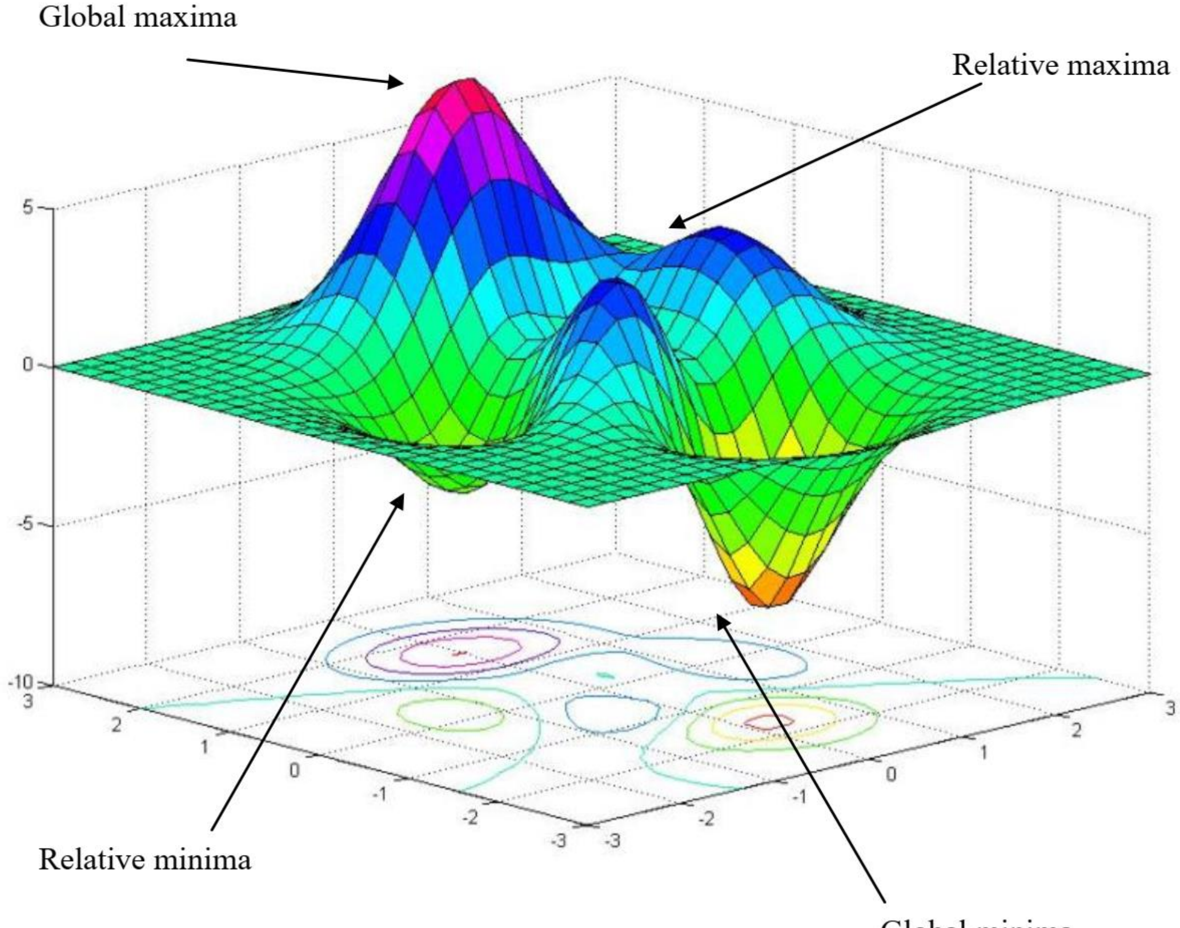

Fig. 7 Global and relative optimal points

Sufficient conditions

Consider the following second order derivatives:

$$\frac{\partial^2 f}{\partial x_1^2};\frac{\partial^2 f}{\partial x_2^2};\frac{\partial^2 f}{\partial x_1 \partial x_2} \tag{9}$$

The Hessian matrix defined by **H** is formed using the above second order derivatives.

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}_{[x_1, x_2]} \tag{10}$$

If **H** is positive definite then the point $\mathbf{X} = [x_1, x_2]$ is a point of local minima.

If **H** is negative definite then the point $\mathbf{X} = [x_1, x_2]$ is a point of local maxima.

If **H** is neither then the point $\mathbf{X} = [x_1, x_2]$ is neither a point of maxima nor minima.

A square matrix is positive definite if all its eigen values are positive and it is negative definite if all its eigen values are negative. If some of the eigen values are positive and some negative then the matrix is neither positive definite or negative definite.

To calculate the eigen values λ of a square matrix then the following equation is solved.

$$|\mathbf{A}-\lambda\mathbf{I}|=0 \tag{11}$$

The above rules give the sufficient conditions for the optimization problem of two variables.

Example 5.

Locate the stationary points of $f(\mathbf{X})$ and classify them as relative maxima, relative minima or neither based on the rules discussed in the lecture.

$f(\mathbf{X}) = 2x_1^3/3 - 2x_1x_2 - 5x_1 + 2x_2^2 + 4x_2 + 5$

Solution

$$\Delta_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(X^*) \\ \frac{\partial f}{\partial x_2}(X^*) \end{bmatrix} = \begin{bmatrix} 2x_1^2 - 2x_2 - 5 \\ -2x_1 + 4x_2 + 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

From $\frac{\partial f}{\partial x_2}(X) = 0$, $x_1 = 2x_2 + 2$

From $\frac{\partial f}{\partial x_1}(X) = 0$

$$2(2x_2+2)^2 - 2x_2 - 5 = 0$$

$$8x_2^2 + 14x_2 + 3 = 0$$

$$(2x_2+3)(4x_2+1) = 0$$

$$x_2 = -3/2 \quad \text{or} \quad x_2 = -1/4$$

so the two stationary points are

$$X_1 = [-1,-3/2]$$

and

$$X_2 = [3/2,-1/4]$$

The Hessian of $f(\mathbf{X})$ is

$$\frac{\partial^2 f}{\partial x_1^2} = 4x_1; \frac{\partial^2 f}{\partial x_2^2} = 4; \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -2$$

$$\mathbf{H} = \begin{bmatrix} 4x_1 & -2 \\ -2 & 4 \end{bmatrix}$$

$$|\lambda\mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda - 4x_1 & 2 \\ 2 & \lambda - 4 \end{vmatrix}$$

At $X_1 = [-1,-3/2]$,

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda+4 & 2 \\ 2 & \lambda-4 \end{vmatrix} = (\lambda+4)(\lambda-4) - 4 = 0$$

$$\lambda^2 - 16 - 4 = 0$$

$$\lambda^2 = 12$$

$$\lambda_1 = +\sqrt{12} \qquad \lambda_2 = -\sqrt{12}$$

Since one eigen value is positive and one negative, X_1 is neither a relative maximum nor a relative minimum.

At X_2 = [3/2,-1/4]

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda-6 & 2 \\ 2 & \lambda-4 \end{vmatrix} = (\lambda-6)(\lambda-4) - 4 = 0$$

$$\lambda^2 - 10\lambda + 20 = 0$$

$$\lambda_1 = 5 + \sqrt{5} \quad \lambda_2 = 5 - \sqrt{5}$$

Since both the eigen values are positive, X_2 is a local minimum.

Minimum value of f(x) is -0.375.

Example 6

The ultimate strength attained by concrete is found to be based on a certain empirical relationship between the ratios of cement and concrete used. Our objective is to maximize strength attained by hardened concrete, given by $f(\mathbf{X}) = 20 + 2x_1 - x_1^2 + 6x_2 - 3x_2^2/2$, where x_1 and x_2 are variables based on cement and concrete ratios.

Solution

Given $f(\mathbf{X}) = 20 + 2x_1 - x_1^2 + 6x_2 - 3x_2^2/2$; where $\mathbf{X} = x_1, x_2$

The gradient vector $\Delta_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(X^*) \\ \frac{\partial f}{\partial x_2}(X^*) \end{bmatrix} = \begin{bmatrix} 2-2x_1 \\ 6-3x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, to determine stationary point $\mathbf{X}^*$.

Solving we get $\mathbf{X}^*$ = [1,2]

$$\frac{\partial^2 f}{\partial x_1^2} = -2; \frac{\partial^2 f}{\partial x_2^2} = -3; \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

$$\mathbf{H} = \begin{bmatrix} -2 & 0 \\ 0 & -3 \end{bmatrix}$$

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda+2 & 0 \\ 0 & \lambda+3 \end{vmatrix} = (\lambda+2)(\lambda+3) = 0$$

Here the values of λ do not depend on **X** and λ_1 = -2, λ_2 = -3. Since both the eigen values are negative, *f*(**X**) is concave and the required ratio $x_1{:}x_2$ = 1:2 with a global maximum strength of *f*(**X**) = 27 units.

CONVEXITY AND CONCAVITY OF FUNCTIONS OF ONE AND TWO VARIABLES

The analyst must determine whether the objective functions and constraint equations are convex or concave. In real-world problems, if the objective function or the constraints are not convex nor concave, the problem is usually mathematically intractable.

Functions of one variable

Convex function

A real-valued function *f* defined on an interval (or on any convex subset *C* of some vector space) is called **convex**, if for any two points *a* and *b* in its domain *C* and any *t* in [0,1], we have

$$f(ta+(1-t)b) \leq tf(a)+(1-t)f(b) \tag{12}$$

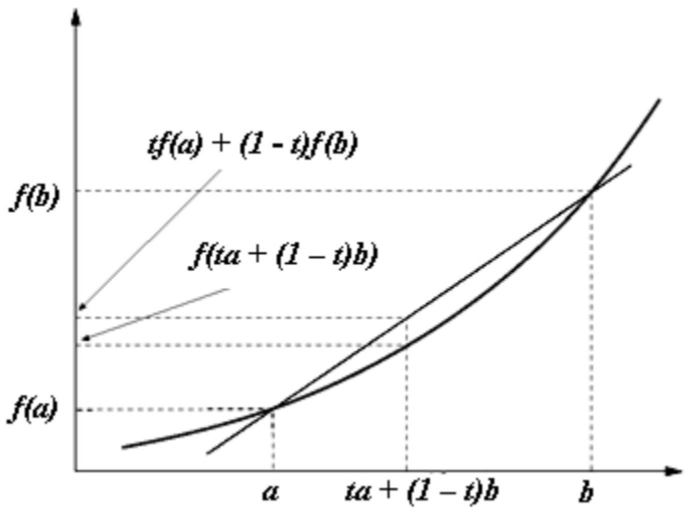

Fig. 8 Convex function

In other words, a function is convex if and only if its epigraph (the set of points lying on or above the graph) is a convex set. A function is also said to be *strictly convex* if

$$f(ta+(1-t)b) < tf(a)+(1-t)f(b) \tag{13}$$

for any *t* in (0,1) and a line connecting any two points on the function lies completely above the function. These relationships are illustrated in Fig. 8.

Testing for convexity of a single variable function

A function is convex if its slope is not decreasing or $\partial^2 f/\partial x^2 \geq 0$. It is strictly convex if its slope is continually increasing or $\partial^2 f/\partial x^2 > 0$ throughout the function.

Properties of convex functions

A convex function *f*, defined on some convex open interval *C*, is continuous on *C* and differentiable at all or at most or at many countable points. If *C* is closed, then *f* may fail to be continuous at the end points of *C*.

A continuous function on an interval *C* is convex if and only if

$$f\left(\frac{a+b}{2}\right)\leq\frac{f(a)+f(b)}{2} \tag{14}$$

for all *a* and *b* in *C*.

A differentiable function of one variable is convex on an interval if and only if its derivative is monotonically non-decreasing on that interval.

A continuously differentiable function of one variable is convex on an interval if and only if the function lies above all of its tangents: $f(b)\geq f(a)+f'(a)(b-a)$ for all a and b in the interval.

A twice differentiable function of one variable is convex on an interval if and only if its second derivative is non-negative in that interval; this gives a practical test for convexity. If its second derivative is positive then it is strictly convex, but the converse does not hold, as shown by $f(x) = x^4$.

More generally, a continuous, twice differentiable function of several variables is convex on a convex set if and only if its Hessian matrix is positive semi definite on the interior of the convex set.

If two functions *f* and *g* are convex, then so is any weighted combination $a f + b g$ with non-negative coefficients *a* and *b*. Likewise, if *f* and *g* are convex, then the function max{*f*,*g*} is convex.

A *strictly convex* function will have only one minimum which is also the global minimum.

Examples

- The second derivative of x^2 is 2; it follows that x^2 is a convex function of *x*.
- The absolute value function $|x|$ is convex, even though it does not have a derivative at $x = 0$.
- The function *f* with domain [0,1] defined by $f(0)=f(1)=1$, $f(x)=0$ for $0<x<1$ is convex; it is continuous on the open interval (0,1), but not continuous at 0 and 1.
- Every linear transformation is convex but not strictly convex, since if *f* is linear, then $f(a + b) = f(a) + f(b)$. This implies that the identity map (i.e., $f(x) = x$) is convex but not strictly convex. The fact holds good if we replace "convex" by "concave".

- An affine function ($f(x) = ax + b$) is simultaneously convex and concave.

Concave function

A differentiable function f is **concave** on an interval if its derivative function f' is decreasing on that interval: a concave function has a decreasing slope.

A function that is convex is often synonymously called **concave upwards**, and a function that is concave is often synonymously called **concave downward**.

For a twice-differentiable function f, if the second derivative, $f''(x)$, is positive (or, if the acceleration is positive), then the graph is convex (or concave upward); if the second derivative is negative, then the graph is concave (or concave downward). Points, at which concavity changes, are called inflection points.

If a convex (i.e., concave upward) function has a "bottom", any point at the bottom is a minimal extremum. If a concave (i.e., concave downward) function has an "apex", any point at the apex is a maximal extremum.

A function $f(x)$ is said to be **concave** on an interval if, for all a and b in that interval,

$$\forall t \in [0,1], f(ta+(1-t)b) \geq tf(a)+(1-t)f(b) \tag{15}$$

Additionally, $f(x)$ is **strictly concave** if

$$\forall t \in [0,1], f(ta+(1-t)b) > tf(a)+(1-t)f(b) \tag{16}$$

These relationships are illustrated in Fig. 9

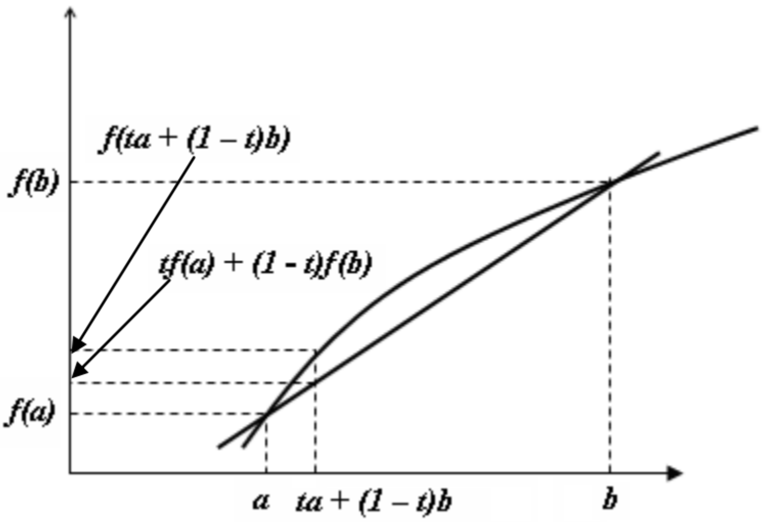

Fig. 9 Concave function

Testing for concavity of a single variable function

A function is concave if its slope is not increasing or $\partial^2 f/\partial x^2 \leq 0$. It is strictly concave if its slope is continually decreasing or $\partial^2 f/\partial x^2 < 0$ throughout the function.

Properties of a concave functions

A continuous function on C is concave if and only if

$$f\left(\frac{a+b}{2}\right) \geq \frac{f(a)+f(b)}{2} \quad (17)$$

for any x and y in C.

Equivalently, $f(x)$ is concave on $[a, b]$ if and only if the function $-f(x)$ is convex on every subinterval of $[a, b]$.

If $f(x)$ is twice-differentiable, then $f(x)$ is concave if and only if $f''(x)$ is non-positive. If its second derivative is negative then it is strictly concave, but the opposite is not true, as shown by $f(x) = -x^4$.

A function is called **quasiconcave** if and only if there is an x_0 such that for all $x < x_0$, $f(x)$ is non-decreasing while for all $x > x_0$ it is non-increasing. x_0 can also be $\pm\infty$, making the function non-decreasing (non-increasing) for all x. The opposite of quasiconcave is **quasiconvex**.

Example 7

Consider the example 3 for a function of two variables. Locate the stationary points of $f(x) = 12x^5 - 45x^4 + 40x^3 + 5$ and find out if the function is convex, concave or neither at the points of optima based on the testing rules discussed above.

Solution

$$f'(x) = 60x^4 - 180x^3 + 120x^2 = 0$$

$$=> x^4 - 3x^3 + 2x^2 = 0$$

or $\quad x = 0, 1, 2$

Consider the point $x = x^* = 0$

$$f''(x^*) = 240(x^*)^3 - 540(x^*)^2 + 240x^* = 0 \text{ at } x^* = 0$$

$$f'''(x^*) = 720(x^*)^2 - 1080x^* + 240 = 240 \text{ at } x^* = 0$$

Since the third derivative is non-zero $x = x^* = 0$ is neither a point of maximum or minimum but it is a point of inflection. Hence the function is neither convex nor concave at this point.

Consider $x = x^* = 1$

$$f''(x^*) = 240(x^*)^3 - 540(x^*)^2 + 240x^* = -60 \text{ at } x^* = 1$$

Since the second derivative is negative, the point $x = x^* = 1$ is a point of local maxima with a maximum value of $f(x) = 12 - 45 + 40 + 5 = 12$. At this point the function is concave since $\partial^2 f / \partial x^2 < 0$.

Consider $x = x^* = 2$

$$f''(x^*) = 240(x^*)^3 - 540(x^*)^2 + 240x^* = 240 \text{ at } x^* = 2$$

Since the second derivative is positive, the point $x = x^* = 2$ is a point of local minima with a minimum value of $f(x)$ = -11. At this point the function is convex since $\partial^2 f / \partial x^2 > 0$.

Functions of two variables

A function of two variables, $f(\mathbf{X})$ where X is a vector = [x_1,x_2], is strictly convex if

$$f(tX_1 + (1-t)X_2) < tf(X_1) + (1-t)f(X_2) \tag{18}$$

where $\mathbf{X_1}$ and $\mathbf{X_2}$ are points located by the coordinates given in their respective vectors. Similarly a two variable function is strictly concave if

$$f(tX_1 + (1-t)X_2) > tf(X_1) + (1-t)f(X_2) \tag{19}$$

Contour plot of a convex function is illustrated in Fig. 10.

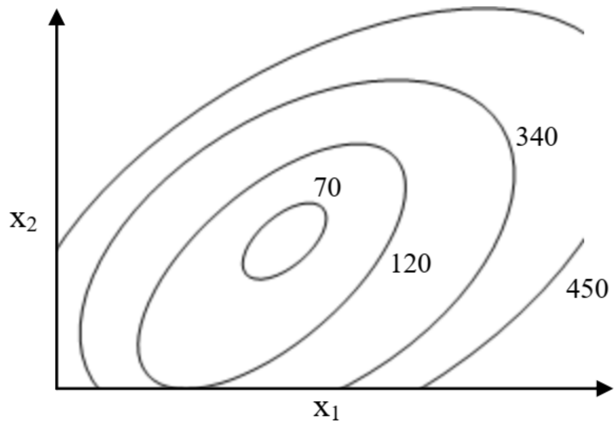

Fig. 10 Contour plot of a convex function

Contour plot of a concave function is shown in Fig. 11.

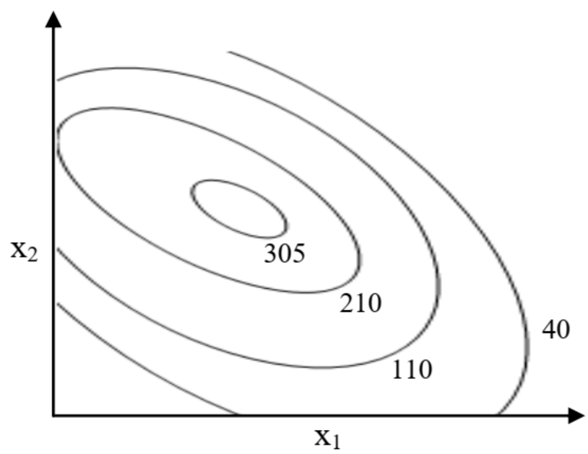

Fig. 11 Contour plot of a concave function

To determine convexity or concavity of a function of multiple variables, the eigen values of its Hessian matrix are examined and the following rules apply.

(a) If all eigen values of the Hessian matrix are positive the function is strictly convex.

(b) If all eigen values of the Hessian matrix are negative the function is strictly concave.

(c) If some eigen values are positive and some are negative, or if some are zero, the function is neither strictly concave nor strictly convex.

Example 8

Consider the example 5 for a function of two variables. Locate the stationary points of *f*(**X**) and find out if the function is convex, concave or neither at the points of optima based on the rules discussed in this lecture.

$f(\mathbf{X}) = 2x_1^3/3 - 2x_1x_2 - 5x_1 + 2x_2^2 + 4x_2 + 5$

Solution

$$\Delta_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(X^*) \\ \frac{\partial f}{\partial x_2}(X^*) \end{bmatrix} = \begin{bmatrix} 2x_1^2 - 2x_2 - 5 \\ -2x_1 + 4x_2 + 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving the above the two stationary points are

$$X_1 = [-1,-3/2]$$

and

$$X_2 = [3/2,-1/4]$$

The Hessian of *f*(**X**) is

$$\frac{\partial^2 f}{\partial x_1^2} = 4x_1; \frac{\partial^2 f}{\partial x_2^2} = 4; \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -2$$

$$\mathbf{H} = \begin{bmatrix} 4x_1 & -2 \\ -2 & 4 \end{bmatrix}$$

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda - 4x_1 & 2 \\ 2 & \lambda - 4 \end{vmatrix}$$

At X_1

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda + 4 & 2 \\ 2 & \lambda - 4 \end{vmatrix} = (\lambda + 4)(\lambda - 4) - 4 = 0$$

$$\lambda^2 - 16 - 4 = 0$$

$$\lambda^2 = 12$$

$$\lambda_1 = +\sqrt{12} \qquad \lambda_2 = -\sqrt{12}$$

Since one eigen value is positive and one negative, X_1 is neither a relative maximum nor a relative minimum. Hence at X_1 the function is neither convex nor concave.

At X_2 = [3/2,-1/4]

$$|\lambda \mathbf{I} - \mathbf{H}| = \begin{vmatrix} \lambda - 6 & 2 \\ 2 & \lambda - 4 \end{vmatrix} = (\lambda - 6)(\lambda - 4) - 4 = 0$$

$$\lambda^2 - 10\lambda + 20 = 0$$

$$\lambda_1 = 5 + \sqrt{5} \quad \lambda_2 = 5 - \sqrt{5}$$

Since both the eigen values are positive, X_2 is a local minimum, and the function is convex at this point as both the eigen values are positive.

BIBLIOGRAPHY / FURTHER READING:

1. Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, N.J, 1957.
2. Hillier F.S. and G.J. Lieberman, *Operations Research*, CBS Publishers & Distributors, New Delhi, 1987.
3. Loucks, D.P., J.R. Stedinger, and D.A. Haith, *Water Resources Systems Planning and Analysis*, Prentice-Hall, N.J., 1981.
4. Rao S.S., *Engineering Optimization – Theory and Practice*, Fourth Edition, John Wiley and Sons, 2009.
5. Taha H.A., *Operations Research – An Introduction*, 8th edition, Pearson Education India, 2008.
6. Vedula S., and P.P. Mujumdar, *Water Resources Systems: Modelling Techniques and Analysis*, Tata McGraw Hill, New Delhi, 2005.