Chapter 1 Introduction to Optimization and Classification

1- Introduction to Optimization

1.1 Aim and Scope

In this chapter we start by stating the general optimization problem, and discussing how optimization problems can be classified by their functional forms, variable types and their structure.

1.2 Objective function and constraints

Optimization is the art of finding the ***best and most efficient solution*** from a collection of ***many other alternatives***. Optimization has numerous applications in science, engineering, finance, medicine and economics. In all these applications, we model our ***decision*** using a set of ***independent variables***, which may be **constrained** to lie in some region or to satisfy some restricted and conditioned functions. To reach this **best** and efficient solution, we must first of all define and identify an **objective**. This **objective** is defined as the ***quantitative measure*** of the ***performance*** to be ensured by the obtained solution.

Example: we can give as an example of the objectives:

- ✓ Profit.
- ✓ Time.
- ✓ Cost.
- ✓ Potential energy.

Consequently, the ***goal*** behind solving optimization problem is to find and identify values of these **independent variables** (*design variables*) that **maximize** (or **minimize**) a ***performance measure***, which we call the **objective (cost) function** that describes the problem being solved as well as, it gives a quantitative measure on the quality of the obtained **solution**. Formally, we formulate optimization problem of the following general form:

$$
\begin{cases}
\underset{x}{\text{minimize}}\, f(x) & (1.a) \\
\text{subject to:} \begin{cases} l_c \le c(x) \le u_c & (1.b) \\ l_A \le A^t x \le u_A & (1.c) \\ l_x \le x \le u_x & (1.d) \\ x \in \chi & (1.e) \end{cases}
\end{cases}
$$

Here, we are mainly concerned with finite dimensional optimization problems, where: $x \in \mathbb{R}^n$.

We have two types of functions that describe and fully define a given optimization problem. These are:

- The ***objective function***: $f(x), : f: \mathbb{R}^n \to \mathbb{R}$.
- The ***constraint function***: $c(x);\ c: \mathbb{R}^n \to \mathbb{R}^m$.

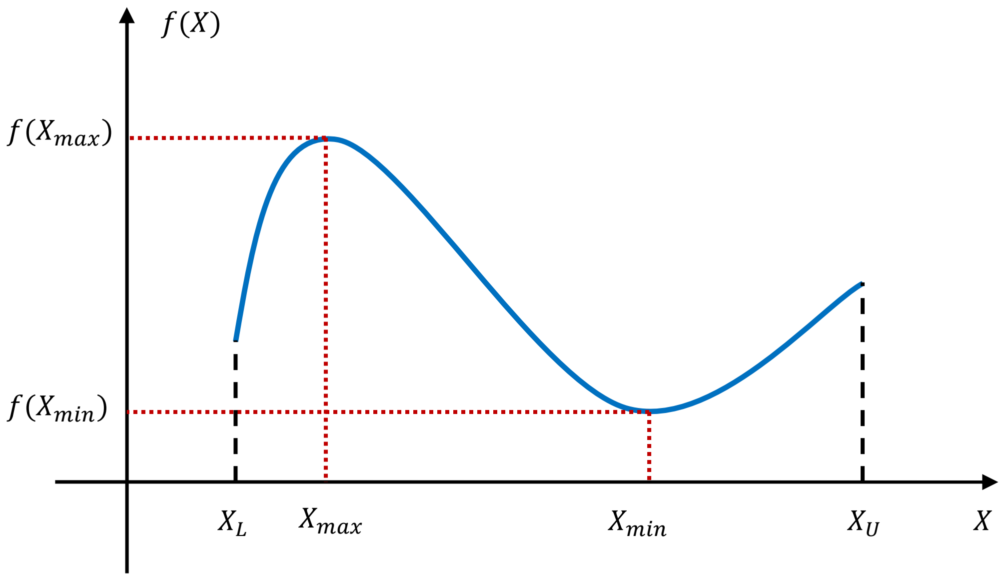
These two functions are particularly assumed ***smooth*** that is, typically ***twice continuous differentiable*** over an open set containing χ.

- The set $\chi \subset \mathbb{R}^n$: imposes further restrictions on the variables.
- $A \in \mathbb{R}^{n \times p}$: is a matrix.
- $l_c, u_c, l_A, u_A, l_x, u_x$: are lower and upper bounds.

1.3 Principle of Optimization

Optimisation theory is, in general, based on the principle of finding the extreme points (minima or maxima) of functions of one or more variables. The simplest and most common application of optimization theory is to find the **global minimum** of a scalar function $f(x)$ of a single variable x.

To illustrate this principle, consider the following given single variable function $f(x)$, which is graphically represented as:



1.4 Minimize vs. Maximize

It is well known that:

$$\max_x f(x) \equiv \min_x -f(x)$$

It turns out that maximization problem can be ***converted*** to a minimization one. Therefore, we will be interested only with minimization problem.

2- Preliminary Mathematics for Optimization

2.1 One-dimensional Function

One dimensional function is the function defined over single independent variable. It is denoted by $f(x)$ and mathematically defined as:

$$f: K \subseteq \mathbb{R} \to \mathbb{R}$$
$$x \longmapsto f(x)$$

2.1.1 First and second derivative of one dimensional functions

If the function of single variable $f(x)$ is continuously twice differentiable, the first and second derivatives are defined as:

$$\begin{cases} \frac{df(x)}{dx} & (1) \\ \frac{d^2f(x)}{dx^2} & (2) \end{cases}$$

2.2 Multidimensional Functions

A multidimensional function is defined as a function of several (more than one) independent real variables. In general, a function of n real variables is denoted by $f(\mathrm{x}) = f(x_1, x_2, \ldots, x_n)$ and mathematically defined as:

$$f: K \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$
$$\mathrm{x} \longmapsto f(\mathrm{x})$$

with:

$\mathrm{x} = [x_1 \quad x_2 \quad \dots \quad x_n]^{\mathrm{T}}$ is the vector of independent variables.

2.2.1 The Gradient

If the function is of n real variables, that is:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$$

Then the column vector formed of the first partial derivatives of the function $f(\mathbf{x})$ is called the **gradient** of the function denoted by $\nabla_x f(\mathbf{x})$, and defined as:

$$\nabla_x f(\mathrm{x}) = \frac{df(\mathrm{x})}{d\mathrm{x}} = \begin{bmatrix} \frac{\partial f(\mathrm{x})}{dx_1} \\ \frac{\partial f(\mathrm{x})}{dx_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f(\mathrm{x})}{dx_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdot & \cdot & \cdot & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$$

The gradient of the function is a vector that points in the direction of the greatest rate of the function's change. When the gradient of the function is equal to zero, this signifies that no change of the function is happened; meaning that the corresponding points are **stationary points** (candidate minima). We write this as:

$$x_1, x_2, \dots, x_n \; are \; stationary \; points \; \Leftrightarrow \nabla_x f(\mathrm{x}) = \frac{\partial f(\mathrm{x})}{\partial x_i} = 0, \quad i = 1,2,3, \dots \dots, n$$

2.2.2 The Hessian of a function

The **Hessian** of a function with n real variables denoted by: $\boldsymbol{H_x}$ is an $n \times n$ matrix consisting of the second order partial derivatives with respect to $(i,j)th$ element as it is expressed below:

$$\{\boldsymbol{H_x}\}_{i,j} = \frac{\partial^2}{\partial x_i x_j} f(\mathrm{x})$$

Also defined as:

$$H(\boldsymbol{X}) = \nabla^2 f(\mathrm{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{i=1,2,\dots,n;\, j=1,2,\dots,n} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdot & \cdot & \cdot & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdot & \cdot & \cdot & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdot & \cdot & \cdot & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

2.2.3 Positive definite and Negative Definite of a Matrix

A matrix A is ***positive definite*** if all of its eigenvalues are positive; that is all the values λ that satisfy the determinantal equation:

$$|\boldsymbol{A}-\lambda\boldsymbol{I}|=0$$

Should be **positive (strictly positive:** > 0**)**.

Similarly, the matrix A will be **negative definite** if all its eigenvalues are negative.

Other alternative:

Another test that can be used to find the definiteness of a matrix A of order n involves evaluation of the determinants as follows:

If we have the matrix A defined under the general form as:

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

Its all possible determinants are:

$$A_1=|a_{11}|$$

$$A_2=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

$$A_2=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$\vdots$$

$$A_n=\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

- The matrix A is said to be ***positive definite*** if and only if the values of all its determinants $(A_1, A_2, A_3, \ldots, A_n)$ are **positive**.
- The matrix A is said to be ***negative definite*** if and only if the sign of the determinant A_j is $(-1)^j,\ for\ j = 1,2,3,\ldots,n$.
- If some of the determinants A_j are positive and the remaining A_j are zero, the matrix is said to be **positive semidefinite**.

3- Convex and Concave Functions

3.1 Convex Function

A function $f(\boldsymbol{X})$ is said to be convex if for any pair of points:

$$\boldsymbol{X}_1 = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix}, \quad \boldsymbol{X}_2 = \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \vdots \\ x_n^{(2)} \end{bmatrix}$$

And all $\lambda,\ 0 \le \lambda \le 1$, then:

$$f(\lambda \boldsymbol{X}_2 + (\mathbf{1} - \lambda)\boldsymbol{X}_1) \le \lambda f(\boldsymbol{X}_2) + (\mathbf{1} - \lambda) f(\boldsymbol{X}_1) \qquad (1)$$

That is, the function is convex if the segment joining the two points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ lies entirely above or on the graph of the function $f(\boldsymbol{X})$.

The figures 1 and 2 illustrate a convex function in one and two dimensions.

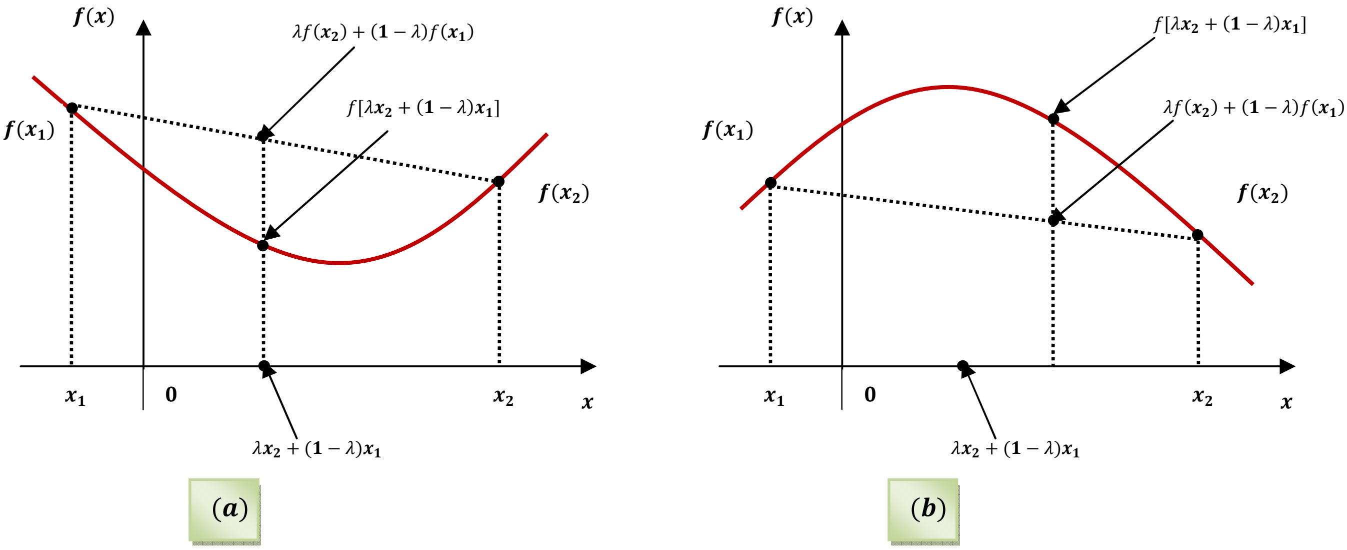


Fig.1 functions of one variable: (a) convex function, (b) concave function

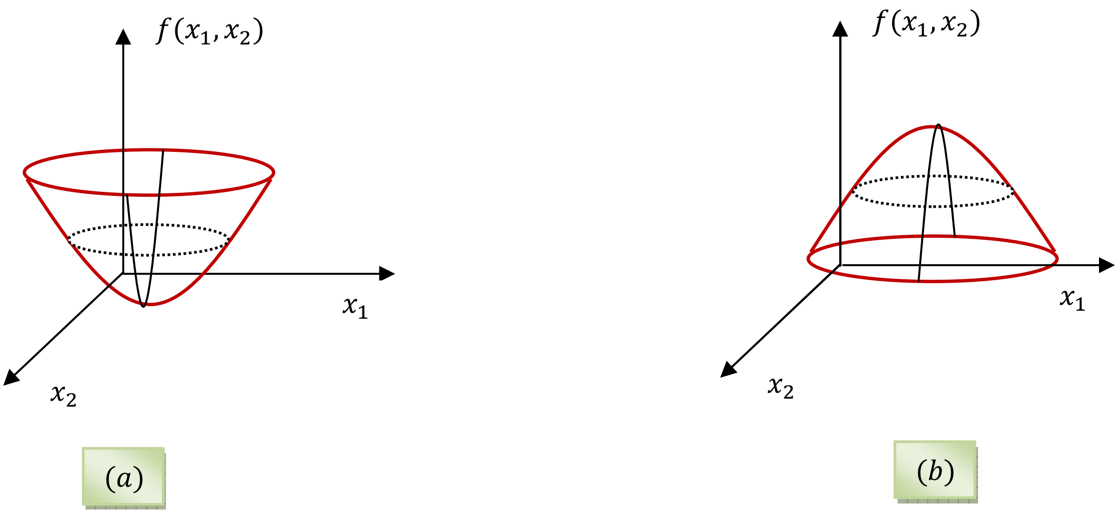


Fig.2 functions of two variables: (a) convex function, (b) concave function

It can be seen that a convex function always bending upward and hence it is apparent that the local minimum of a convex function is also a global minimum.

3.2 Concave function

A function $f(\boldsymbol{X})$ is called concave if for any two points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and for all $\lambda,\ 0 \le \lambda \le 1$:

$$f(\lambda \boldsymbol{X_2} + (\mathbf{1} - \lambda)\boldsymbol{X_1}) \geq \lambda f(\boldsymbol{X_2}) + (\mathbf{1} - \lambda) f(\boldsymbol{X_1}) \qquad (2)$$

That is, the function is concave if the line segment joining the two points lies below or on the graph of the function $f(\boldsymbol{X})$.

The fig 1. (b) and fig 2. (b) give and illustrate a concave function in one and two dimensions respectively.

It is apparent that the concave function, in general, bends downward and hence the local maximum will be always a global maximum.

- It is also obvious that the negative of the convex function is a concave function, and vice versa.
- We can note also that the sum of convex functions is a convex function.
- The sum of concave functions is a concave function.
- The function $f(\boldsymbol{X})$ is strictly convex or strictly concave if the strict inequality holds in equations (1) and (2) for any two different points, $\boldsymbol{X_1} \neq \boldsymbol{X_2}$.
- A function can be convex in a region and concave elsewhere, as it can be illustrated in the figure below:

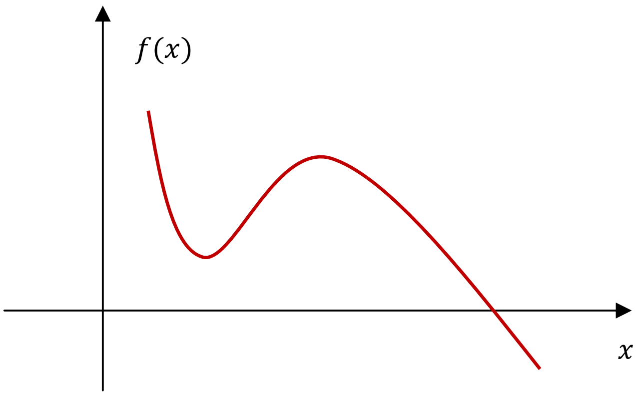


Fig.3 function that is convex over certain region and concave over other region

3.3 Testing for Convexity and Concavity

In addition to the preceding given definitions for convex and concave function, there are relations that can be used to identify whether an objective function is convex or concave.

3.3.1 Theorem 1:

A function $f(\boldsymbol{X})$ is convex if for any two different points $\boldsymbol{X_1} \neq \boldsymbol{X_2}$ we have:

$$f(\boldsymbol{X_2}) \geq f(\boldsymbol{X_1}) + \nabla f^T(\boldsymbol{X_1})(\boldsymbol{X_2} - \boldsymbol{X_1})$$

3.3.2 Theorem 2:

A function $f(\boldsymbol{X})$ is convex if its Hessian matrix:

$$H(\boldsymbol{X}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{X})$$

Is **positive semidefinite**.

Hence we can state the following two rules:

Rule 1:
A function $f(\boldsymbol{X})$ is convex if the Hessian matrix $H(\boldsymbol{X})$ is **positive semidefinite**.
Rule 2:
A function $f(\boldsymbol{X})$ is concave if the Hessian matrix $H(\boldsymbol{X})$ is **negative semidefinite**.

The following theorem establishes a very important relation, namely, any local minimum is global minimum for a convex function

3.3.3 Theorem 3:

If a function $f(\boldsymbol{X})$ is convex (concave), any **local** minimum (maximum) is **global** minimum (maximum).

Example

Determine whether the following functions are convex or concave.

(a) $f(x) = e^x$

(b) $f(x) = -8x^2$

(c) $f(x_1, x_2) = 3{x_1}^3 - 6{x_2}^2$

(d) $f(x_1, x_2, x_3) = 4{x_1}^2 + 3{x_2}^2 + 5{x_3}^2 + 6x_1x_2 + x_1x_3 - x_1 - 2x_2 + 15$

Solution

(a) $f(x) = e^x$

Since the function is of one variable, so it is convex or concave depends on the second derivative.

We have:

$$\frac{d^2}{dx^2}f(x) = \frac{d^2}{dx^2}[e^x] = \frac{d}{dx}[e^x] = e^x > 0, \forall x \in \mathbb{R}$$

That is $f(x) = e^x$ is strictly convex function.

(b) $f(x) = -8x^2$

Similarly, since the function is of one variable, we calculate the second derivative, instead of the Hessian, to identify it (whether convex or concave).

We have:

$$\frac{d^2}{dx^2}f(x) = \frac{d^2}{dx^2}[-8x^2] = \frac{d}{dx}[-16x] = -16 < 0, \forall x \in \mathbb{R}$$

Therefore, the function $f(x) = -8x^2$ is strictly concave.

(c) $f(x_1, x_2) = 3{x_1}^3 - 6{x_2}^2$

Here we calculate the Hessian matrix of the function:

We have:

$$H(\boldsymbol{X}) = \frac{\partial^2}{\partial x_i \partial x_j}f(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1}(3{x_1}^3 - 6{x_2}^2) & \frac{\partial^2}{\partial x_1 \partial x_2}(3{x_1}^3 - 6{x_2}^2) \\ \frac{\partial^2}{\partial x_2 \partial x_1}(3{x_1}^3 - 6{x_2}^2) & \frac{\partial^2}{\partial x_2 \partial x_2}(3{x_1}^3 - 6{x_2}^2) \end{bmatrix}$$

$$H(\boldsymbol{X}) = \begin{bmatrix} 18x_1 & 0 \\ 0 & -12 \end{bmatrix}$$

Using the method of determinants, we get:

$$H_1 = |18x_1| = 18x_1$$

$$H_2 = \begin{vmatrix} 18x_1 & 0 \\ 0 & -12 \end{vmatrix} = -144x_1$$

Here we have to study the cases:

We distinguish two cases depending on x_1

- If $x_1 < 0$, then $H_1 < 0, H_2 > 0$, that is the Hessian is negative semidefinite, and hence the function is strictly concave.
- If $x_1 > 0$, then $H_1 > 0, H_2 < 0$, that is the Hessian is indefinite, and hence the function is neither convex nor concave.

(d) $f(x_1, x_2, x_3) = 4{x_1}^2 + 3{x_2}^2 + 5{x_3}^2 + 6x_1x_2 + x_1x_3 - x_1 - 2x_2 + 15$

We calculate the Hessian matrix:

$$H(\boldsymbol{X}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_1 \partial x_2} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_1 \partial x_3} f(\boldsymbol{X}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_2 \partial x_2} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_2 \partial x_3} f(\boldsymbol{X}) \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_3 \partial x_2} f(\boldsymbol{X}) & \frac{\partial^2}{\partial x_3 \partial x_3} f(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{bmatrix}$$

Using the methods of determinants to determine the nature of the Hessian matrix as follows:

$$H_1 = |8| = 8 > 0$$

$$H_2 = \begin{vmatrix} 8 & 6 \\ 6 & 6 \end{vmatrix} = 12 > 0$$

$$H_3 = \begin{vmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{vmatrix} = 114 > 0$$

Hence all determinants of the Hessian matrix are positive, which indicate that it is positive semidefinite and the function is strictly convex.

4- Classification of Optimization Problems

In general, an optimization problem at the mathematical level is given and defined by an objective (cost) function:

$$f: K \rightarrow \mathbb{R}$$

And one search for the optimal point:

$$x^* \in K$$

Such that:

$$f(x^*) \leq f(x), \quad for\ all\ x \in K$$

Often, K is a subset of $\mathbb{R}^n$, that is:

$$K \subset \mathbb{R}^n$$

Which is defined by constraints.

Then, the optimization problem is formulated as:

$$\min_{x \in \mathbb{R}^n} f(x), \qquad (1.a)$$

$$s.t.$$

$$c_j(x) \geq 0 \quad j \in I \qquad (1.b)$$

$$c_j(x) = 0 \quad j \in E \qquad (1.c)$$

Where:

I: is an index set denoting the inequality constraints, and:

E: is an index set denoting the equality constraints of the optimization problem.

We denote the sizes of I and E respectively as:

$$\begin{cases} m_I = |I| \\ m_E = |E| \end{cases}$$

And:

$$c_j: \mathbb{R}^n \to \mathbb{R}$$

Are the constraints of the optimization problem.

When classifying optimization problems, one can roughly divide them into the following categories:

4.1 Unconstrained Optimization problems

In general, unconstrained optimization problem corresponds to the case where:

$$m_I = m_E = 0$$

In other words, **unconstrained optimization** problems are problems without any additional constraints; that is relations **(1.b)** to **(1.c)** are not present. Hence, **unconstrained optimization** problems are simply defined as follows:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\, f(x)$$

The ***unconstrained optimization*** problem represents, in fact, a category that encompasses many types of optimization problems, where we can define:

4.1.1 Unconstrained Quadratic optimization problems

In which, the objective (cost) function f is quadratic, that is, a function defined as:

$$\boldsymbol{f(x) = x^T A x - b^T x}$$

With: $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a symmetric square matrix.

The **quadratic optimization** problems require the solution of a **linear system**, where the **conjugate Gradient** algorithm may be conveniently used to solve it.

4.1.2 Unconstrained General nonlinear optimization problems

In this optimization problem, the objective function f is neither **linear** nor **quadratic.**

4.2 Constrained Optimization problems

The constrained optimization problems are those problems in which the constraint functions $c_j(x)$ is not zero. These problems are divided to the following:

4.2.1 Constrained Optimization Problem with Equality constraints

This corresponds to the case of $m_I = 0$ in the previously general formulation of the optimization problem.

4.2.2 Constrained optimization problems with Inequality constraints

This corresponds to the case of $m_E = 0$ in the previously general formulation of the optimization problem.

Furthermore, Constrained Optimization problems are found under the following common groups or types:

4.2.3 Linear programming optimization problems (LP)

These are characterized by:

- Both objective function and constraints are linear.
- They have the general statement form as:

$$\min_X C^T \boldsymbol{X}$$
$$s.t. \quad A\boldsymbol{X} \le b$$
$$and \quad X \ge 0$$

4.2.4 Quadratic Programming Optimization Problems (QP)

These optimization problems are characterized by:

- The objective function is quadratic and the constraints are linear.
- They have the following general statement form:

$$\min_{X} \boldsymbol{X^T} Q\boldsymbol{X} + C^T\boldsymbol{X}$$
$$s.t. \quad A\boldsymbol{X} \le b$$
$$and \quad X \ge 0$$

4.2.5 Nonlinear Programming Optimization Problems (NLP)

These are characterized by:

- The objective function is nonlinear or at least one constraint is nonlinear.

4.3 Optimal control problems

These are optimization problems where the variable x is function of one or several parameters.

4.4 Combinatorial optimization problems:

These are optimization problems where the set $\boldsymbol{K} \subset \mathbb{R}^n$ is discrete or even finite. We will not address the *combinatorial optimization problems*, because the techniques used to solve such problems are substantially different from the techniques used for all the other problems. Optimal control problems can be however handled by the techniques discussed in this chapter after discretization.

5- Solution strategy (Method) for a Given Optimization Problem

The solution strategy or algorithm for a given optimization problem depends on its class; that is, whether it is linear, nonlinear or quadratic. Consequently, each problem class requires its own algorithms.

6- First Optimization Example

Consider the design of a reinforced concrete beam to support a load (more complex examples, where we consider structures such as houses can be readily derived). We are interested in minimizing the cost of the reinforced beam, which is the sum of the steel reinforcement, and the concrete. The variables are the area of the re-reinforcement, x_1, and the width and depth of the beam, x_2 and x_3. The full problem can be defined as:

$$\begin{cases} \underset{x}{\text{minimize}} \quad f(x) = 29.4x_1 + 0.6x_2x_3 & \textit{cost of beam (cost function)} \\ \textit{subject to:} \begin{cases} c(x) = x_1 - 1x_2 - 7.735\dfrac{{x_1}^2}{x_2} \ge 180 & \textit{load constraint (constraint function)} \\ x_3 - 4x_2 \le 0 & \textit{width/ depth rati} \\ 40 \le x_1 \le 77, x_2 \ge 0, x_3 \ge 0 & \textit{simple bounds} \end{cases} \end{cases}$$

Where the **objective function** adds the cost of the reinforcement and the concrete, the nonlinear constraint describes the load, the linear constraint describes the desired width to depth ratio, and the simple bounds describe positivity and size constraints. In practice, some of the variables such as the size of the reinforcement may be integer, because they have to be chosen from a set of prefabricated units.