Chapter one

Basic concepts and preliminaries
"The tremendous power of the Simplex Method is a constant surprise to me."

G. B. Dantzig

1.1 Introduction

The first problem in linear programming was formulated in 1940 by a Russian mathematician Kantorovich, L. V. [86,87] and an American economist Hitchcock, F. L. [80] in 1941. They dealt with a well-known transportation problem which forms a branch of linear programming. Even though the French mathematician Jean-Baptiste-Joseph Fourier seemed to be aware of the subject potential as early as 1823, Kantorovich, L. V. published an extensive monograph [85] in 1939 and is credited with being the first to recognize that certain important broad classes of scheduling problems had well-defined mathematical structures. In 1941, an English economist, Stigler, G. [139] described yet another linear programming problem that of determining an optimal diet, not only at the minimum cost but also to satisfy minimum requirements.

Intensive work began in 1947 in US Air Force, when under the compulsions of World War II, a United Stated Air Force project called SCOOP (Scientific Computation of Optimum Programs) was setup under the leadership of Dantzig, G. B. The Simplex Algorithm and much of the related theory was developed by Dantzig, G. B. [37] and his team. Further work on special problems and methods continued throughout the next decade by Dantzig group in U. S. A. and by others in Europe.

Before 1947 all practical planning was characterized by a series of authoritatively imposed rules of procedure and priorities. General objectives were never stated, probably because of the impossibility of performing the calculations necessary to minimize an objective function under constraints.
In 1947 a method composed of successive tests for optimality at extreme points and intervening linear movements along polygon edges called the Simplex Computational Method was introduced by Dantzig, G. B. [37] which turned out to be an efficient method. Interest in linear programming grew rapidly and by 1951 its use spread to industry, Dantzing, G. B. [38], Koopmans, T. C. [94,95] etc. Afterwards, significant contribution in Linear Programming were made by Dantzig, G. B. [41]. Hadley, G. [75], Gale, D. [62], Gass, S. I. [64], Simmonard, M. [137], Llewelly, R. W. [107], Luenberger, D. G. [108], Barrodale, I. [13] and Murty, K. G. [116] etc.

Other forms of Linear Programming problems further studied includes:

Integer programming problems were introduced by Gomory, R. E. [71,72] in 1960. Unlike the earlier work on the traveling salesman problem by Dantzig, G. B., Fulkerson, D. R. et al. [40] showed how to systematically generate the ‘cutting’ planes. Cuts are extra necessary conditions which when added to an existing system of inequalities guarantee that the optimization solution will result in integers. Subsequently several techniques of generating cutting planes and Branch and Bound approach to solve integer programming were developed by different authors, Benders, J. R.[22], Balinski, M. L.[11], Balas, E.[8,9,10], Bellmore, M.[21], Glover, F. [69], Bradley, G. H. and Wahi, P. N. [24], Davis, R. E. [42], Bradley, G. H. [23], Glover, F. [70], Koopmans, T. C. [96] and Trotter, L. E. [143], etc. Ideas of Gomory, R. E. [71,72], Balas, E. [8,9,10] and that of others were extended by IBM, to develop clever elimination schemes for solving covering problems. Branch-and-Bound method has turned out to be one of the most successful ways to solve practical integer programs.
Linear Fractional Programming problems were introduced by Charnes, A., Cooper, W. W. [30] in 1962. These types of problems consist of linear type fractional objective function and linear constraints. Other authors who contributed this field were Dorn, W. S., [44], Schaible, S. [133,134] etc.

For a long time it was not known whether or not, linear programs belonged to a non-polynomial class called “hard” (such as the one, the traveling salesman problem belongs to) or to an “easy” polynomial class (like the one that the shortest path problem belongs to). Klee, V. and Minty, G. J. [92] created an example that showed that the classical Simplex Algorithm would require an exponential number of steps to solve a worst-case linear program. In 1979, the Russian mathematician Khachian, L. G. [91] developed a polynomial-time algorithm for solving linear programs. It is an interior method using ellipsoids inscribed in the feasible region. He proved that the computing time is guaranteed to be less that a polynomial expression in the dimensions of the problem and the number of digits of input data. Although polynomial, the bound he established turned out to be too high for his algorithm to be used to solve practical problems.

Karmarkar’s algorithm [88,89] was an important improvement on the theoretical results of Khachian, L. G. [91] that a linear program can be solved in polynomial time. Moreover, his algorithm turned out to be one which could be used to solve practical linear programs. In 1990, interior algorithms were in open competition with variants of the Simplex Method. It appears likely that commercial software for solving linear programs would eventually combine pivot type moves used in the Simplex Methods with interior type moves.

The various forms of Nonlinear Programming further studied during the next decades are:

a) **Stochastic Programming** Stochastic Programming were developed in 1955 which has been greatly extended by Walkup, D. W. and Wets, R. J. B. [150]. Important contributions to this field have been made by Charnes, A. and Cooper, W. W. [29,31] in the late 1950’s using chance constraints, i.e., constraints which hold with a stated probability. The other authors who contributed in this field are Bawa, V. S. [17], Cocks, K. D. [33], Evers, W. H. [51], Garstka, S. J. [63] and Williams, A. C. [154] etc.

b) **Quadratic Programming** Quadratic programming was introduced by Wolfe, P. [61,157], who provided algorithm for positive definite and positive semi definite cases. Other authors who have contributed to this field are Beale, E. M. L. [19], Eaves, B. C. [48], Lemke, C. E. [105] and Van de Panne, C. [145,146] etc.

c) **Dynamic Programming** Discrete Dynamic Programming was developed in 1950’s through the work of Bellman, R. E. [20] and the continuous Dynamic Programming by White, L. S. [153]. Other
authors, contributing in this field are Howard, R. A. [81], Greenberg, H. J. [74] and Nemhauser, G. L. [117] etc.

**d) Geometric Programming** Geometric Programming had its beginning in 1967 at the hands of Duffin, R. J., Peterson, E. L., and Zener, C. [47]. Duffin developed the theory while Zener worked mostly on applications. The method initially was restricted to posynomials and less than or equal to one constraints because it was linked to the algebraic inequality of arithmetic and geometric mean. The extension of the method to negative terms and arbitrary inequalities was achieved by Passy, U. and Wild, D. J. [120] on the basis of lagrangian function and Kuhn-Tucker conditions.

### 1.1.1 Mathematical Programming in Statistics

In the development of statistical methods, one is often faced with an optimization problems. The techniques for solving such problems can be broadly classified as classical, numerical, variational methods and mathematical programming. The fundamental paper by Charnes, A., Cooper, W. W. et al. [28] introduced the application of mathematical programming to statistics. As an alternative to the least-squares approach to linear regression, they chose to minimize the sum of the absolute deviations (MINMAD) and showed the equivalence between the MINMAD problem and a linear programming problem. MINMAD estimator was studied as early as 1757 by Boscovitch, R. G. (later discussed by Eisenhart, C. [50]). Edgeworth, F. Y. [49] presented a method for the simple regression, with MINMAD estimator. However, Turner, H. H. [144] questioned Edgeworth’s claim of his method’s computational superiority over the least-squares method and also pointed out the non-uniqueness of the MINMAD estimator. It was only after the work of Charnes, A., Cooper, W. W. et al. [28], that a
renewed interest in using the MINMAD estimator for regression problems was created. They showed the equivalence between a MINMAD problem and a linear programming problem. Wagner, H. M. [148] suggested solving the problem through the dual approach. Afterwards, by the introduction of the efficient modification of the Simplex Method for solving the MINMAD problem by Barrodale, I. and Roberts, F. D. K. [13], the possibility has further increased of using MINMAD as an alternative to least squares.

The MINMAD problem with additional linear restrictions is considered along the same lines in Barrodale, I. and Robert, F. D. K. [14]. Special purpose algorithms for the MINMAD problem have also been given by Armstrong, R. D. and Hultz, J. W. [4], Bartels, R. H. and Conn, A. R. [15]. Computer comparisons have established the Barrodale, I. and Roberts, F. D. K. algorithms as an efficient method for solving the MINMAD problem. Revised Simplex version of this algorithm by Armstrong, R. D. et al. [5] is claimed to be even more efficient than the Barrodale, I. and Robert, F. D. K. algorithm.

Taylor, L. D. [142] suggested the combination of MINMAD and least square in which MINMAD should be applied first as a means of identifying outliers to be terminated and then the least squares applied after the trimming has been done. He also gives excellent arguments for the use of MINMAD in econometric analysis, Wilson, H. G. [155] via Monte Carlo sampling investigated the cases in which the disturbances are normally distributed with constant variance except for one or more outliers whose disturbance are generated from normal distribution with large variance. Among other results, it was found that MINMAD estimation retains its advantage over least squares under different conditions such as variations in outlier variance, number of independent variables, number of observations, and number of outliers.
Wagner, H. M. [148] gives a linear programming formulation of the MINMAXAD problem. He also suggests the dual approach for solving this problem. Steifel, E. [138] considers this problem and also brings out the connections between linear programming and Jordan elimination. Additionally, he gives examples to bring out the geometrical aspects of the problem. Collatz, L. and Wetterling, W. [34] also discuss this problem in the context of Chebyshev approximation theory.


The historical development of the theory of testing statistical hypotheses Lehmann, E. L. [104], and the fundamental work by Neyman, J. and Pearson, E. S. [119]. Barankin, E. W. [12] was the first to observe that linear programming might be used in testing of hypothesis. A good treatment of knapsack problems with exhaustive references can be found in Salkin, H. [131]. Schaafsma, W. [132] considers maximum tests and suggest the use of linear programming.

The mathematical programming approach to the generalized Neyman-Pearson problem has been considered by Francis, R. L. and Wright, G. [60], Meeks, H. D. and Francis, R. L. [113], Pukelsheim, F. [124] and Krafft, O. [97]. The result on the duality relationship between the Lagrangian problem and the primal problem is due to Francis, R. L. and Wright, G. [60]. Wagner,
D. H. [149] considered nonlinear functional variations of the Neyman-Pearson lemma, and discussed a number of applications of Neyman-Pearson problem.

Related problems in decision theory have also received the mathematical programming treatment. Weiss, L. [152] shows the use of a Simplex Method for solving minmax decision function. Similar duality results appear in Witting, H. [156], Krafft, O. and Witting, H. [98], Schaafsma, W. [132], Baumann, V [16] and Krafft, O. and Schmitz, N. [99].

Optimal allocation in stratified sampling has been considered by Dalenius, T. [36], Folks, J. L. and Antle, C. E. [55], Stock, J. S. and Frankel, L. R. [140], Ghosh, S. P. [73], Kokan, A. R. and Khan, S. [93] etc. Allocation of total sample size among different strata, when the sample means are required to have the sampling variance as much as possible in a given ratio so as reflect different degrees of importance in the various data, is considered by Chaddha, R. L. et al. [26]. Approximating nonlinear separable objective function by piecewise linear function and formulating the problem as a restricted linear programming problem are given by Hadley, G [76] also discussed the sufficiency of solving the problem as a linear programming problem when the objective function is convex.

experiments with resource constraints; Foody, W. and Hedayat, A. [56] in the construction of BIB design with repeated blocks.

Regression analysis includes techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

The earliest form of regression ever known was the method of least squares introduced by Legendre, A. M. [103] in 1805 and by Gauss, C. F. [65] in 1809. The term "regression" was coined by Francis, G. [58] in the nineteenth century to describe a biological phenomenon. The phenomenon dealt with the heights of descendants of tall ancestors that tend to regress down towards a normal average (a phenomenon also known as regression towards the mean) Mogull, R. G. [115], Francis, G. [59]. For Francis, regression had only this biological meaning [57, 58], but his work was later extended by Pearson, K., Yule, G. U. et al. [121] and to a more general statistical context, Yule, G. Udny. [160]. In the work of Pearson, K., Yule, G. U. the joint distribution of the response and explanatory variables is assumed to be Gaussian. This assumption was weakened by Fisher, R. A. in his works of 1922 and 1925, Fisher, R.A. [53, 54] and Aldrich, J. [2]. Fisher assumed that the conditional distribution of the response variable is Gaussian, but the joint distribution need not be Gaussian. In this respect, Fisher's assumption is closer to Gauss's formulation [66].

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be
used to infer causal relationship between the independent and dependent variables.

A large number of techniques for carrying out regression analysis has been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be finite-dimensional.

Mathematical optimization is used as an aid to a human decision maker, system designer, or system operator, who supervises the process, checks the results, and modifies the problem (or the solution approach) when necessary. This human decision maker also carries out any actions suggested by the optimization problem, e.g., buying or selling assets to achieve the optimal portfolio.

In data fitting, the task is to find a model, from a family of potential models, that best fits some observed data and prior information. Here the variables are the parameters in the model, and the constraints can represent prior information or required limits on the parameters (such as non-negativity).

Mathematical Programming plays a vital role in engineering application like, Design of aircraft, optimal trajectory of space vehicles, Optimum design of electric network, Planning the best strategy to obtain maximum profit, etc.

The commercial applications of non-linear Programming were made in 1952 by Charnes, A., Cooper, W. W. et al. [27] with their optimal blending of petroleum products to make gasoline. Applications quickly
spread to other commercial areas and soon eclipsed the military applications which started the field.

Applications of mathematical programming are everywhere, many real-life problems can be converted into mathematical form, Arnold, N. [6] and optimized by using the techniques of mathematical programming.

In our economy, we always try to maximize profit and sales, and costs should be as low as possible. Therefore, optimization is one of the oldest of sciences which even extends into daily life, Arnold, N. [6].

Mathematical programming also plays a vital role in statistics, many problems in regression analysis, sample surveys, cluster analysis, construction of designs, estimation, decision theory and so on can be viewed as mathematical programming problems.

1.1.1.1 Estimation Problem

The general problem of estimation is one of choosing a density function belonging to a specified family of density functions. on the basis of observed data. For this purpose a function of observations called “estimator” is defined so that the value of the estimator for a given observed datum is the estimate of the unknown density function. When interested only in estimating certain parameters of the density function. We might not estimate the entire density function. Such problems are indeed optimization problems, Rao, C. R. [127].

The problem of testing statistical hypotheses was considered originally by Neyman, J. and Pearson, E. S. [119]. Connections between the well-known Neyman-Pearson Lemma for constructing the uniformly most powerful test of a sample hypothesis having a singly alternative, and optimization with linear models can be seen from the following passage from Dantzing’s Linear Programming and Extensions [41].
1.1.1.2 Sample Survey Methodology

Since there is a need for reliable data to understand the world, statisticians have to devise methods of collecting such data. Information on a population may be collected either by complete enumerations or by sample enumeration. The cost of conducting sample enumeration is in general less than that of complete enumerations. On the other hand, precision suffers when too small a sample is considered. Thus the fundamental problem in sample survey is to choose a sampling design that either assures the maximum precision for a given cost of the survey or assumes the minimum cost for a given level of precision. Thus the root of sample-survey methodology lies an optimization problem of considerable importance. Similarly, the consideration behind the statistical design versus efficiency of the design chosen.

Mathematical programming problems have received the attention of researchers in mathematics, economics and operations research for over three decades. Since the development of Simplex Method for efficiently solving the linear programming problem, both the theory and the methods of mathematical programming have seen unprecedented growth. Also the emphasis has turned for solving certain problems, towards finding efficient methods suitable for computers.

1.2 Preliminaries and Definitions

Mathematical Programming may be described in terms of its mathematical structures and computational procedures or in terms of the broad class of important decision problems, problems having certain aim/goal which needs to be optimized under certain conditions that can be formulated/converted into mathematical functions of several variables
(defined decision variables). The word – programming used in mathematical programming is not to be confused with programming as used for the task of preparing a sequence of instructions for a computer but its origin lies in planning/scheduling the quantity and timing of the various activities of an organization such as a factory, national economy or world trade etc.

A mathematical programming problem can be stated as:

\[(\text{MPP}): \quad \text{Maximize } Z = f_0(x) \]

subject to

\[f_i(x) \leq b_i, \quad i = 1, \ldots, m.\]
\[x_j \geq 0 \quad j = 1, \ldots, n.\]

Here the vector \( x=(x_1, \ldots, x_n) \) contains the unknown decision variables of the problem, the function \( f_0 : R_n \rightarrow R \) is the objective function, the functions \( f_i : R_n \rightarrow R, \quad i = 1, \ldots, m, \) are the constraint functions, and the constants \( b_1, \ldots, b_m \) are the limits or bounds for the constraints. The values of the decision variables that satisfy the set of all constraints and non-negative conditions will form the set of basic feasible solutions known as the feasible region of mathematical programming problem. A vector \( x \) is called optimal, or a solution of the problem (MPP), if it has the greatest objective value among all vectors that satisfy the constraints; for any \( z \) with \( f_i(z) \leq b_i, \ldots, f_m(z) \leq b_m \), we have \( f_0(z) \geq f_0(x) \) with non-negativity condition included.

Since a maximization problem can be expressed in a minimization problem, so we will consider only maximization problem throughout the work.
1.2.1 Linear Programming

The mathematical programming (MPP) is called linear programming problem if the objective function and the constraints both are linear:

\[(\text{MPL}): \quad \text{Maximize} \quad Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n,\]

Subject to

\[a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1,\]
\[a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2,\]
\[a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m,\]
\[x_1 \geq 0, \; x_2 \geq 0, \ldots, x_n \geq 0,\]

where \(c_j, a_{ij}\) and \(b_i\) are given constants. \((i = 1, \ldots, m, \; j = 1, \ldots, n.\)

The values of the decision variables \(x_1, x_2, \ldots, x_n\) that satisfy all the constraints of (MPL) and non negativity conditions simultaneously are said to form a feasible solution to the linear programming problem. Any solution belonging to feasible region is called a **basic feasible solution** to (MPL). A feasible solution which in addition optimizes the objective function of (MPL) is called an **optimal solution** of the problem. There are three possible situations which may emerge.

i) The linear program could be infeasible, meaning that there are no values of the decision variables, \(x_1, x_2, \ldots, x_n\) that simultaneously satisfy all the constraints.

ii) It could have an unbounded solution, meaning that, if we are maximizing the value of the objective function, then the value of
the objective function can be increased indefinitely without violating any of the constraints or if we are minimizing, the value of the objective function may be decreased indefinitely. Such problems usually are in fact poorly formulated.

iii) In most cases, it will have at least one finite optimal solution, in case it has many optimal solutions then we call these solutions as alternate optimal solutions or multiple optimal solutions.

1.2.2 Multiobjective Linear Programming

The mathematical programming problem (MPP) is called Multiobjective Linear Programming problem if the problem has more than one linear objective function that are to be optimized simultaneously and constraints are also linear functions with non negativity conditions included.

(MOLP): Maximize $Z = \{f_0(x_1), f_0(x_2), \ldots, f_0(x_n)\}$

subject to

$f_1(x_1, x_2, \ldots, x_n) \leq b_1,$

$f_2(x_1, x_2, \ldots, x_n) \leq b_2,$

\ldots

\ldots

\ldots

$f_m(x_1, x_2, \ldots, x_n) \leq b_m,$

$x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0.$

If the maximum values of all $f_0(x)$ are achieved at the same point $X^*$, then $X^*$ is called an ideal solution of the problem. An ideal solution is the solution with all the objective functions simultaneously maximize. However, in real life problems such a situation would be very rare.
In the absence of an ideal solution, we may prefer a solution regarded best by some other suitable criterion. One such criterion is to look for a point from where the value of any of the objective functions cannot be decreased without increasing the value of at least one of the other objective functions. In other words, to find a feasible point $X^*$ such that there is no other feasible point in moving to which one of the objective functions decreases and none of the other objective functions increase. Such a point is called an efficient point, and the corresponding solution an efficient solution.

### 1.2.3 Non-linear Programming

Another fundamental extension of the problem (MPP) is to allow the objective function or the constraints or both to be nonlinear functions.

\textbf{(MPNLP):} Maximize $Z = f_0(x_1, x_2, \ldots, x_n)$

subject to

\begin{align*}
  f_1(x_1, x_2, \ldots, x_n) & \leq b_1, \\
  f_2(x_1, x_2, \ldots, x_n) & \leq b_2, \\
  \vdots & \vdots \\
  \vdots & \vdots \\
  f_m(x_1, x_2, \ldots, x_n) & \leq b_m, \\
  x_1 & \geq 0, x_2 \geq 0, \ldots, x_n \geq 0.
\end{align*}

Often in nonlinear programming the right hand-side values are included in the definition of the function $f_i(x_1, x_2, \ldots, x_n)$, leaving the right hand side to be zeroes. In order to solve a nonlinear programming problem, some assumptions must be made about the shape and behavior of the functions involved the nonlinear functions must be rather well-behaved in
order to have computationally efficient means of finding a solution. The nonlinear programming with one objective function as above is called single objective nonlinear programming.

1.2.4 Multiobjective Nonlinear Programming

The mathematical programming (MPP) is called Multiobjective Nonlinear Programming problem if the set of objective functions \( f_0(x) \) or the constraints or both are nonlinear functions which are to be optimized. Multi-objective optimization (or programming), also known as multi-criteria or multi-attribute optimization, is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints and non negativity conditions.

\[ \text{(MONLP): Maximize } \mathbf{Z} = \{ f_0(x_1), f_0(x_2), \ldots, f_0(x_n) \} \]

subject to

\[
\begin{align*}
    f_1(x_1, x_2, \ldots, x_n) & \leq b_1, \\
    f_2(x_1, x_2, \ldots, x_n) & \leq b_2, \\
    \vdots \ & \vdots \ & \vdots \\
    \vdots \ & \vdots \ & \vdots \\
    f_m(x_1, x_2, \ldots, x_n) & \leq b_m, \\
    x_1 & \geq 0, x_2 \geq 0, \ldots, x_n \geq 0.
\end{align*}
\]

Application of multiobjective optimization can be found in various fields like, product and process design, finance, aircraft design, the oil and gas industry, automobile design, or wherever optimal decisions need to be taken in the presence of trade-offs between two or more conflicting objectives. Maximizing profit and minimizing the cost of a product;
maximizing performance and minimizing fuel consumption of a vehicle; and
minimizing weight while maximizing the strength of a particular component
are examples of multi-objective optimization problems.

1.3 Regression Approach

In linear regression, the model specification is that the dependent
variable, \( y_i \) is a linear combination of the parameters (but need not be linear
in the independent variables). For example, in simple linear regression for
modeling ‘n’ data points there is one independent variable \( x_i \), and two
parameters, \( \beta_0 \) and \( \beta_1 \):

\[
Y_i = \beta_0 + \beta_1 x_i + e_i \quad i = 1, 2, \ldots, n.
\]

In the more general multiple regression model, there are ‘p’ independent
variables:

\[
Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i,
\]

where \( x_{ij} \) is the \( i^{th} \) observation on the \( j^{th} \) independent variable, and the first
independent variable takes the value 1 for all \( i \) (so \( \beta_1 \) is the regression
intercept).

The least squares parameter estimates are obtained from \( p \)- normal
equations. The residual can be written as

\[
e_i = Y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip},
\]

The normal equations are

\[
\sum_{i=1}^{n} \sum_{k=1}^{p} X_{ij} X_{ik} \beta_k = \sum_{i=1}^{n} X_{ij} Y_i, \quad j = 1, \ldots, p.
\]

In matrix notation, the normal equations are written as
\( (X^T X) \hat{\beta} = X^T Y, \)

where the \( i^{\text{th}} j^{\text{th}} \) element of \( X \) is \( x_{ij} \), the \( i \) element of the column vector \( Y \) is \( y_i \), and the \( j \) element of \( \hat{\beta} \) is \( \hat{\beta}_j \). Thus \( X \) is \((n \times p)\), \( Y \) is \((n \times 1)\), and \( \hat{\beta} \) is \((p \times 1)\).

and \( \hat{\beta} = (X^T X)^{-1} X^T Y. \)

### 1.3.1 Coefficient of Determination

Coefficient of determination (R\(^2\)) is a statistical measure that will give some information about the goodness of fit of a model. In regression, the R\(^2\), coefficient of determination is a statistical measure of how well the regression line approximates the real data points. The coefficient of determination R\(^2\) is a measure of the global fit of the model. Specifically, R\(^2\) is an element of \([0, 1]\) and represents the proportion of variability in \( Y_i \) that may be attributed to some linear combination of the regressors (explanatory variables) in \( X \).

In case of the linear model of the form

\[ Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i, \]

where, for the \( i^{\text{th}} \) case, \( Y_i \) is the response variable, \( X_{i1}, \ldots, X_{ip} \) are \( p \) regressors, and \( \epsilon_i \) is a mean zero error term. The quantities \( \beta_0, \ldots, \beta_p \) are unknown coefficients, whose values are determined by least squares.

R\(^2\) is often interpreted as the proportion of response variation "explained" by the regressors in the model. Thus, \( R^2 = 1 \) indicates that the fitted model explains all variability in \( y \), while \( R^2 = 0 \) indicates no 'linear'
relationship (for straight line regression, this means that the straight line model is a constant line (slope=0, intercept=$\bar{y}$) between the response variable and regressors.

A caution that applies to $R^2$, as to other statistical descriptions of correlation and association is that "correlation does not imply causation." In other words, while correlations may provide valuable clues regarding causal relationships among variables, a high correlation between two variables does not represent adequate evidence that changing one variable has resulted, or may result, from changes of other variables.

In case of a single regressor, fitted by least squares, $R^2$ is the square of the pearson product–moment coefficient relating the regressor and the response variable. More generally, $R^2$ is the square of the correlation between the constructed predictor and the response variable.

A data set has values $y_i$, each of which has an associated modeled value ‘$f_i$’ (also sometimes referred to as $\hat{y}_i$). Here, the values $y_i$ are called the observed values and the modeled values ‘$f_i$’ are sometimes called the predicted values.

The "variability" of the data set is measured through different sums of squares:

$$SS_{tot} = \Sigma (y_i - \bar{y})^2,$$

This is the total sum of squares (proportional to the sample variance);

$$SS_{reg} = \Sigma ((f_i - \bar{y})^2,$$
This is the regression sum of squares, also called the explained sum of squares.

\[ SS_{err} = \sum_i (y_i - \hat{f}_i)^2, \]

This is the sum of squares of residuals, also called the residual sum of squares.

In the above \( \bar{y} \) is the mean of the observed data:

\[ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]

where \( n \) is the number of observations.

The most general definition of the coefficient of determination is

\[ R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \]

### 1.4 Algorithms for Solving Mathematical Programming Problems

Generally, optimization algorithms can be divided into two basic classes: deterministic and probabilistic algorithms. Deterministic algorithms are most often used if a clear relation between the characteristics of the possible solutions and their utility for a given problem exists. Then, the search space can efficiently be explored using, for example, a divide and conquer scheme. If the relation between a solution candidate and its “fitness” are not so obvious or too complicated, or the dimensionality of the search space is very high, it becomes harder to solve a problem deterministically. Trying it would possibly result in exhaustive enumeration of the search space, which is not feasible even for relatively small problems. Thus, probabilistic algorithms come into play their role. The initial work in this area which now has become one of most important research field in optimization was started in 1958 by Richard, M. F. [128], Woodrow, W. et
al. [159], Hans, J. Bremermann and others [79,77]. An especially relevant family of probabilistic algorithms are the Monte Carlo-based approaches. They trade in guaranteed correctness of the solution for a shorter runtime. This does not mean that the results obtained using them are incorrect- they may just not be the global optima.

Various Techniques are used to solve mathematical programming problems, the techniques used depends upon the various factors like whether the problem has constraints or not, whether the problem is linear or nonlinear, whether the objective function and constraints are differentiable or not and so on. Some of the techniques mostly used are:

1.4.1 Classical Techniques

The classical optimization techniques are useful in finding the optimum solution of unconstrained maxima or minima of continuous and differentiable functions. These are analytical methods and make use of differential calculus in locating the optimum solution. The classical methods have limited scope in practical applications as some of them involve objective functions which are not continuous and/or differentiable. Yet, the study of these classical techniques of optimization form a basis for developing most of the numerical techniques that have evolved into advanced techniques more suitable to today’s practical problems. These methods assume that the function is differentiable twice with respect to the decision variables and the derivatives are continuous.

Three main types of problems can be handled by the classical optimization techniques:

i) single variable functions

ii) multivariable functions with no constraints

iii) multivariable functions with both equality and inequality constraints
In problems with equality constraints, the Lagrange multiplier method can be used. A necessary condition for a function $f(x)$ subject to constraints $g_i(x) = 0$, $i=1, 2, \ldots, m$ to have a relative minimum at a point $X^*$ is that the first partial derivative of the Lagrange function defined by $L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m)$ with respect to each of its argument must be zero.
A sufficient condition for \( f(x) \) to have a relative minimum at \( x^* \) is that the quadratic form ‘\( Q \)’ defined by

\[
Q = \sum_{i=1}^{n} \sum_{j=1}^{n} (d^2 L/dx_idx_j)d_{xi}d_{xj}
\]  

(1.1)

evaluated at \( x = x^* \) must be positive definite for all values of \( d_x \) for which the constraints are satisfied. If \( Q = \sum_{i=1}^{n} \sum_{j=1}^{n} (d^2 L/dx_idx_j)(x^*, \lambda)d_{xi}d_{xj} \) is negative for all choices of the admissible variations \( d_{xi} \), \( x^* \) will be a constrained maximum of \( f(x) \).

The necessary condition for the quadratic form \( Q \), defined by (1.1), to be positive (negative) definite for all admissible variations \( d_x \) is that each root of the polynomial, \( z_i \), defined by the following determinantal equation be positive (negative):
The expansion leads to a \((n-m)^{th}\) order polynomial in \(z\). If some of the roots of this polynomial are positive while the others are negative, the point \(x^*\) is not an extreme point.

If the problem has inequality constraints, the Kuhn-Tucker conditions [100] can be used to identify the optimum solution, the inequality is converted into equations by using nonnegative slack variables, \(S_i^2 (\geq 0)\) be the slack quantity added to the \(i^{th}\) constraints \(g_i(x) \leq 0\) and \(S^2 = (s_1^2, s_2^2, \ldots, s_m^2)^T\).
where ‘m’ is the total number of inequality constraints. The Lagrangean function is thus given by

\[ L(X, S, \lambda) = f(x) - \lambda [g(x) + S^2] \]

given the constraints \{g(x)\} are \( \leq 0 \)
a necessary condition for optimality is that \( \lambda \) be nonnegative (nonpositive) for maximization (minimization) problems. The justification about this is that the \( \lambda \) measures the rate of variation of \( f(x) \) with respect to \( g(x) \) that is

\[ \lambda = \frac{\partial f}{\partial g} \]
as the right hand side of the constraint \( g(x) \leq 0 \) (maximization case) increases above zero, the solution space becomes less constrained and hence \( f(x) \) cannot decrease. This means that \( \lambda \geq 0 \). Similarly for minimization case, as a resource increase, \( f(x) \) cannot increase, which implies that \( \lambda \leq 0 \). If the constraints are equalities, that is \( g(x) = 0 \) then \( \lambda \) becomes unrestricted in sign.

The restrictions on \( \lambda \) must hold as part of the Kuhn-Tucker necessary conditions. Also

\[ \frac{\partial L}{\partial x} = \nabla f(x) - \lambda \nabla g(x) = 0 \]

\[ \frac{\partial L}{\partial x} = -2\lambda_i S_i = 0, \quad i = 1, 2, \ldots, m \]

\[ \frac{\partial L}{\partial x} = -(g(x) + S^2) = 0 \]
It means that

I. If $\lambda_i$ is not zero, then $S_i^2 = 0$. This means that the corresponding resource is scarce, and consequently it is consumed completely (equality constraints).

II. If $S_i^2 > 0$, $\lambda_i = 0$. This means the $i^{th}$ resource is not scarce and, consequently, it does not affect the value of $f(x)$ (i.e., $\lambda_i = \frac{\partial f}{\partial g_i(x)} = 0$).

The Kuhn-Tucker necessary conditions are also sufficient if the objective function and the solution space satisfy certain conditions regarding convexity and concavity. These conditions are in Table 1.3

<table>
<thead>
<tr>
<th>Sense of optimality</th>
<th>Required conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Objective function</strong></td>
</tr>
<tr>
<td><strong>Maximization</strong></td>
<td>Concave</td>
</tr>
<tr>
<td><strong>Minimization</strong></td>
<td>Convex</td>
</tr>
</tbody>
</table>

Table 1.3
1.4.2 Methods for Linear Programming

Various methods are available that are used to solve linear programming problems like graphical methods especially for two dimensional cases. Simplex Method for n-dimensional cases and Karmarker Method [88,89] etc. are some of the well known methods for linear cases. The method proposed by Karmarkar, N. [88,89] is an interior point algorithm. The Simplex Method [41] is an iterative procedure for solving a linear program and provides all the information about the program. Also, it indicates whether or not the program is feasible and if the program is feasible, it either provides an optimal solution or indicates that an unbounded solution exists.

The Ellipsoid Method established by Khachian, L. G. [91], on the other hand, investigates the interior points of a feasible region until it reaches an optimal point on the boundary, and is essentially an interior point algorithm. The basic idea for Karmarker Algorithm is to use the steepest descent method. When the objective is to find the minimum of the problem:

I) It is advisable to move in the direction of steepest descent if the current (approximate) interior point is near the centre of the polytope describing the feasible region.

II) It is possible to transform the feasible region so as to place the current point near the centre of the polytope, without changing the problem in any essential way.

Optimization models can be the subject of various classifications depending on the point of view, we adopt according to the number of time periods considered in the model, optimization models can be classified as astatic (single time period) or multistage (multiple time periods). Even when
all relationships are linear and several time periods are incorporated in the model, the resulting linear program could become prohibitively large for solution by standard computational methods. Fortunately, in most of these cases, the problem exhibits some form of special structure that can be adequately exploited by the application of special types of mathematical programming methods. Dynamic programming is one approach for solving multistage problems. Further, there is a considerable research effort underway today, in the field of large-scale linear programming, to develop special algorithms to deal with multistage problems.

1.4.3 Methods for Non-linear Programming

Like linear programming problems it is difficult to solve nonlinear programming problems. Our main aim is to search for the optimal points (local or global). Lagrange multipliers method is a special case of the more general optimality conditions of the so called Kuhn-Tucker (K-T) conditions [100].

The K-T conditions from the basis of many algorithms for nonlinear programming problems. The necessary K-T conditions [100] are primarily useful in the negative sense. In other words, if a point does not satisfy them, it cannot then be a optimal solution.

Several factors have to be considered in deciding a particular method to solve a given nonlinear programming problem. Some of them are:

i) The type of problem to be solved (general nonlinear programming problem, geometric programming, etc).

ii) The availability of a readymade computer programme.

iii) The calendar time necessary for the development of a programme.

iv) The necessity of derivatives of the functions f and g.
v) The available knowledge about the efficiency of the method.

vi) The accuracy of the solution desired.

vii) The programming language and quality of coding desired.

viii) The dependability of the method in finding the true optimum solution.

ix) The generality of the programme for solving other problems.

x) The ease with which the programme can be used and its output interpreted.

For problems involving explicit (nonlinear) expressions for ‘f’ and ‘g’, and small or moderate number of variables, the use of penalty function algorithms are expected to work most efficiently. Out of these, the interior penalty function algorithms is more efficient. The exterior penalty function algorithms will be less efficient since even a feasible starting point \((X_1)\) leads to an infeasible point \((X_1^*)\) at the end of the minimization of function. As the sequence of the points \(X_1^*, X_2^*, \ldots, X_n^*\) lies in the infeasible region, and approaches the optimum point and feasibility simultaneously, this algorithm is useful only when a starting feasible point \((X_1^*)\) cannot be found. If all the constraints of the optimization problem are linear, the gradient projection method will be the best one. If the problem involves f and g that are implicitly dependent on the design vector \(X\) (i.e. an analysis is needed to evaluate ‘f’ and ‘g’), the derivatives of the functions ‘f’ and ‘g’ cannot be obtained in closed form. If these derivatives can be obtained by finite difference formulas, the Zoutendijk’s algorithm \([161]\) of feasible directions will be more efficient than the penalty function algorithm. However, if one intends to use approximations in evaluating ‘f’ and ‘g’, the evaluation of ‘f’ and ‘g’
is extremely difficult and if one is interested in finding only a near optimal solution, the interior penalty function algorithm is the obvious choice.

### 1.4.4 Methods for Quadratic Programming

When the objective function of mathematical programming problem (MPP) is quadratic and the constraints are linear the mathematical programming is called quadratic programming problem. i.e.

\[(QP1): \quad f_0(x) = C^tX + \frac{1}{2} X^tGX\]

subject to

\[f_i(x) \leq b_i,\]

\[i = 1, \ldots, m \quad \text{and} \quad G \text{ is } (n \times n) \text{ matrix.}\]

The matrix G can be taken as non-null since otherwise (QP) is a linear program. Also without loss of any generality, G can be taken as a symmetric matrix since \(X^tGX = X^t \frac{1}{2} (G + G^t)X\) and \(\frac{1}{2} (G + G^t)\) is symmetric. The form \(X^tGX\) is convex if and if G is a positive semi-definite matrix and it is strictly convex if and if G is positive definite. Thus, a quadratic program is convex when G is a positive semidefinite matrix. Some of the important Algorithms for quadratic programming are:

#### 1.4.4.1 Beale’s Algorithm

Beale’s algorithm [19] uses the classical calculus results rather than Kuhn-Tucker conditions, it is applicable to any quadratic program of the form:

\[(QP2): \quad f_0(x) = C^tX + \frac{1}{2} X^tGX\]

subject to

\[Ax = b\]
where ‘A’ is (m x n), ‘b’ is (m x 1), ‘C’ is (n x 1) and ‘G’ is symmetric (n x n) matrix.

The Beale method requires partitioning of the variables into basic and nonbasic variables at each iteration and expressing the objective function in terms of only the nonbasic variables.

1.4.4.2 Fletcher’s Algorithm

The Fletcher algorithm can be viewed as belonging to the general class of methods known as feasible direction methods derived by Zountendijk. G., [161] it can also be regarded as a special case of the projected gradient method for solving linearly constrained problems.

Fletcher’s algorithm is an iterative method that uses an active set strategy. An active set is a list of those constraints that are satisfied as equalities during an iteration. The algorithm generates a sequence of equality constrained quadratic programs which differ only in active constraints, usually in that a constraint is either added to or removed from the set of active constraints.

1.4.5 Methods for Separable Programming

Mathematical programming problem (MPP) is called separable programming if the objective function $f_0(x)$ of the mathematical programming problem can be expressed as the sum of ‘n’ single variable functions.

$$f_0(x) = f_0(x_1) + f_0(x_2) + f_0(x_3) + \ldots + f_0(x_n)$$

Separable programming deals with nonlinear problems in which the objective function and the constraints are separable, i.e. we have to divide
the nonlinear functions into individual parts. The idea is to construct a constrained optimization model that linearly approximates the original problem. The approximations enlarge the size of the technique, the method has considerable practical significance Miller, C.E. [114] as the approach can be used equally well to approximate a nonlinear objective function and nonlinear constraints.

1.5 Software’s for Mathematical Programming

A large number of software’s are available to solve the different types of mathematical programming. Some of important software used for solving the mathematical programming problems are:

LP-Optimizer is a simplex-based code for linear and integer programs, written by Markus Weidenauer.

SoPlex is an object-oriented implementation of the primal and dual Simplex Algorithms, developed by Roland Wunderling.

EXLP solves linear programs of moderate size in exact rational arithmetic, using the GNU Multiple Precision Arithmetic Library.

PCX The Optimization Technology Center at Argonne National Laboratory and Northwestern University has developed PCx, an interior-point code.

GIPALS Denis Smirnov has developed GIPALS, an environment that incorporates an interior-point solver and simple graphical user interface.

LINGO, LINDO API The methods used for these softwares are Generalized Reduce Gradient, Successive (sequential) linear Programming and global search.