MDSC - 103

Sri Sathya Sai Institute of Higher Learning

(Deemed to be University)

Department of Mathematics and Computer Science

Muddenahalli Campus

Course: **M.Sc. Data Science and Computing** Date: **October 5, 2021**

Subject: **Optimization Techniques** Module : **Classical Optimization Theory**

Contents

1. Introduction

In this chapter we discuss, how to find optimal points for given function either in one variable or several variables. This is often called *classical optimization*, even though many of the techniques are of recent origin. Occasionally these techniques can be used to solve real-world problems. These techniques frequently used to optimize the cost function in Machine Learning and Deep Learning algorithms.

2. Optimization of Unconstrained Functions

2.1. **Optimization of Unconstrained Functions of One Variable.** We start with some definitions:

Definition 2.1. Consider a continuous function $f : I = (a, b) \to \mathbb{R}$.

(1) We say that f has a global minimum/ absolute minimum at $x_1 \in I$ if
$$f(x_1) \leq f(x) \ \forall x \in I$$
(2) We say that f has a global maximum/ absolute maximum at $x_1 \in I$ if
$$f(x_1) \geq f(x) \ \forall x \in I$$
(3) We say that f has a global extreame/ absolute extreame at $x_1 \in I$ if f has either a global minimum or a global maximum at x_1.
(4) We say that f has a local minimum/ relative minimum at $x_1 \in I$ if there exists a real number $\delta > 0$ such that
$$f(x_1) \leq f(x) \ \forall x \in (x_1 - \delta, x_1 + \delta)$$
(5) We say that f has a local maximum/ relative maximum at $x_1 \in I$ if there exists a real number $\delta > 0$ such that
$$f(x_1) \geq f(x) \ \forall x \in (x_1 - \delta, x_1 + \delta)$$
(6) We say that f has a local extreame/ relative extreame at $x_1 \in I$ if f has either a local minimum or a local maximum at x_1.
(7) We say that f has a stationary/ critical point at $x_1 \in I$ if f is differentiable at x_1 and $f'(x_1) = 0$.

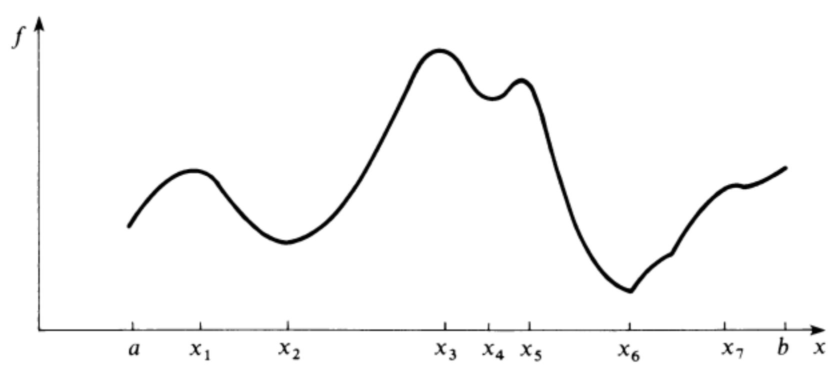


Figure 7.1. Points x_1, x_3, and x_5 are local maxima. Point x_3 is a global maximum. Points x_2, x_4, and x_6 are local minima. Point x_6 is a global minimum.

2.1.1. *A Necessary Condition for Local Extreme.* Our interest is to find global extreme of a given function f. Unfortunately it's not easy to find the global extreme directly. Instead of computing global extreme, first find all local extreme. The following theorem gives the necessary condition for existence of local extreme.

Necessary Condition for local extreme

Theorem 2.2. *If $f : I \to \mathbb{R}$ is differentiable at x_1, the f has a local extreme at x_1 then $f'(x_1) = 0$.*

We have the following definition:

Definition 2.3. f has a point of inflection at x_1 if f has a stationary point at x_1 but f does not have a local extreme at x_1 and f' has a local extreme at x_1.

Example 2.4. Consider the function $f(x) = x^3$. The point $x = 0$ is the point of inflection of f.

2.1.2. *Sufficient Condition for Local Extreme.* We can compute the stationary points by solving $f'(x) = 0$. One need to examine which of these stationary points are local extreme.The following result useful to determine the local extreme out of stationary points.

Sufficient Condition for local extreme

Theorem 2.5. *If $f^{(k)}(x_1) = 0,\ k = 1, 2, \dots, n$ and $f^{(n+1)}(x_1) \neq 0$ and $f^{(n+1)}(x)$ is continuous in the neighborhood of x_1, then f has a local extreme at x_1 if and only if $(n+1)$ is even. Further, f has a local minimum at x_1 if $f^{(n+1)}(x_1) > 0$ and f has a local maximum at x_1 if $f^{(n+1)}(x_1) < 0$.*

Remark 2.6. From above theorem, we can conclude that, x_1 is point of inflection if n is odd, and $f^{(k)}(x_1) = 0,\ k = 1, 2, \dots, n$ and $f^{(n+1)}(x_1) \neq 0$.

Example 2.7. Consider

$$f(x) = x^3 - 9x^2 + 27x - 27$$

To find the extreme of this function, we need to calculate the derivative of f and is given below:

$$f'(x) = 3x^2 - 18x + 27$$

By solving $f'(x) = 0$, we have unique zero at $x_1 = 3$. The double derivative of f is given below:

$$f''(x) = 6x - 18$$

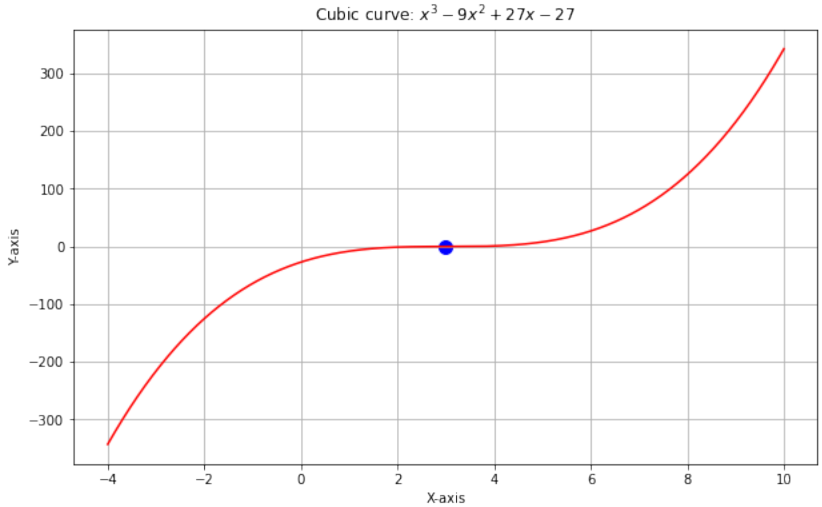
and

$$f''(3) = 0.$$

Now

$$f^{(3)}(x) = 6 \neq 0.$$

Thus the point $x_1 = 3$ is point of inflection for the function f.



Example 2.8. Consider

$$f(x) = x^4 - 8x^3 + 24x^2 - 32x + 16$$

The first derivative is

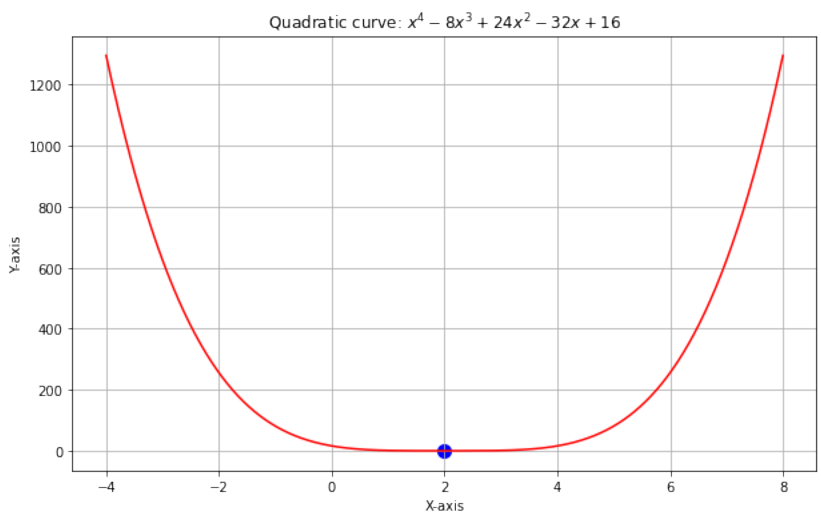
$$f'(x) = 4x^3 - 24x^2 + 48x - 32$$

which has unique zero at $x_1 = 2$. This is the only candidate for an extreme. However,

$$f''(x) = 12x^2 - 48x + 48$$

and

$$f''(2) = 0.$$



Also,

$$f^{(3)}(x) = 24x - 48$$

and

$$f^{(3)}(2) = 0.$$

But

$$f^{(4)}(2) = 24$$

and

$$f^{(4)}(2) = 24 \neq 0.$$

Therefore, (n+1) = 4, which is even, hence f has a local extreme at $x_1 = 2$. As $f^{(4)} = 24 > 0,\ x_1 = 2$ is a local minimum.

2.1.3. *The Taylor's Theorem.* Let f be a differentiable function on $(x, x+h)$, for $h > 0$. The mean value theorem states that

$$f(x+h) = f(x) + hf'(\theta x + (1-\theta)(x+h)), \text{ for some } \theta,\ 0 < \theta < 1.$$

This result can be generalized as follows if kth derivative of f is continuous on $[x, x+h]$ and $(k+1)$ th derivative exists on $(x, x+h)$.

Taylor's Theorem

Theorem 2.9. *If f^k is continuous on $[x, x+h]$ and $f^{(k+1)}$ exists on $(x, x+h)$, then*

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \cdots + \frac{h^{k+1}}{(k+1)!}f^{(k+1)}(\theta x + (1-\theta)(x+h)), \text{ for some } \theta,\ 0 < \theta < 1$$

2.1.4. *The Newton-Raphson Method: One Variable Case.* To compute the stationary points of the function $f(x)$, we need to solve $f'(x) = 0$. Some time this is highly nonliner and, hence difficult to solve it. The Newton-Raphson method is an iterative algorithm for solving nonlinear equations.

Consider the following equation

$$f(x) = 0$$

Let x_k be a given point. Then by Taylor's expansion is given as

$$f(x) \approx f(x_k) + f'(x_k)(x - x_k)$$

Thus the above equation can be written as

$$f(x_k) + f'(x_k)(x - x_k) = 0$$

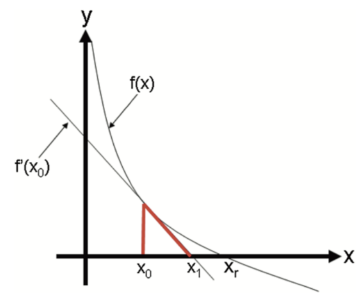
By solving for x, we have the following:

$$x = x_k - \frac{f(x_k)}{f'(x_k)} \quad \text{if } f'(x_k) \neq 0.$$

The idea of the method is to start from an initial point x_0, and then use the equation above to determine a new point. The process may or may not converge depending on the selection of the starting point. Convergence occurs when two successive points, x_k and x_{k+1}, are approximately equal (within specified acceptable tolerance).

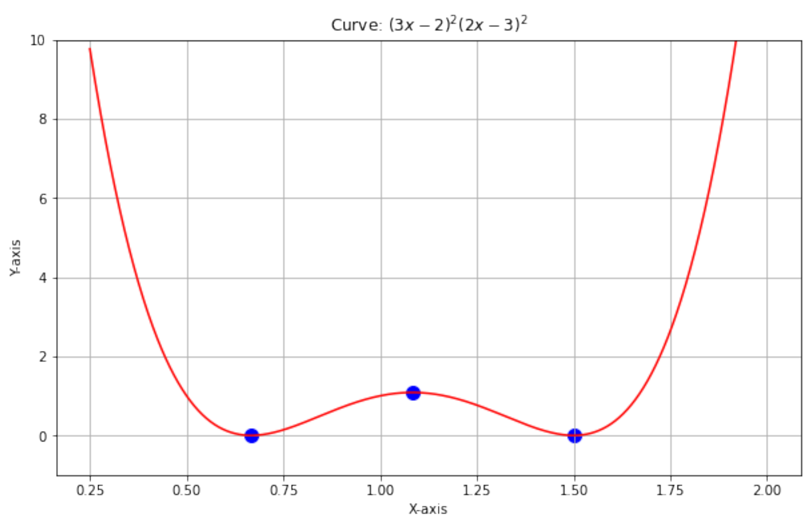
An illustration of the iterative process in the Newton-Raphson method is given below: Written generally, a Newton step computes an improved guess, x_{k+1}, using a previous guess x_k, and is given by the equation

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$



Example 2.10. Consider the function

$$g(x) = (3x-2)^2(2x-3)^2$$



By observing above plot, we conclude that there are three stationary points (x = 2/3, 13/12, and 3/2) for given function. These points can be investigated using Newton-Raphson method using suitable initial point.

To determine the stationary points of $g(x)$, we need to solve

$$f(x) = g'(x) = 72x^3 - 234x^2 + 241x - 78$$

Thus, for the Newton-Raphson method, we have

$$f'(x) = 216x^2 - 468x + 241$$

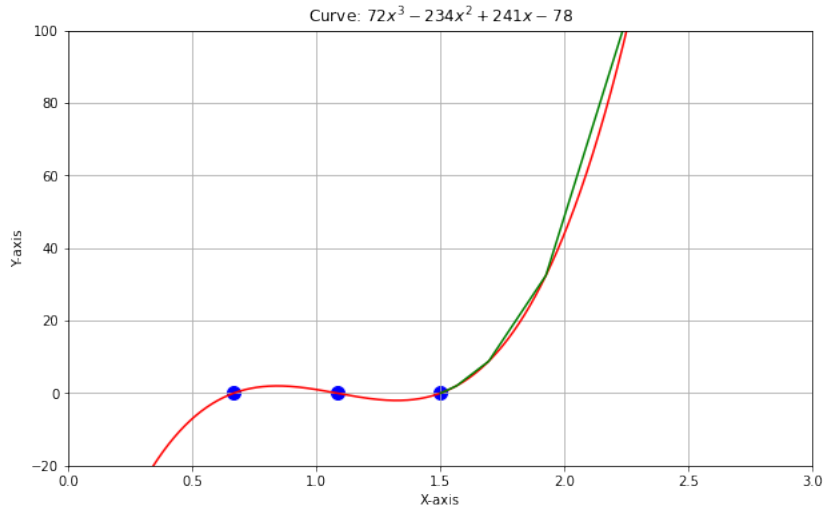
$$x_{k+1} = x_k - \frac{72x^3 - 234x^2 + 241x - 78}{216x^2 - 468x + 241}$$

Starting with $x_0 = 10$, the following table provides the successive iterations. The sequence x_k converges to 1.500432 with tolerance 1e-3, where $h(x_k) = f(x_k)/f'(x_k)$.

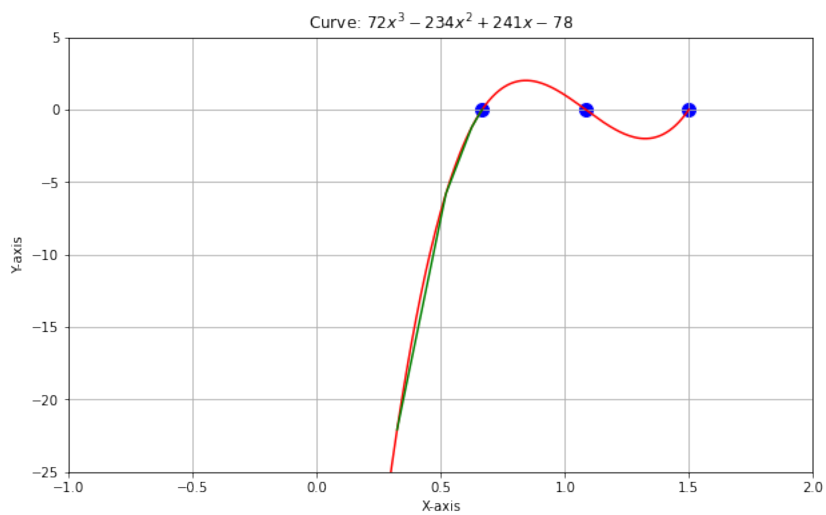
Geometrically, you can see in the below figure

Suppose, if you start with $x_0 = 0$, the sequence of x_k converges to 0.66657 with tolerance 1e-4 and the iterations given below. Similarly, if you start with $x_0 = 1$, then the sequence x_k converges to 13/12.

```
*** Newton Raphson Method Implementation***


Iteration,  xk,  h(xk),  xk+1
0  10.00000   2.967892   7.03211
1  7.03211   1.976429   5.05568
2  5.05568   1.314367   3.74131
3  3.74131   0.871358   2.86995
4  2.86995   0.573547   2.29641
5  2.29641   0.371252   1.92515
6  1.92515   0.230702   1.69445
7  1.69445   0.129000   1.56545
8  1.56545   0.054156   1.51130
9  1.51130   0.010864   1.50043

Estimated root is:  1.500432
```




```
*** Newton Raphson Method Implementation***


Iteration,  xk,  h(xk),  xk+1
0  0.00000   -0.323651   0.32365
1  0.32365   -0.196783   0.52043
2  0.52043   -0.103780   0.62421
3  0.62421   -0.037209   0.66142
4  0.66142   -0.005148   0.66657

Estimated root is:  0.66657
```



2.2. **Optimization of Unconstrained Functions of Several Variables.** Of course many models of real-world problems involve functions of many variables. In this section we generalize the results obtained in the earlier sections of this chapter.

We denote the a vector of n tuples by $\mathbf{X} = (x_1, \ldots, x_n)^T$. Consider a function $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a region. The Definitions (2.1) hold for multidimensional functions with $\mathbf{X}$ replacing x.

The following definition is useful in $n-$ dimensional case.

Definition 2.11 (Hessian Matrix)**.** The Hessian matrix is defined as

$$H(\mathbf{X}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{\mathbf{X}}$$

Taylor's theorem can be extended to functions of several variables:

Theorem 2.12. *(Taylor's Theorem for Several Variables) If the second partial derivatives of f are continuous and $\mathbf{X}$ and $\mathbf{X}+h$ are two points in the domain of f, then*

$$f(\mathbf{X}+h) = f(\mathbf{X}) + \nabla f(\mathbf{X})^T h + \frac{1}{2} h^T H(\theta \mathbf{X} + (1-\theta)(\mathbf{X}+h))h, \text{ for some } \theta,\ 0 < \theta < 1.$$

2.2.1. *A Necessary and Sufficient Condition for Local Extreme.* The following result is a necessary condition for the existence of local extreme in several variable case.

Necessary condition for local extreme

Theorem 2.13. *If $\partial f(\mathbf{X})/\partial x_j$ exists for all $\mathbf{X} \in S$ and for all $j = 1, 2, \ldots, n$ and if f has a local extreme at X^* in the interior of S, then*

$$\frac{\partial f(\mathbf{X}^*)}{\partial x_j} = 0, \quad j = 1, 2, \ldots, n.$$

Equivalently,

$$\nabla f(\mathbf{X}^*) = 0$$

Note that the simultaneous equations will give the stationary points. To determine the local extreme, the sufficient conditions useful. The following result is sufficient condition for a local extreme in $n-$ dimensional case.

Sufficient Condition for local extreme

Theorem 2.14. *If*

$$\frac{\partial f(\mathbf{X}^*)}{\partial x_j} = 0, \quad j = 1, 2, \ldots, n,$$

for some $\mathbf{X}^*$ in the interior of S, and if $H(X^*)$, the Hessian matrix of f evaluated at X^*, then

(1) f has a local maximum at X^* if the Hessian matrix $H(X^*)$ is negative definite

(2) f has a local minimum at X^* if the Hessian matrix $H(X^*)$ is positive definite.

Remark 2.15. (1) If all the eigen values of a symmetric are positive (negative) then it is defined as positive (negative) definite matrix.

(2) If all upper left $k \times k$ sub-matrix determinants of a symmetric are positive, the matrix is positive definite.

(3) A matrix A is positive definite if $x^T A x > 0$ for all vectors $x \neq 0$.

Example 2.16. Consider

$$f(\mathbf{X}) = x_1^4 + x_2^4 - 2x_1^2 + 4x_1 x_2 - 2x_2^2,\ \mathbf{X} \in \mathbb{R}^2$$

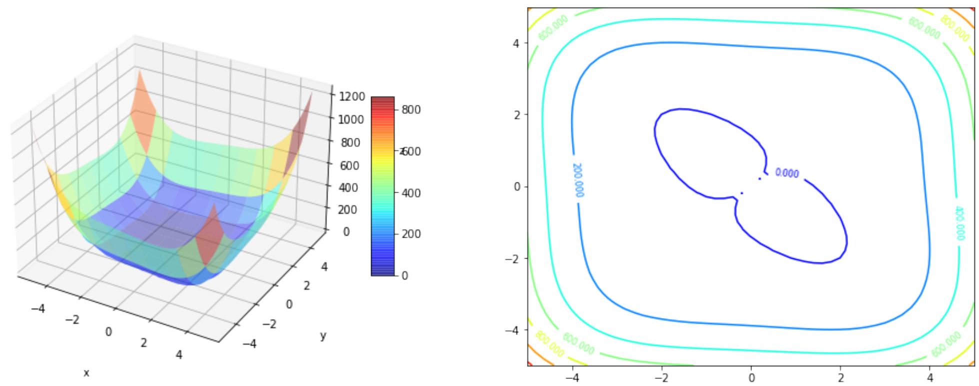


Figure 1. Surface

From above contour figure, we can conclude that it has two local minima. That we can investigate it now.

Observe that

$$f_x = 4x^3 - 4x + 4y;\ f_y = 4y^3 + 4x - 4y;$$

$$f_{xx} = 12x^2 - 4;\ f_{xy} = 4;\ f_{yy} = 12y^2 - 4$$

The Hessian matrix for f is given below:

$$H = \begin{bmatrix} 12x^2 - 4 & 4 \\ 4 & 12y^2 - 4 \end{bmatrix}$$

To find the stationary points, the necessary condition is

$$f_x = 0 = f_y$$

Thus

$$x^3 - x + y = 0;\ y^3 + x - y = 0$$

Adding above two equations, we have:

$$x^3 + y^3 = 0 \text{ or } y = -x$$

Putting $y = -x$ in $f_x = 0$, we obtain $x^3 - 2x = 0$. This implies

$$x = \sqrt{2}, -\sqrt{2}, 0.$$

Therefore, corresponding values of y are $-\sqrt{2},\ \sqrt{2}, 0$. Thus, possible local extreme are

$$(0,0), (\sqrt{2}, -\sqrt{2}) \text{ and } (-\sqrt{2}, \sqrt{2})$$

At $(\sqrt{2}, -\sqrt{2})$, the Hessian matrix is given as

$$H = \begin{bmatrix} 20 & 4 \\ 4 & 20 \end{bmatrix}$$

As the determinant of the matrix is non-zero. Thus the matrix is positive definite and hence the point $(\sqrt{2}, -\sqrt{2})$ has local minimum for the function f.

We have the same Hessian matrix as above at $(-\sqrt{2}, \sqrt{2})$. Thus the point $(-\sqrt{2}, \sqrt{2})$ has a local minimum for the function f.

If you evaluate the Hessian matrix at $(0,0)$, we have,

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

As the determinant of H is zero, the point $(0,0)$ is not an extreme. Thus it is saddle point.

2.3. **The Newton-Raphson Method: Several Variable Case.** To determine the stationary points in several variable function $f(\mathbf{X})$, the necessary condition is $\nabla f(\mathbf{X}) = 0$. This gives the system of equations and the solution of this system is known as stationary points in several variables. As $\nabla f(\mathbf{X}) = 0$ highly nonlinear and difficult to solve. This can be solved by using iterative technique called *The Newton-Raphson Method.*

Now, we discuss the Newton-Raphson method in several variable case.

Assume that $\nabla f(\mathbf{X}) = 0$ implies the following system of equations:

$$g_j(\mathbf{X}) = 0, \quad j = 1, 2, \ldots, m.$$

Let $\mathbf{X}_k$ be the given point. Then by Taylor's expansion we have

$$g_j(\mathbf{X}) \approx g_j(\mathbf{X}_k) + \nabla g_j(\mathbf{X}_k)(\mathbf{X} - \mathbf{X}_k), \quad j = 1, 2, \ldots, m$$

Thus, the original equations, $g_j(\mathbf{X}) = 0, \quad j = 1, 2, \ldots, m$ may be approximated as

$$g_j(\mathbf{X}_k) + \nabla g_j(\mathbf{X}_k)(\mathbf{X} - \mathbf{X}_k) = 0, \quad j = 1, 2, \ldots, m$$

These equations may be written in matrix notation as

$$A_k + J_k(\mathbf{X} - \mathbf{X}_k) = 0$$

where

$$\mathbf{A}_k = \begin{bmatrix} g_1(\mathbf{X}_k) \\ g_2(\mathbf{X}_k) \\ \vdots \\ g_m(\mathbf{X}_k) \end{bmatrix}$$

and the Jacobian matrix

$$\mathbf{J} = \begin{bmatrix} \nabla g_1(\mathbf{X}_k) \\ \nabla g_2(\mathbf{X}_k) \\ \vdots \\ \nabla g_m(\mathbf{X}_k) \end{bmatrix}_{m \times 1} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}_{m \times n}$$

If the Jacobean J is non-singular, then we have the following solution

$$\mathbf{X} = \mathbf{X}_k - \mathbf{J}^{-1}\mathbf{A}_k.$$

Thus we use the following iterative formulae, in the Newton-Raphson method for several variables

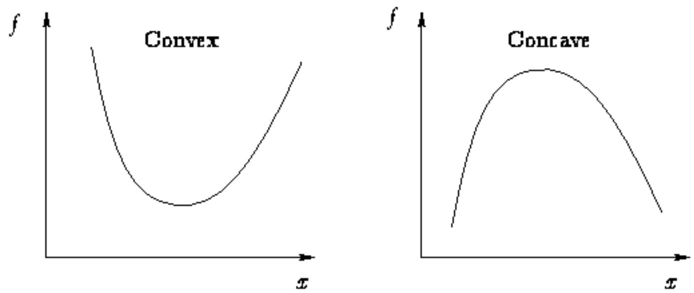
$$\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{J}^{-1}\mathbf{A}_k.$$

Example 2.17. Find the stationary point of the following problem using Newton-Raphson method:

$$f(\mathbf{X}) = x_1^2 + x_2^2$$

3. Convex and Concave functions

We discuss the definitions of convex and concave functions. Also we see some important properties of these functions to compute its global extreme.



Definition 3.1. (1) A function f defined on a simply connected region $S \subset \mathbb{R}^n$, is said to be *concave* on S if for all $\alpha \in [0,1]$, and for all $\mathbf{X_1}, \mathbf{X_2} \in S$,

$$f(\alpha \mathbf{X_1} + (1-\alpha)\mathbf{X_2}) \geq \alpha f(\mathbf{X_1}) + (1-\alpha) f(\mathbf{X_2}).$$

(2) A function f defined on a simply connected region $S \subset \mathbb{R}^n$, is said to be *convex* on S if $-f$ is if for all $\alpha \in [0,1]$, and for all $\mathbf{X_1}, \mathbf{X_2} \in S$,

$$f(\alpha \mathbf{X_1} + (1-\alpha)\mathbf{X_2}) \leq \alpha f(\mathbf{X_1}) + (1-\alpha) f(\mathbf{X_2}).$$

Example 3.2. Examples in $\mathbb{R}$

Convex

(1) affine line: $\{z | z = ax + b,\ a, b \in \mathbb{R}\}$
(2) exponential $e^{\alpha x},\ \alpha \in \mathbb{R}$
(3) negative entropy: $x \log x,\ x \in \mathbb{R}^+$

Concave

(1) affine line: $\{z | z = ax + b,\ a, b \in \mathbb{R}\}$
(2) power $x^{\alpha},\ x \in \mathbb{R}^+,$ for $0 \leq \alpha 1$
(3) $\log x,\ x \in \mathbb{R}^+$

Example 3.3. Which of the following sets are convex?

(1) The space $\mathbb{R}^n$
(2) A line through two given vectors $\mathbf{X_1}$ and $\mathbf{X_2}$

$$L(\mathbf{X_1}, \mathbf{X_2}) = \{z | z = \alpha \mathbf{X_1} + (1-\alpha)\mathbf{X_2},\ \alpha \in \mathbb{R}\}$$

(3) The positive orthant $\{\mathbf{X} \in \mathbb{R}^n | \mathbf{X} \succcurlyeq 0\}$.
(4) The set $\{\mathbf{X} \in \mathbb{R}^2 | x_1 x_2 = 0\}$.

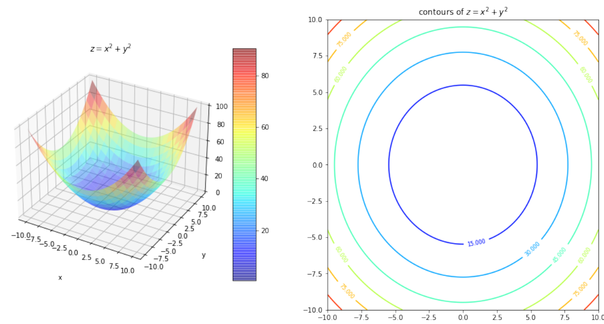


Figure 2. Convex

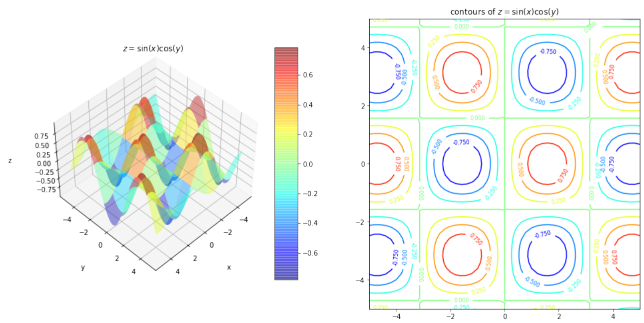


Figure 3. Non-Convex

3.1. **Graphical Illustrations of Convexity.** The main benefit of knowing whether a function is convex, as far as optimization is concerned, is provided by the following theorem.

Theorem 3.4.
If f is concave (convex) on a simply connected region $S \subset \mathbb{R}^n$ with a local maximum (local minimum) $\mathbf{X} \in S$ then f has a global maximum (global minimum) at $\mathbf{X}$.

Therefore, determine the convexity of a function is important. The following subsection gives some criteria to determine the convexity of function f.

3.1.1. *Determine the convexity of a function.* For the single-variable function $f(x)$, the following two characteristics are useful to determine the convexity of a function.

(1) The graph of a convex function $f(x)$ lies above each of its tangent lines. That is, if $f(x)$ is convex on an interval I and $x_1, x_2 \in I$, then
$$f(x_1) + f'(x_1)(x_2 - x_1) \leq f(x_2).$$
(2) If $f(x)$ is twice differentiable on an interval I, then $f(x)$ is convex on I if and only if $f''(x) \geq 0$ on I.

We have similar results in several variable case as well.

Theorem 3.5. *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$ is a function. The function f is convex on S if and only if S is convex and for $\mathbf{X}, \mathbf{Y} \in S$,*
$$f(\mathbf{X}) + \nabla f(\mathbf{X}).(\mathbf{Y} - \mathbf{X}) \leq f(\mathbf{Y}).$$

Corollary 3.6. *Let $S \subset \mathbb{R}^n$ is convex and f is a convex function. Then any stationary point of $f(\mathbf{X})$ is a global minimum for $f(\mathbf{X})$.*

The following theorem is very useful for determining whether a function is convex or strictly convex.

Theorem 3.7. *Let $S \subset \mathbb{R}^n$ open and convex and $f : S \to \mathbb{R}$ is a twice continuously differentiable function on S.*

(1) If the Hessian matrix $H(\mathbf{X})$ of f is positive semi-definite then f is convex on S
(2) If the Hessian matrix $H(\mathbf{X})$ of f is positive definite then f is strictly convex on S.

Example 3.8. Let

$$f(\mathbf{X}) = 2x_1^2 + x_2^2 + x_3^2 + 2x_2x_3$$

This function has the Hessian matrix

$$H(\mathbf{X}) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

As the minors has the non-negative values ($\triangle_1 = 4, \triangle_2 = 8$, and $\triangle_3 = 0$), the Hessian matrix $H(\mathbf{X})$ is positive semi-definite. Thus f is convex.

4. Optimization of Constrained Functions

This section deals with the optimization of constrained continuous functions. There are two cases are here. The first case is optimization of equal constraints and the other is optimization of inequality constraints.

General optimization problem with constraints is given below:

$$\text{Maximum or Minimum } z = f(\mathbf{X})$$

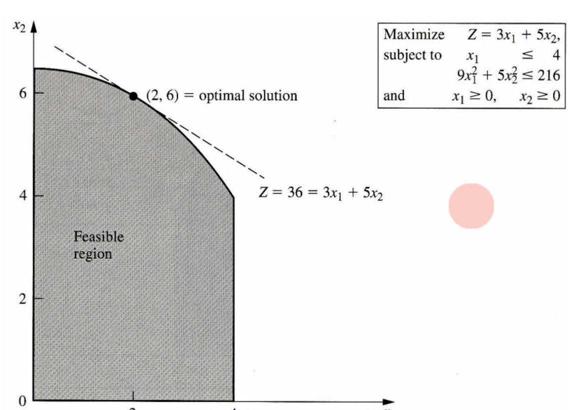
subject to

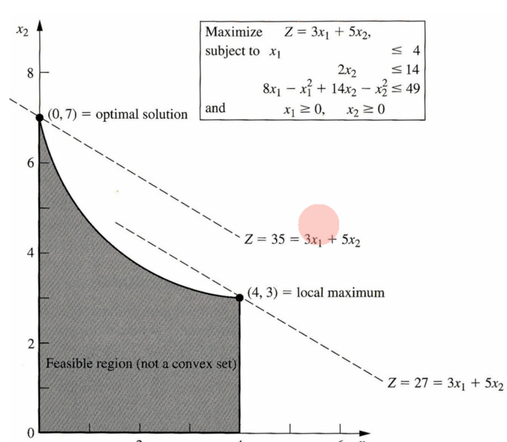
$$g_i(\mathbf{X}) \leq 0, \ i = 1, 2, \ldots, r$$

$$g_i(\mathbf{X}) \geq 0, \ i = r+1, \ldots, p$$

$$g_i(\mathbf{X}) = 0, \ i = p+1, \ldots, m$$

For constrained problem, we say that $\mathbf{X}$ is **feasible** if it satisfies all the constraints of the problem. The set of all feasible points constitute the feasible region of the problem. The graphical illustrations of feasible regions for constrained problems given below:





4.1. **Optimization of Equality Constraints.** In this section, we discuss two methods: the Jacobi and the Lagrange. The Lagrange method can be developed logically from the Jacobi. We remark that the Lagrange method used widely in Machine Learning algorithms like SVM.

Optimization Problem with Equality Constraints

Consider the problem

$$Maximize\ z = f(\mathbf{X})$$

subject to

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

where $\mathbf{X} = (x_1, x_2, \ldots, x_n)$

One obvious approach is to use the equations to eliminate some of the variables from the problem. For instance consider the following problem:

$$\text{Maximize: } z = x_2 + 2(x_2 - 4)^2 + 8$$

$$\text{subject to: } x_1 - x_2^2 + 4 = 0$$

This problem leads the following unconstrained problem in one dimension:

$$\text{Maximize:} f(x_2) = (x_2^2 - 4)^2 + 2(x_2 - 4)^2 + 8,$$

which is easier to solve. Of course, this approach of elimination will be successful in reducing the number of variables in the problem only if it is possible to express a solution for one or more of the variables explicitly. Often, however, this cannot be done.

It can be shown that when the variables of the objective function must satisfy constraints which are equations, the optimal point must lie on the boundary of the feasible region F. There are a number of methods available for locating optima which lie in the interior of F. We now present the Jacobi method and Lagrange's method, which both transform such a problem into one with its optima all lying in the interior of F.

4.1.1. *The Jacobi Method/ Constrained Derivatives Method.* We now present a method which solves the following problem

$$Maximize\ z = f(\mathbf{X})$$

subject to

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

where $\mathbf{X} = (x_1, x_2, \ldots, x_n)$

Assume that the functions $f(\mathbf{X})$ and $g_i(\mathbf{X}),\ i = 1, 2, \ldots, m$, are twice continuously differentiable.

The strategy of the Jacobi method is to find a suitable expression for ∇f at all points which satisfies $g_(\mathbf{X}) = 0,\ i = 1, \ldots, m$. The feasible stationary points of f are the ones among these for which $\nabla f(\mathbf{X}) = 0$.

Consider any point $\mathbf{X}$ which satisfies $g_i(\mathbf{x}) = 0$. Expanding f and $g_i,\ i = 1, \ldots, m$, in a Taylor series about $\mathbf{x}$, we get

$$f(\mathbf{X} + h) = f(\mathbf{X}) + \nabla f(\mathbf{X})^T h + \frac{1}{2} h^T H_f(\theta \mathbf{X} + (1 - \theta)(\mathbf{X} + h))h,$$

$$g_i(\mathbf{X} + h) = g_i(\mathbf{X}) + \nabla g_i(\mathbf{X})^T h + \frac{1}{2} h^T H_{g_i}(\theta \mathbf{X} + (1 - \theta)(\mathbf{X} + h))h,\ i = 1, \ldots, m,$$

for some $\theta,\ 0<\theta<1$. As $\mathbf{X}+h$ approaches $\mathbf{X}$, we get

$$f(\mathbf{X}+h)\approx f(\mathbf{X})+\nabla f(\mathbf{X})^T h$$

$$g_i(\mathbf{X}+h)\approx g_i(\mathbf{X})+\nabla g_i(\mathbf{X})^T h,\ i=1,\dots,m.$$

Therefore

$$\partial f(\mathbf{X})\approx \nabla f(\mathbf{X})^T\partial\mathbf{X}$$

$$\partial g_i(\mathbf{X})\approx \nabla g_i(\mathbf{X})^T\partial\mathbf{X},\ i=1,\dots,m.$$

Note that the equality constrains $g_i(\mathbf{X})=0$ satisfies the following equations:

$$\partial g_i(\mathbf{X})=0,\ i=1,\dots,m.$$

Thus

$$\nabla g_i(\mathbf{X})^T\partial\mathbf{X}=0,\ i=1,\dots,m.$$

Finally, we have the following system of $(m+1)$ equations with $(n+1)$ variables $\partial x_1,\ \partial x_2,\dots,\partial x_n,\partial f(\mathbf{X})$:

$$\partial f(\mathbf{X})=\nabla f(\mathbf{X})^T\partial\mathbf{X}$$

$$\nabla g_i(\mathbf{X})^T\partial\mathbf{X}=0,\ i=1,\dots,m.$$

Note that $\nabla f(\mathbf{X})$ and $\nabla g_i(\mathbf{X}),\ i=1,\dots,m$ consist of known constants.

If $m>n$, at least $(m-n)$ equations are redundant. Eliminating redundancy, the system reduces to $m\le n$. If $m=n$, the solution is $\partial\mathbf{X}=0$, which implies that there are no feasible points other than $\mathbf{X}$ in any neighborhood of $\mathbf{X}$. That is the set of feasible points is discrete. Hence we can assume that

$$m<n.$$

We redefine $\mathbf{X}=(x_1,x_2,\dots,x_n)$ as

$$\mathbf{X}=(w_1,w_2,\dots,w_m,y_1,\dots,y_{n-m}).$$

The variables $w_j,\ j=1,\dots,m$ are called *state variables* and the variables $y_j,\ j=1,\dots,(n-m)$ are called *decision variables*. Therefore the above system of equations can be expressed in terms of state and decision variables as follows:

$$\sum_{j=1}^{m}\frac{\partial f(\mathbf{X})}{\partial w_j}\partial w_j+\sum_{j=1}^{n-m}\frac{\partial f(\mathbf{X})}{\partial y_j}\partial y_j=\partial f(\mathbf{X})$$

$$\sum_{j=1}^{m}\frac{\partial g_i(\mathbf{X})}{\partial w_j}\partial w_j+\sum_{j=1}^{n-m}\frac{\partial g_i(\mathbf{X})}{\partial y_j}\partial y_j=0,\ i=1,2,\dots,m.$$

Suppose now that the $\partial y_i,\ i=1,2,\dots,(n-m)$ are given arbitrary values. Using these values, we can get unique values of $\partial w_i,\ i=1,\dots m$. Using the values of $\partial w_i,\ \partial y_i$ one can see if $\partial f(\mathbf{X})>0$, i.e., the new point $\mathbf{X}+h$ is an improvement over $\mathbf{X}$.

Define the matrix

$$\mathbf{J}=\begin{bmatrix}\nabla_w g_1\\ \nabla_w g_2\\ \vdots\\ \nabla_w g_m\end{bmatrix}$$

is called the *Jacobian matrix*, and the matrix

$$\mathbf{C} = \begin{bmatrix} \nabla_y g_1 \\ \nabla_y g_2 \\ \vdots \\ \nabla_y g_m \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_{n-m}} \\ \frac{\partial g_2}{\partial y_1} & \cdots & \frac{\partial g_2}{\partial y_{n-m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_{n-m}} \end{bmatrix}$$

is called the *control matrix*. The Jacobean $\mathbf{J}$ is assumed non-singular. This is always possible because the given m equations are independent by definition. The components of the vector $(w_1, \ldots, w_m)$ must thus be selected such that $\mathbf{J}$ is non-singular.

Now let

$$\mathbf{W} = (w_1, \ldots, w_m)$$

$$\mathbf{Y} = (y_1, \ldots, y_{n-m})$$

Then the system of equations can be written as

$$\nabla_w f^T \partial \mathbf{W} + \nabla_y f^T \partial \mathbf{Y} = \partial f(W, Y)$$

and

$$\mathbf{J}\partial\mathbf{W} + \mathbf{C}\partial\mathbf{Y} = 0.$$

As $\mathbf{J}$ is non-singular, we have

$$\partial\mathbf{W} = -\mathbf{J}^{-1}\mathbf{C}\partial\mathbf{Y}.$$

Thus

$$\partial f(W, Y) = (\nabla_y f^T - \nabla_w f^T \mathbf{J}^{-1}\mathbf{C})\partial\mathbf{Y}$$

From above equation, we can form what is known as the *constrained gradient* of f w.r.t Y, which is

$$\nabla_y^c f = \frac{\partial^c f(w, y)}{\partial^c y} = \nabla_y f^T - \nabla_w f^T \mathbf{J}^{-1}\mathbf{C}$$

Each element of $\nabla_y^c f$, namely $\frac{\partial^c f}{\partial^c y_i}$, $i = 1, 2, \ldots, (n-m)$, is called a *constrained derivative*. The necessary condition for $\mathbf{X}^*$ be a feasible maximum if

$$\nabla_y^c f(\mathbf{X}^*) = 0.$$

Above equation is used to identify all the stationary points; it remains to find which one is the global maximum. To do this construct Hessian matrix H which contain the constrained second derivatives with respect to the independent variables $y_1, \ldots, y_{n-m}$ only, and not $w_1, \ldots, w_m$. The complete method will be illustrated with a numerical example.

Example 4.1. Consider the following problem:

$$\text{Maximize } f(\mathbf{X}) = -2x_1^2 - x_2^2 - 3x_3^2$$

subject to

$$g_1(\mathbf{X}) = x_1 + 2x_2 + x_3 - 1 = 0$$

$$g_2(\mathbf{X}) = 4x_1 + 3x_2 + 2x_3 - 2 = 0$$

Here $m = 2,\ n = 3$, and we define

$$W = (x_1, x_2),\ Y = (x_3)$$

$$\nabla_w f = (-4x_1, -2x_2)^T$$

$$\nabla_y f = (-6x_3)$$

$$\mathbf{J} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \text{ and } \mathbf{J}^{-1} = \begin{bmatrix} -3/5 & 2/5 \\ 4/5 & -1/5 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The constrained gradient of f is given as

$$\begin{aligned} \nabla_y^c f(\mathbf{X}) &= \nabla_y f^T - \nabla_w f^T \mathbf{J}^{-1}\mathbf{C} \\ &= -6x_3 - (-4x_1, -2x_2)\begin{bmatrix} -3/5 & 2/5 \\ 4/5 & -1/5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= -6x_3 + \frac{4}{5}x_1 + \frac{4}{5}x_2 \\ &= 0, \end{aligned}$$

Combining this equations with the two original constraints, we have

$$\frac{4}{5}x_1 + \frac{4}{5}x_2 - 6x_3 = 0$$

$$x_1 + 2x_2 + x_3 = 1$$

$$4x_1 + 3x_2 + 2x_3 = 2,$$

Solving this system of equations, we have unique solution:

$$\mathbf{X}^* = (\frac{5}{27}, \frac{10}{27}, \frac{2}{27}),$$

which is a stationary point.

Now we determine whether this point is a maximum point.

$$\nabla_y^c f = \frac{\partial^c f}{\partial^c y_1} = \frac{4}{5}x_1 + \frac{4}{5}x_2 - 6x_3$$

$$\frac{(\partial^c)^2 f}{\partial^c y_1^2} = \frac{4}{5}\frac{dx_1}{dx_3} + \frac{4}{5}\frac{dx_2}{dx_3} - 6$$

$$\begin{aligned} \begin{bmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{bmatrix} &= \frac{\partial w}{\partial y} = -\partial J^{-1}\partial J \\ &= -\begin{bmatrix} -3/5 & 2/5 \\ 4/5 & -1/5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{5} \\ -\frac{2}{5} \end{bmatrix} \end{aligned}$$

Therefore

$$\frac{(\partial^c)^2 f}{\partial^c y_1^2} = \frac{4}{5}\frac{-1}{5} + \frac{4}{5}\frac{-2}{5} - 6 < 0$$

Thus $\mathbf{X}^*$ is indeed a maximum point.

Example 4.2. Let's consider another problem

$$\text{Minimize } f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2$$

subject to

$$g_1(\mathbf{X}) = x_1 + x_2 + 3x_3 - 2 = 0$$

$$g_2(\mathbf{X}) = 5x_1 + 2x_2 + x_3 - 5 = 0$$

We determine the constrained extreme points as follows. Let

$$\mathbf{W} = (x_1, x_2), \text{ and } \mathbf{Y} = (x_3)$$

Thus,

$$\nabla_w f = \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1}\Big)^T = (2x_1, 2x_2),\ \nabla_y f = \frac{\partial f}{\partial x_3} = 2x_3$$

$$\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 5 & 2 \end{bmatrix}, \mathbf{J}^{-1} = \begin{bmatrix} -2/3 & 1/3 \\ 5/3 & -1/3 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Hence,

$$\nabla^c f = \frac{\partial^c f}{\partial^c x_3} = 2x_3 - (2x_1, 2x_2)\begin{bmatrix} -2/3 & 1/3 \\ 5/3 & -1/3 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$= \frac{10}{3}x_1 - \frac{28}{3}x_2 + 2x_3$$

The equations for determining the stationary points are thus given as

$$\nabla^c f(\mathbf{X}) = 0$$

$$g_1(\mathbf{X}) = 0$$

$$g_2(\mathbf{X}) = 0$$

By solving these, we have

$$\mathbf{X}^* \approx (0.81, 0.35, 0.28)$$

This is the only candidate for the choice of local extrema. Now we investigate further to determine nature of local extrema. Given that x_3 is the independent variable, it follows from $\nabla^c f$ that

$$\frac{(\partial^c)^2 f}{\partial^c x_3^2} = (\frac{10}{3})(\frac{dx_1}{dx_3}) - (\frac{28}{3})(\frac{dx_2}{dx_3}) + 2$$

From the Jacobin Method,

$$\begin{bmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{bmatrix} = \mathbf{J}^{-1}\mathbf{C} = \begin{bmatrix} 5/3 \\ -14/3 \end{bmatrix}$$

Substitution gives $\frac{(\partial^c)^2 f}{\partial^c x_3^2} = \frac{460}{9} > 0$. Hence, $\mathbf{X}^*$ is the minimum point.

4.1.2. *The Method of Lagrange.* The following method was developed by Lagrange in 1761. From the arguments that developed in Jacobi method, we have

$$\partial f(W, Y) = \nabla_w f^T \partial \mathbf{W} + \nabla_y f^T \partial \mathbf{Y}$$

and

$$\partial g = \mathbf{J}\partial \mathbf{W} + \mathbf{C}\partial \mathbf{Y},$$

where $g = (g_1, g_2, \ldots, g_m)^T$, the vector of constraint functions. Note that

$$\partial \mathbf{W} = \mathbf{J}^{-1}\partial g + \mathbf{J}^{-1}\mathbf{C}\partial \mathbf{Y}$$

Then

$$\partial f(\mathbf{W}, \mathbf{Y}) = \nabla_w f^T \mathbf{J}^{-1}\partial g + \nabla_y f^T \partial \mathbf{Y} - \nabla_w f^T \mathbf{J}^{-1}\mathbf{C}\partial \mathbf{Y}$$

and recall constrained gradient of f from Jacobi method,

$$\nabla_y^c f = \nabla_y f^T - \nabla_w f^T \mathbf{J}^{-1}\mathbf{C}$$

Thus

$$\partial f(\mathbf{W}, \mathbf{Y}) = \nabla_w f^T \mathbf{J}^{-1}\partial g + \nabla_y^c f^T \partial \mathbf{Y}$$

If $\mathbf{X}^*$ is a local maximum, then $\nabla_y^c f(\mathbf{X}^*) = 0$. Therefore, we have

$$\partial f(\mathbf{W}^*, \mathbf{Y}^*) = \nabla_w f^T \mathbf{J}^{-1}\partial g$$

hence

$$\frac{\partial f(\mathbf{W}^*, \mathbf{Y}^*)}{\partial g} = \nabla_w f^T \mathbf{J}^{-1}.$$

The above equation is useful in allowing one to analyze the rate at which $f(\mathbf{W}^*, \mathbf{Y}^*)$, the optimal value, changes when g is perturbed. Let

$$\lambda = (\lambda_1, \ldots, \lambda_m) = \frac{\partial f(\mathbf{W}^*, \mathbf{Y}^*)}{\partial g} = \nabla_w f^T \mathbf{J}^{-1}$$

which implies

$$\partial f(\mathbf{W}^*, \mathbf{Y}^*) = \lambda \partial g.$$

Define F, the Lagrangian, as

$$F(\mathbf{X}, \lambda) = f(\mathbf{X}) - \lambda g.$$

The we have the following the system of equations:

$$\frac{\partial F}{\partial \lambda_j} = 0, \quad j = 1, 2, \ldots, m$$

$$\frac{\partial F}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n$$

It is necessary that any stationary point satisfies the above equations, which constitute a system of $(m+n)$ equations in $(m+n)$ unknowns, $\lambda_1, \ldots, \lambda_m, x_1, \ldots, x_n$. Any stationary point will produce a unique set of values for the elements of λ, as long as the above equations are independent. Hence these values are independent of which members of $\mathbf{X}$ are assigned to $\mathbf{W}$ and which to $\mathbf{Y}$.

Example 4.3. Determine all the stationary points of the following constrained problem

$$\text{Minimize } f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2$$

subject to:

$$x_1 + x_2 + 3x_3 = 2$$

$$5x_1 + 2x_2 + x_3 = 5$$

The Lagrangian function for the problem:

$$F(\mathbf{X}, \lambda) = x_1^2 + x_2^2 + x_3^2 + \lambda_1(x_1 + x_2 + 3x_3 - 2) + \lambda_2(5x_1 + 2x_2 + x_3 - 5)$$

By the method of Lagrangian, we have the following equations:

$$\frac{\partial F}{\partial x_1} = 2x_1 + \lambda_1 + 5\lambda_2 = 0$$

$$\frac{\partial F}{\partial x_2} = 2x_2 + \lambda_1 + 2\lambda_2 = 0$$

$$\frac{\partial F}{\partial x_3} = 2x_3 + 3\lambda_1 + \lambda_2 = 0$$

$$\frac{\partial F}{\partial \lambda_1} = 5x_1 + 2x_2 + x_3 - 5 = 0$$

We solve them:

$$\mathbf{X} = (0.8043, 0.3478, 0.2826), \text{ and } \lambda = (-0.0870, -0.3044)$$

We have only one KKT point, namely $\mathbf{X} = (0.8043, 0.3478, 0.2826)$, but we still do not know if this is optimal or not. The optimal can be investigated by checking Hessian matrix of f.

We will find that the Hessian is given by a diagonal matrix with diagonal entries equal to 2. The Hessian is positive definite, therefore the point $\mathbf{X}$ is a global minimum.

Example 4.4. Consider the problem

$$\text{Maximize: } f(\mathbf{X}) = -2x_1^2 - x_2^2 - 3x_3^2$$

subject to:

$$g_1(\mathbf{X}) = x_1 + 2x_2 + x_3 - 1 = 0$$

$$g_2(\mathbf{X}) = 4x_1 + 3x_2 + 2x_3 - 2 = 0$$

Now

$$F(\mathbf{X}, \lambda) = F(x_1, x_2, x_3, \lambda_1, \lambda_2)$$

$$= -2x_1^2 - x_2^2 - 3x_3^2 - \lambda_1(x_1 + 2x_2 + x_3 - 1) - \lambda_2(4x_1 + 3x_2 + 2x_3 - 2)$$

$$\frac{\partial F}{\partial x_1} = -4x_1 - \lambda_1 - 4\lambda_2 = 0$$

$$\frac{\partial F}{\partial x_2} = -2x_2 - 2\lambda_1 - 3\lambda_2 = 0$$

$$\frac{\partial F}{\partial x_3} = -6x_3 - \lambda_1 - 2\lambda_2 = 0$$

$$\frac{\partial F}{\partial \lambda_1} = -(x_1 + 2x_2 + x_3 - 1) = 0$$

$$\frac{\partial F}{\partial \lambda_2} = -(4x_1 + 3x_2 + 2x_3 - 2) = 0$$

If you solve these system of equations, we have the following:

$$(x_1^*, x_2^*, x_3^*, \lambda_1^*, \lambda_2^*) = (5/27, 10/27, 2/27, -4/27, -4/27).$$

4.2. **Optimization with Inequality Constraints.** General form of the problem in this section we consider as follows:

$$\text{Maximize:} \quad f(\mathbf{X})$$

$$\text{Subject to:} \quad g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m$$

Necessary condition for $\mathbf{X}$ to be a stationary point for above problem were developed by Kuhn and Tucker (1951). We discuss this technique now.

4.2.1. *The Karush–Kuhn-Tucker (KKT) Conditions.* Consider the problem:

$$\text{Maximize:} \quad f(\mathbf{X})$$

$$\text{Subject to:} \quad g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m$$

Adding non-negative slack variables s_j to left-hand side of the inequality constraint in the problem, we have

$$g_j(\mathbf{X}) + s_j = 0, \quad j = 1, 2, \ldots, m$$

$$s_j \geq 0, \quad j = 1, 2, \ldots, m$$

Now, we apply the method of Lagrange's and for m the Lagrangian:

$$F(\mathbf{X}, \lambda) = f(\mathbf{X}) - \sum_{j=1}^{m} \lambda_j (g_(\mathbf{X}) + s_j).$$

Necessary KKT conditions for maximization

Theorem 4.5. *If f has a local maximum $\mathbf{X}$ in the feasible region $\mathbf{R}$ of the problem*

$$\textit{Maximize: } f(\mathbf{X})$$

$$\textit{Subject to: } g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m,$$

where f and g_j, $j = 1, 2, \ldots, m$ have continuous first derivatives, and R is well-behaved at its boundary, then it is necessary that

$$\nabla f(\mathbf{X}) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{X}) = 0$$

$$g(\mathbf{X}) \leq 0$$

$$\lambda_j g_j(\mathbf{X}) = 0, \quad j = 1, 2, \ldots, m$$

$$\lambda_j \geq 0, \quad j = 1, 2, \ldots, m$$

for some set of real numbers $\lambda = (\lambda_, \ldots, \lambda_m)$.

The values $\lambda = (\lambda_, \ldots, \lambda_m)$ are called *generalized Lagrange multipliers.* The conditions in above theorem other than $g(\mathbf{X}) \leq 0$ are called KKT conditions.

Example 4.6. Consider the following problem

$$\text{Maximize} \quad f(\mathbf{X}) = 2x_1^2 - x_2^2 - 3x_3^2$$

$$\text{subject to:} \quad g_1(\mathbf{X}) = x_l + 2x_2 + x_3 - 1 \leq 0$$

$$g_2(\mathbf{X}) = 4x_1 + 3x_2 + 2x_3 - 2 \leq 0$$

It can be shown that the feasible region defined by $g_1(\mathbf{X})$ and $g_2(\mathbf{X})$ obeys the constraint qualification and so we can apply the result of theorem 4.5.

Let $\mathbf{X} = (x_1, x_2, x_3)$ be the local maximum, then:

$$(4x_1,\ -2x_2,\ -6x_3) - \lambda_1(1,2,1) - \lambda_2(4,3,2) = 0$$

$$x_l + 2x_2 + x_3 - 1 \leq 0$$

$$4x_1 + 3x_2 + 2x_3 - 2 \leq 0$$

$$\lambda_1(x_1 + 2x_2 + x_3 - 1) = 0$$

$$\lambda_2(4x_1 + 3x_2 + 2x_3 - 2) = 0$$

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0$$

Solving above equations we have the following:

$$\begin{aligned}
&-4x_1 - \lambda_1 - 4\lambda_2 = 0\\
&-2x_2 - 2\lambda_1 - 3\lambda_2 = 0\\
&-6x_3 - \lambda_1 - 2\lambda_2 = 0\\
&\lambda_1 x_1 + 2\lambda_1 x_2 + \lambda_1 x_3 = \lambda_1\\
&4\lambda_2 x_1 + 3\lambda_2 x_2 + 2\lambda_2 x_3 = 2\lambda_2
\end{aligned}$$

which has the following solution:

$$\mathbf{X} = (x_1, x_2, x_3) = (0, 0, 0),$$

and

$$(\lambda_1, \lambda_2) = (0, 0)$$

Hence $\mathbf{X}^* = 0$ is a local maximum.

Now we state the necessary conditions for the problem with a minimization objective.

Necessary KKT conditions for minimization

Theorem 4.7. *If f has a local minimum $\mathbf{X}$ in the feasible region $\mathbf{R}$ of the problem*

$$\text{Maximize: } f(\mathbf{X})$$

$$\text{Subject to: } g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m,$$

where f and $g_j,\ j = 1, 2, \ldots, m$ have continuous first derivatives, and R is well-behaved at its boundary, then it is necessary that

$$\nabla f(\mathbf{X}) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{X}) = 0$$

$$g(\mathbf{X}) \leq 0$$

$$\lambda_j g_j(\mathbf{X}) = 0, \quad j = 1, 2, \ldots, m$$

$$\lambda_j \leq 0, \quad j = 1, 2, \ldots, m$$

for some set of real numbers $\lambda = (\lambda_, \ldots, \lambda_m)$.

4.2.2. *When the KKT conditions are Sufficient.* The following theorems gives the sufficient conditions to have global extreme.

Sufficient KKT conditions for maximization

Theorem 4.8. *If f has a local maximum $\mathbf{X}$ in the feasible region $\mathbf{R}$ of the problem*

$$\text{Minimize: } f(\mathbf{X})$$

$$\text{Subject to: } g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m,$$

if f is concave and g_j is convex for $j = 1, 2, \ldots, m$ and there exist $\mathbf{X}^*$ and $\lambda = (\lambda_, \ldots, \lambda_m)$ which satisfy

$$\nabla f(\mathbf{X}^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{X}^*) = 0$$

$$g(\mathbf{X}) \leq 0$$

$$\lambda_j g_j(\mathbf{X}^*) = 0, \quad j = 1, 2, \ldots, m$$

$$\lambda_j \geq 0, \quad j = 1, 2, \ldots, m$$

then f has a global maximum at $\mathbf{X}^*$.

We state the analogous result for the problem with a minimization objective;

Sufficient KKT conditions for minimization

Theorem 4.9. *If f has a local minimum $\mathbf{X}$ in the feasible region $\mathbf{R}$ of In the problem*

$$\text{Minimize: } f(\mathbf{X})$$

$$\text{Subject to: } g_j(\mathbf{X}) \leq 0, \quad j = 1, 2, \ldots, m,$$

if f is convex and g_j is convex for $j = 1, 2, \ldots, m$ and there exist $\mathbf{X}^*$ and $\lambda = (\lambda_, \ldots, \lambda_m)$ which satisfy

$$\nabla f(\mathbf{X}^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{X}^*) = 0$$

$$g(\mathbf{X}) \leq 0$$

$$\lambda_j g_j(\mathbf{X}^*) = 0, \quad j = 1, 2, \ldots, m$$

$$\lambda_j \leq 0, \quad j = 1, 2, \ldots, m$$

then f has a global minimum at $\mathbf{X}^*$.

Table 1. Sufficiency of the KKT Conditions

Sense of optimization	**Required Conditions**	
	Objective function	Solution space
Maximization	Concave	Convex set
Minimization	Convex	Convex set

Finally, we state the sufficiency of KKT conditions for general nonlinear problem:

$$\text{Maximum or Minimum } z = f(\mathbf{X})$$

subject to

$$g_i(\mathbf{X}) \leq 0, \; i = 1, 2, \ldots, r$$

$$g_i(\mathbf{X}) \geq 0, \; i = r+1, \ldots, p$$

$$g_i(\mathbf{X}) = 0, \; i = p+1, \ldots, m$$

The Lagrangian function given as

$$L(\mathbf{X}, \mathbf{S}, \lambda) = f(\mathbf{X}) - \sum_{i=1}^{r} \lambda_i [g_i(\mathbf{X}) + s_i] - \sum_{i=r+1}^{p} \lambda_i [g_i(\mathbf{X}) - s_i] - \sum_{i=p+1}^{m} \lambda_i g_i(\mathbf{X})$$

The parameter λ_i is the Lagrange multiplier associated with constraint i. The conditions for establishing the sufficiency of the KKT conditions are summarized in Table 2.

The conditions in table 2 gives that Lagrangian function $L(\mathbf{X}, \mathbf{S}, \lambda)$ is concave in the case of maximization and a convex $L(\mathbf{X}, \mathbf{S}, \lambda)$ in case of minimization.

Table 2. Sufficiency of the KKT Conditions

Sense of optimization	**Required Conditions**			
	$f(\mathbf{X})$	$g_i(\mathbf{X})$	λ_i	
Maximization	Concave	*Convex*	≥ 0	$(1 \leq i \leq r)$
		Concave	≤ 0	$(r+1 \leq i \leq p)$
		Linear	Unrestricted	$(p+1 \leq i \leq m)$
Minimization	Convex	*Convex*	≤ 0	$(1 \leq i \leq r)$
		Concave	≥ 0	$(r+1 \leq i \leq p)$
		Linear	Unrestricted	$(p+1 \leq i \leq m)$

5. Python Implementation

Example 5.1. Consider the following problem:
Minimize $f(X) = (x-1)^2 + (y-2.5)^2$
subject to

$$\begin{aligned} x - 2y + 2 &\geq 0 \\ -x - 2y + 6 &\geq 0 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

First we load the required libraries.

```python
import numpy as np
from scipy.optimize import minimize
import matplotlib.pyplot as plt
```

Define the function and constraints:

```python
 # the objective function
 f = lambda x: (x[0]-1)**2 + (x[1]-2.5)**2
 # Constraints
 cons = ({'type': 'ineq','fun': lambda x: x[0]-2*x[1]+2},
         {'type': 'ineq','fun': lambda x: -x[0]-2*x[1]+6}
         )
 # Bounds to the variables
 bnds = ((0,None),(0,None))
```

Now, we solve the above problem using *minimize* function from *scipy.optimize* with initial point $X_0 = (2, 0)$.

```python
 X0 = (2,0)
 result = minimize(f,X0, bounds=bnds, constraints= cons)
 print(result)
```

Output

fun: 0.8000000011920985

jac: array([0.80000002, -1.59999999])

message: 'Optimization terminated successfully'

nfev: 10

nit: 3

njev: 3

status: 0

success: True

x: array([1.4, 1.7])

The optimal solution is $(1.4, 1.7)$ and the objective value at this point is 0.80. Now, we visualize this solution through *matplotlib.pyplot*.

```
x = np.linspace(-1, 4, 100)
y = np.linspace(0, 5, 100)
X, Y = np.meshgrid(x, y)
Z = f(np.vstack([X.ravel(), Y.ravel()])).reshape((100,100))
levels = np.arange(0.0,12,0.4)
cp = plt.contour(X, Y, Z,levels=levels, cmap='gist_rainbow');
plt.clabel(cp)
plt.plot(x, (x+2)/2, 'k:', linewidth=1)
plt.plot(res.x[0],res.x[1],c='r',marker='o')
plt.plot(x, (-x+6)/2, 'k:', linewidth=1)
plt.axis([-1,4,0,4])
plt.show()
```

The ouput as follows:

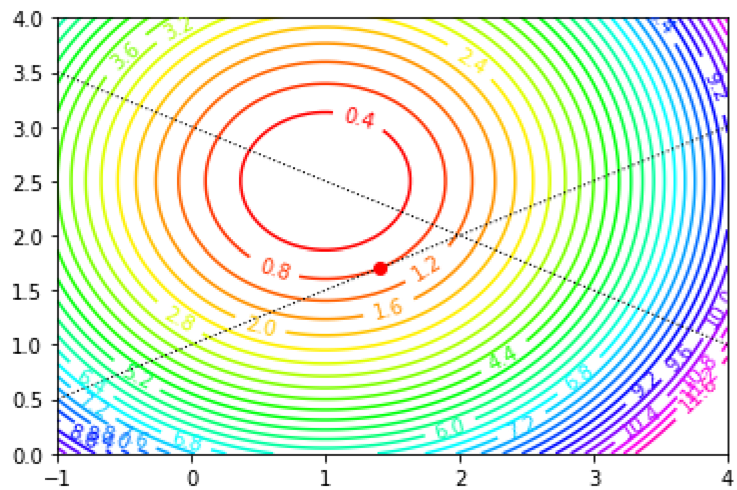


Figure 4. Contours of $f(x)$

6. Problems On Classical Optimization Techniques

(1) Locate all extreme of the following functions and identify the nature of each.
   (a) $f(x) = 6x^4 + 3x^2 + 42$
   (b) $f(x) = x^2 + 4x - 8$
   (c) $f(x) = 6x^5 - 4x^3 + 10$
   (d) $f(\mathbf{X}) = x_1^3 + x_2^3 - 3x_1x_2$
   (e) $f(\mathbf{X}) = x_1^2 - x_1 + 3x_2^2 + 18x_2 + 14$
   (f) $f(\mathbf{X}) = x_1^2 - 6x_1 + x_2^2 - 16x_2 + 25$
   (g) $f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 - 7x_1 - 8x_2^2 - 9x_3 + 101$

(2) Solve the following problem using Newton-Raphson method.
   (a) $f(x) = 4x^4 - x^2 + 5$
   (b) $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2$
   (c) $f(x_1, x_2) = x_1^2 - x_1 + 3x_2^2 + 18x_2 + 14$
   (d) $f(x_1, x_2) = 2x_1^2 + x_2^2 + x_3^2 + 6(x_1 + x_2 + x_3) + 2x_1x_2x_3$

(3) Solve the following problems using Jacobi Method
   (a) Maximize $f(\mathbf{X}) = 6x_1^2 + 3x_2^2 + 4x_1x_2$
   subject to: $x_1x_2 = 7$
   (b) Maximize $f(\mathbf{X}) = 2x_1^2 + x_2^2 + 3x_1 + 4x_2 + 9$
   subject to: $x_1^2 + x_2 + 3x_1x_2 = 11$
   $x_1 + x_2^2 + 4x_1x_2 = 12$
   (c) Minimize $f(\mathbf{X}) = x_1^2 + x_2^2$
   subject to: $x_1x_2 = 8$

(4) Consider the problem

$$\text{Minimize} f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

subject to

$$g_1(\mathbf{X}) = x_1 + 2x_2 + 3x_3 + 5x_4 - 10 = 0$$

$$g_2(\mathbf{X}) = x_1 + 2x_2 + 5x_3 + 6x_4 - 15 = 0$$

   (a) Show that by selecting x_3 and x_4 as independent variables, the Jacobian method fails to provide a solution and state the reason.
   (b) Solve the problem using x_1 and x_3 as independent variables, and apply the sufficiency condition to determine the type of the resulting stationary point.
   (c) Determine the sensitivity coefficients, given the solution in (b).

(5) Solve the following problems using method of Lagrange.
   (a) Maximize $f(\mathbf{X}) = 6x_1^2 + 3x_2^2 + 4x_1x_2$
   subject to: $x_1x_2 = 7$
   (b) Minimize $f(\mathbf{X}) = 5x_1^2 + x_2^2 + 2x_1x_2$
   subject to: $x_1x_2 = 10$

(6) Write the KKT necessary conditions for the following problems:

(a) Maximize $f(\mathbf{X}) = x_1^3 - x_2^2 + x_1 x_3^2$
subject to:

$$x_1 + x_2^2 + x_3 = 5$$
$$5x_1^2 - x_2^2 - x_3 \geq 2$$
$$x_1,\ x_2,\ x_3 \geq 0$$

(b) Minimize $f(\mathbf{X}) = x_1^4 + x_2^2 + 5x_1 x_2 x_3$
subject to:

$$x_1^2 - x_2^2 + x_3^2 \leq 10$$
$$x_1^3 + x_2^2 + 4x_3^2 \geq 20$$

(7) Consider the following problem

$$\text{Maximize } f(\mathbf{X})$$

Subject to

$$g_1(\mathbf{X}) \geq 0,\ g_2(\mathbf{X}) = 0,\ g_3(\mathbf{X}) \leq 0$$

Develop the KKT conditions, and give the stipulations under which the conditions are sufficient.

References

[1] Hamdy A. Taha, *Operations Reasearch: An Introduction*, 10th Ed., Pearson Education Limited 2017
[2] L. R. Foulds, *Optimization Techniques: An Introduction*, Springer-Verlag New York Inc. 1981