Chapter 2 General Unconstrained Optimization Problems

1- Introduction

In this section we address the problem of maximizing (minimizing) a function in the case when there are no constraints on its arguments (design variables). This is not a very interesting case for real problems (economic, industrial, etc.), which typically deals with problems where resources are constrained, but represents a **natural starting point** to solving the more economically relevant constrained optimization problems.

2- Unconstrained Optimization Problem statement and preliminaries

The unconstrained optimization problem is generally stated as follows:

$$\min_{\mathrm{x}\in\mathbb{R}^n} f(\mathrm{x})$$

Where: $f: \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable function.

This notation means that we wish to find the vector $\mathbf{x}$ that minimizes the function $f(\mathrm{x})$. Regarding this statement and definition of an unconstrained optimization problem, we distinguish two cases:

- **Case of:** $\boldsymbol{n} = \mathbf{1}$, which corresponds to the single variable unconstrained optimization problem.
- **Case of:** $\boldsymbol{n} > 1$, which corresponds to the multivariable unconstrained optimization problem

2.1 Single Variable Unconstrained Optimization Problem

2.1.1 Definition

A single variable **optimization problem** is one in which the value of $x = x^*$ is to be found in the interval $[a, b]$ over which the function is defined such that:

$$x^* \; Minimizes \; f(x)$$

In other words, we are interested in finding **minima** (or **maxima**) of this one dimensional function. Before we do so, we need to start with defining what we mean by these concepts.

2.1.2 Local (relative) Min (Max)

- A function of one variable denoted by $f(x), x = scalar$ is said to have a **local** (relative) **minimum** at $x = x^*$ if:

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

For all sufficiently small (close to zero) positive and negative values of h.

If:

$$f(x^*) < f(x^* + h)$$

For all sufficiently small (close to zero) positive and negative values of h.

we say that the local minimum x^* is **strict**.

- Similarly, a single variable function $f(x), x = scalar$ is said to have a **local** (relative) **maximum** at $x = x^*$ if:

$$f(x^*) \geq f(x^* + h)$$

For all sufficiently small (close to zero) positive and negative values of h.

If:

$$f(x^*) > f(x^* + h)$$

For all sufficiently small (close to zero) positive and negative values of h.

we say that the local maximum x^* is **strict**.

Result:

Clearly a function can have many or no local maxima in its domain of definition.

2.1.3 Global (Absolute) Min (Max)

A function of one variable denoted by $f(x), x = scalar$ is said to have a **Global** (Absolute) **Minimum** at $x = x^*$ if:

$$f(x^*) \le f(x), \quad \forall\, x$$

That is for all values of x and not just for all x close to x^*, in the domain over which the function $f(x)$ is defined.

Similarly, a point x^* will be a Global (Absolute) Maximum for the function $f(x)$ if:

$$f(x^*) \ge f(x), \qquad \forall\, x$$

For all values of x in the domain over which the function is defined.

In the following figure it is shown the difference between local and global optimum points.

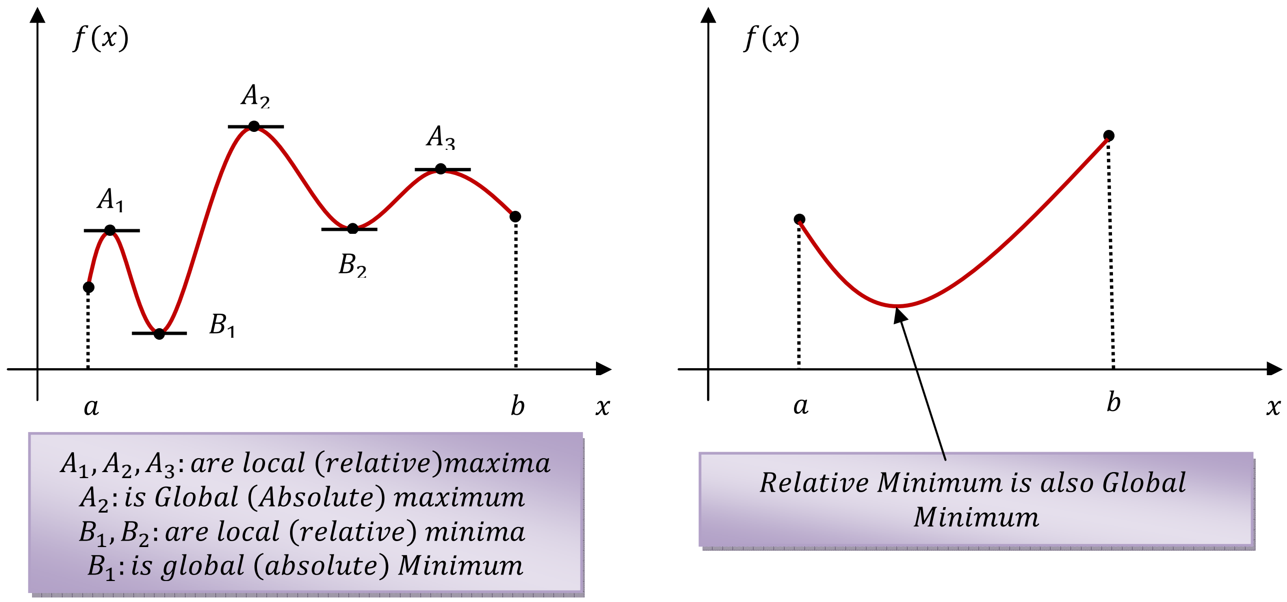

Fig.1 Local and Global Minima

2.1.4 Optimality Conditions

The question is how do we go about finding local (global) maxima (minima)? This is what we call optimality conditions. Most of the time we use differentiation to define the first order and second order optimality conditions, where we set the first derivative to zero but a zero first derivative is neither necessary (e.g., corner maximum; kink maximum), nor sufficient condition for maximum ($f' = 0$ could also correspond to a minimum or other critical point). Thus some care is needed to ensure that what one finds by setting $f' = 0$ is indeed what one is looking for.

Let us call both local maximum and local minimum local Extremum. We will derive these two conditions but before doing this, the following definitions are preliminaries in the context of optimization theory. The following two theorems provide the necessary and sufficient conditions for the local (relative) Extremum of **unconstrained optimization** problems for the univariate function.

2.1.4.1 Theorem 2.1 (First order Necessary Condition of Optimality)

If the single variable function $f(x)$ is defined in the interval $a \le x \le b$ and has a relative (local) minimum at the point $x = x^*$, where $a \le x^* \le b$, and if the first derivative $f'(x) = \frac{d}{dx}f(x)$ exists as a finite number at $x = x^*$, then:

$$f'(x^*) = 0$$

❖ **Important Remarks**

1. This theorem can be proved even if x^* is a relative maximum.

2. The theorem ***does not say*** what happens if a minimum or maximum occurs at point x^* where the derivative fails to exist.

Example:

Consider the example of fig.2.

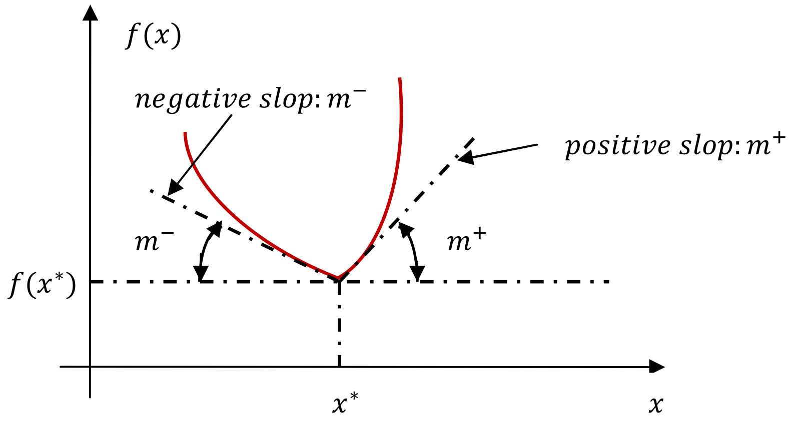

Fig.2 undefined derivative at x^*

In this case, the derivative is calculated as:

$$\lim_{h \to 0} \frac{f(x^* + h) - f(x^*)}{h} = m^+(positive) or\ m^-\ (negative)$$

Depends on whether h approaches zero through positive or negative values, respectively.

Obviously, in this case the derivative does not exist unless the values m^+ and m^- are equal.

If the derivative at the point $x = x^*$ does not exist, the above theorem is not applicable.

3. The theorem ***did not say*** also what happens if a minimum or maximum occurs at the endpoint of the interval over which the function is defined. That is:

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

Exists for positive values of h only or for negative values of h only, in either case, the derivative does not exist and hence it is not defined at the endpoints of the interval.

4. The theorem ***does not say*** that the function necessarily will have minimum or maximum at every point where the derivative is zero.

Example:

Consider the example illustrated in fig.3.

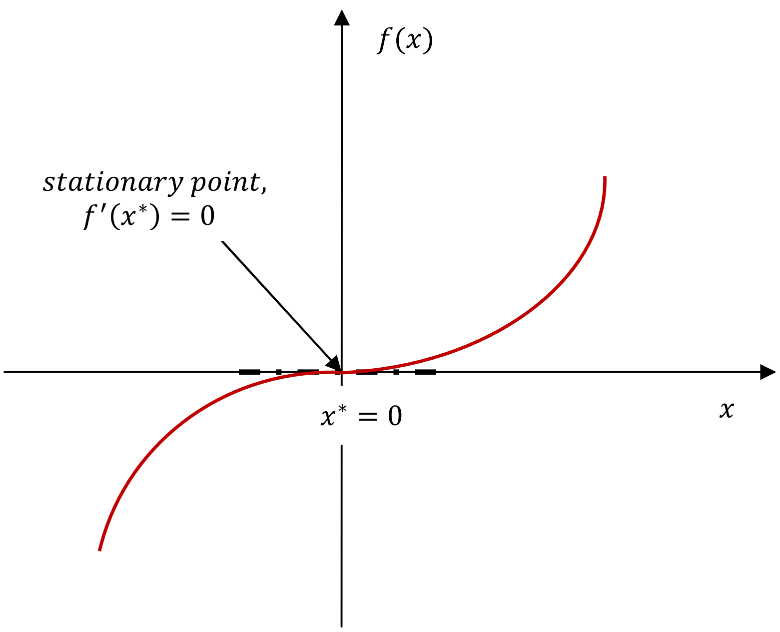

Fig.3 stationary (inflection) point

In this example, we have $f'(x=0)=0$, however and as it can be seen this point is ***neither minimum nor maximum***.
For that reason we say that in general, a point x^* at which the ***first derivative*** equals zero; $f'(x^*)=0$ is called a **stationary point (critical point)** which is just candidate of being Extremum. Therefore, if the function of one variable $f(x)$ possesses continuous derivatives of every order that come in question, in the neighborhood of $x=x^*$, the following theorem provides the sufficient condition for the minimum or maximum of the function $f(x)$.

2.1.4.2 Theorem 2.2 (Sufficient Conditions of Optimality)

Let $f'(x^*)=f''(x^*)=f^{(3)}(x^*)=\cdots=f^{(n-1)}(x^*)=0$, but $f^{(n)}(x^*)\neq 0$, then :

- **(i)** $f(x^*)$ is a minimum value of $f(x)$ if $f^{(n)}(x^*)>0$ and n is even.
- **(ii)** $f(x^*)$ is a maximum value of $f(x)$ if $f^{(n)}(x^*)<0$ and n is even.
- **(iii)** $f(x^*)$ is neither minimum nor maximum value of $f(x)$ if n is odd.

Example:

Determine the minimum and maximum values of the function defined by:

$$f(x)=12x^5-45x^4+40x^3+5$$

Solution:

<u>**Step 1:**</u> calculate the first derivative and determination of stationary ponts :

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

Then :

$$f'(x)=0\Rightarrow 60x^2(x^2-3x+2)=0\Rightarrow\begin{cases}x=0\\ \\ x^2-3x+2=0\end{cases}\Rightarrow\begin{cases}x=0\\x=1\\x=2\end{cases}$$

According to <u>***theorem 2.1***</u>, the points $x=0,1,2$ are stationary points.

To verify the minimum and maximum, we need to calculate the second derivative at these points:

The second derivative of the function $f(x)$ is:

$$f''(x)=240x^3-540x^2+240x=60x(4x^2-9x+4)$$

Step 2: we evaluate the second derivative at each candidate point

We find:

At $x=0$: $f''(x=0)=240x^3-540x^2+240x=0$

At $x=1$: $f''(x=1)=240(1)-540(1)+240(1)=-60$

At $x=2$: $f''(x=2)=240(2)^3-540(2)^2+240(2)=240$

Conclusion:

$f''(x=1)=-60\leq 0$, therefore, $x=1$ is a local (relative) maximum, where $f_{max}=f(x=1)=12$.

$f''(x=2)=240\geq 0$, therefore, $x=2$ is a local (relative) minimum, where $f_{min}=f(x=2)=-11$.

$f''(x=0)=0$, therefore, we need to investigate the third derivative at $x=0$. We find:

$$f^{(3)}(x)=720x^2-1080x+240$$

Hence:

$f^{(3)}(x=0)=240\neq 0$ and n is odd (= 3), which proves that $x=0$ is neither minimum nor maximum.

3- Multivariable Unconstrained Optimization Problem

Now consider more general functions of the type $f:\mathbb{R}^n\to\mathbb{R}$ (multivariate).

2.1 Necessary and Sufficient Optimality Conditions

In this section we consider the necessary and sufficient conditions for the minimum or maximum of an unconstrained function of several variables.

2.1.1 Theorem 2.3 (First order (Necessary) Conditions of Optimality)

If the function f of several variables $\boldsymbol{X}$ **de**noted by $f(\boldsymbol{X})$ has an extreme point (Maximum or Minimum) at $\boldsymbol{X} = \boldsymbol{X}^*$, and if the **first partial derivatives** of that function $f(\boldsymbol{X})$ at $\boldsymbol{X}^*$ exist, then we should have:

$$\nabla f(\mathrm{x}^*) = \frac{\partial f(\mathrm{x}^*)}{\partial x_i} = 0, \ \ i = 1,2\,3, \dots, n \Rightarrow \frac{\partial}{\partial x_1} f(\mathbf{x}^*) = \frac{\partial}{\partial x_2} f(\mathbf{x}^*) = \cdots = \frac{\partial}{\partial x_n} f(\mathbf{x}^*) = 0$$

The above theorem states that at an interior **local Extremum** all **first partial derivatives** must be equal to zero, i.e. we can solve the system of n equations defined by the condition above and look for interior extrema only among its solutions.

Note also that the above can be written equivalently as;

$$\nabla f(\mathrm{x}^*) = 0_{n\times 1}$$

i.e., at interior **local Extremum** the **gradient** of f is **zero**.

Remembering that **the gradient** was a vector pointing in the direction in which the function changes fastest, we see that the above condition implies that at the ***Extremum*** there's no such best direction, i.e. if we go in any direction we will reach a lower functional value (if we are talking about a maximum). Therefore, the first-order condition $\nabla f(x^*) = 0_{n\times 1}$ is **only necessary**. The following theorem states the sufficient conditions for a stationary point to be a local Extremum (minimum or maximum) of an unconstrained multivariable objective function.

2.1.2 Theorem 2.4 (Second order (Sufficient) Conditions of Optimality)

A sufficient condition for a ***stationary point*** $\boldsymbol{X}^*$ to be an extreme (maximum or minimum) point of an unconstrained multivariable objective function $f(\mathbf{x})$ is that the matrix formed of the **second partial derivatives** of the function $f(\mathbf{x})$ which is called ***Hessian Matrix*** $(H(\mathbf{x}))$ evaluated at $\mathbf{x} = \mathbf{x}^*$, **is:**

- **(i) Positive definite** when $\mathbf{x}^*$ is **strict** relative (local) minimum point.
- **(ii) Negative definite** when $\mathbf{x}^*$ is **strict** relative (local) maximum point.

Additionally, we can state the following theorem showing the sufficient conditions for a stationary point $\mathbf{x} = \mathbf{x}^*$ to be just local Extremum of an unconstrained multivariable objective function.

• Theorem 23 (Second-order necessary conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be* smooth function of several variables. If x^* is a local maximum (minimum) of f then:

1) $\nabla f(\mathrm{x}^*) = 0_{n\times 1}$

2) $H(\mathrm{x}^*)$ is Negative semidefinite, n.s.d. (Positive semidefinite, p.s.d.).

2.2 Case of Saddle Point

In case of function of two variables denoted by $f(\boldsymbol{X}), with\ X = (x_1, x_2)$, when the Hessian matrix $H(\boldsymbol{X})$ is neither positive nor negative definite at the point $X^* = ({x_1}^*, {x_1}^*)$, and the necessary condition for relative Min or Max is satisfied; that is:

$$\frac{\partial}{\partial x_1} f(\boldsymbol{X}^*) = \frac{\partial}{\partial x_2} f(\boldsymbol{X}^*) = 0$$

In such a case, the point $X^* = ({x_1}^*, {x_2}^*)$, is called **saddle point**.

2.2.1 Characteristics

The characteristics of a **saddle point** is that it corresponds to relative (local) minimum or maximum of the function $f(\boldsymbol{X})$ with respect to one variable, say x_1 (the other variable being fixed at $x_2 = x_2{}^*$) and a relative minimum or maximum of the function $f(\boldsymbol{X})$ with respect to the second variable (the other variable being fixed at $x_1 = x_1{}^*$).

2.2.2 Example2.1: saddle point

Consider the function of two variables $f(x_1, x_2)$ defined as:

$$f(x_1, x_2) = x_1^2 - x_2^2$$

For this function, we have:

❖ **Necessary condition**:

$$\frac{\partial}{\partial x_1} f(x_1, x_2) = \frac{\partial}{\partial x_1}(x_1^2 - x_2^2) = 2x_1$$

$$\frac{\partial}{\partial x_2} f(x_1, x_2) = \frac{\partial}{\partial x_2}(x_1^2 - x_2^2) = -2x_2$$

Then:

$$\nabla f(x_1, x_2) = 0 \Rightarrow \begin{cases} \frac{\partial}{\partial x_1} f(x_1, x_2) = 0 \\ \frac{\partial}{\partial x_2} f(x_1, x_2) = 0 \end{cases} \Rightarrow \begin{cases} 2x_1 = 0 \\ -2x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases}$$

As a result, from the necessary condition, the candidate relative minimum or maximum of the function $f(x_1, x_2)$ is at the point $X^* = (x_1{}^*, x_2{}^*) = (0,0)$

❖ **Sufficient condition**:

We compute the Hessian matrix $H(\boldsymbol{X})$ of the function $f(x_1, x_2)$ at the point $X^* = (x_1{}^*, x_2{}^*) = (0,0)$.

$$H(\boldsymbol{X}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{X}), \quad for \begin{cases} i = 1,2, \dots, n \\ j = 1,2, \dots, n \end{cases}$$

$$\Rightarrow H(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(X^*) & \frac{\partial^2}{\partial x_1 \partial x_2} f(X^*) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(X^*) & \frac{\partial^2}{\partial x_2 \partial x_2} f(X^*) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

We then check for the definiteness of the Hessian matrix:

Method 1: The eigenvalues:

If λ is an eigenvalues of H, then:

$$|\boldsymbol{H} - \lambda \boldsymbol{I}| = 0 \Rightarrow \left| \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right| = 0 \Rightarrow \left| \begin{bmatrix} 2-\lambda & -\lambda \\ -\lambda & -2-\lambda \end{bmatrix} \right| = 0$$

$$\Rightarrow (2-\lambda)(-2-\lambda) - \lambda^2 = 0 \Rightarrow -4 - 2\lambda + 2\lambda + \lambda^2 - \lambda^2 = 0 \Rightarrow -4 = 0$$

Which impossible case, therefore, the Hessian matrix is neither positive definite nor negative definite.

The point (0, 0) is a saddle point.

Method 2: the Principal Minors of the matrix H

$$H_1 = |a_{11}| = 2$$

$$H_2 = \begin{vmatrix} 2 & 0 \\ 0 & -2 \end{vmatrix} = -4$$

Is it positive definite?

- Not all the determinants of H are positive, so the Hessian matrix is not positive definite.

Is it negative definite?

To negative definite, the sign of the determinants should verify:

$$(-1)^j, \quad for\; j = 1,2,3, \dots, n.$$

We have, for: $j = 1 \Rightarrow H_1 > 0$

And for $j = 2 \Rightarrow H_2 < 0$

Hence, the matrix H is ***not negative definite***.

As a result, the Hessian matrix is neither positive definite nor negative definite, which reveals that the point (0, 0) is a **saddle point** for the function $f(x_1, x_2) = x_1^2 - x_2^2$.

It can be seen that $f(x_1, x_2{}^*) = f(x_1, 0) = x_1^2$ has a relative minimum at (0, 0) and $f(x_1{}^*, x_2) = f(0, x_2) = -x_2^2$ has a relative maximum at the point (0, 0).

Remark:

Saddle points may also exist for functions of several variables greater than 2. The characteristic of a saddle point stated above still hold provided that the two variables (x_1, x_2) are interpreted as vectors in multidimensional case.

3.3 Example 2.2:

Find the extreme points of the function defined as:

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

Solution:

Step 1: **necessary condition**

The necessary conditions for the existence of the extreme points of the function $f(x_1, x_2)$ is that the Gradient vanishes. That is;

$$\nabla f(x_1, x_2) = 0 \Rightarrow \begin{cases} \frac{\partial}{\partial x_1} f(x_1, x_2) = 0 \\ \frac{\partial}{\partial x_2} f(x_1, x_2) = 0 \end{cases}$$

We have:

$$\frac{\partial}{\partial x_1} f(x_1, x_2) = 0 \Rightarrow 3x_1^2 + 4x_1 = 0 \Rightarrow x_1(3x_1 + 4) = 0 \Rightarrow \begin{cases} x_1 = 0 \\ x_1 = -\frac{4}{3} \end{cases}$$

$$\frac{\partial}{\partial x_2} f(x_1, x_2) = 0 \Rightarrow 3x_2^2 + 8x_2 = 0 \Rightarrow x_2(3x_2 + 8) = 0 \Rightarrow \begin{cases} x_2 = 0 \\ x_2 = -\frac{8}{3} \end{cases}$$

These equations are simultaneously satisfied at the points: $(0, 0), \left(0, -\frac{8}{3}\right), \left(-\frac{4}{3}, 0\right), \left(-\frac{4}{3}, -\frac{8}{3}\right)$

To know exactly the nature of each point, we need to verify the sufficient condition.

Step 2: sufficient conditions

In general the sufficient conditions for the existence of extreme points for functions of several variables is the Hessian matrix $H(\boldsymbol{X})$.

The Hessian matrix is given as:

$$\Rightarrow H(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(X^*) & \frac{\partial^2}{\partial x_1 \partial x_2} f(X^*) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(X^*) & \frac{\partial^2}{\partial x_2 \partial x_2} f(X^*) \end{bmatrix} = \begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$$

Now we evaluate this matrix at the candidate extreme points:

If:

$$H_1 = |a_{11}| = |6x_1 + 4|$$

$$H_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{vmatrix}$$

The values of H_1 and H_2 with the according nature of the extreme points are summarized in the following table:

Point $X^* = (x_1{}^*, x_2{}^*)$	Value of H_1	Value of H_2	Nature of $H(\boldsymbol{X})$	Nature of $X^* = (x_1{}^*, x_2{}^*)$	Value of $f(\boldsymbol{X})$ at $X^* = (x_1{}^*, x_2{}^*)$
$(0, 0)$	$+4$	$+32$	Positive definite	Local minimum	6
$\left(0, -\frac{8}{3}\right)$	$+4$	-32	indefinite	Saddle point	$\frac{418}{27}$
$\left(-\frac{4}{3}, 0\right)$	-4	-32	indefinite	Saddle point	$\frac{194}{27}$
$\left(-\frac{4}{3}, -\frac{8}{3}\right)$	-4	$+32$	Negative definite	Local maximum	$\frac{50}{3}$

Example 2.3:

As an example, consider two special functions, namely an (affine) linear function, $l(x)$ and quadratic function $q(x)$ given respectively by:

$$\begin{cases} l(x) = a^t x + b \\ q(x) = \frac{1}{2} x^t G x + b^t x + c \end{cases}$$

Then, it follows that the gradient and Hessian matrix of $l(x)$ and $q(x)$ are as follows:

$$\nabla l(x) = \begin{bmatrix} \frac{\partial l}{\partial x_1} & \frac{\partial l}{\partial x_2} & . & . & . & \frac{\partial l}{\partial x_n} \end{bmatrix}^t = [a^t \quad a^t \quad . \quad . \quad . \quad a^t]^t$$

$$\nabla l(x) = [a^t]^t . [1 \quad 1 \quad . \quad . \quad . \quad 1]^t = a$$

Hence, the Hessian matrix is:

$$\nabla^2 l(x) = \left[\frac{\partial^2 l(x)}{\partial x_i \partial x_j} \right]_{i=1,2,\dots,n; j=1,2,\dots,n} = 0$$

For the quadratic function, we have:

$$\nabla q(x) = \begin{bmatrix} \frac{\partial q}{\partial x_1} & \frac{\partial q}{\partial x_2} & . & . & . & \frac{\partial q}{\partial x_n} \end{bmatrix}^t = Gx + b$$

And,

$$\nabla^2 q(x) = \left[\frac{\partial^2 q(x)}{\partial x_i \partial x_j} \right]_{i=1,2,\dots,n; j=1,2,\dots,n} = G$$