CHAPTER 3
RESPONSE SURFACE METHODOLOGY AND GENETIC ALGORITHM

3.1. Introduction of Response Surface Methodology (RSM)

Box and Wilson were began the usage of Response Surface Methodology (RSM) in the Journal of Royal Statistical Society (1951). This tool RSM was used in their investigation to establish the number of experiments capably with suitable design matrix and finally concludes the best set of optimal responses with manageable process parameters. The main aim of the RSM is to develop the objective function using the different models like Linear, 2F, first degree and second degree polynomial models with the assumption of constant output responses. From 1970 onwards RSM is used in the various applications and different experimentation conditions [42].

RSM is a group of mathematical and statistical approach used for various models to analyze and optimize the process variables. The main objective with the RSM is the optimization of output variable. This statistical method is applied to analyze the data obtained from the experiments pertained to any problem because the outputs are always influenced by several input parameters. After analysis it helps in modelling the problem also [43].

For, instance, if any researcher would need to know the levels for control variables \( x_1, x_2 \) in order to yield the maximum value \( (y) \) of the process, then its process is shown as

\[
y = f(x_1, x_2) + \varepsilon \tag{3.1}
\]

Here \( \varepsilon \) = error of the system.

If the estimated response is written as \( E(y) = f(x_1, x_2) = \eta \), then the response surface is denoted as \( \eta = f(x_1, x_2) \) in case of response curve, single control factor relates to output response.

If two control variables relate to the output response, it will be termed as response surface [43]. The common practice is representation of response surfaces in the form of plots (i.e. contours, individual plots & 3D Plots).
3.2. Assumptions

In the process, several assumptions are applied and listed below.
1. Existence of some type of relationship (function) between the control factors and the responses.
2. The relationship or the function (s) is moreover unidentified or complicated.
3. The control factors are quantitative, continuous, controlled closely and can be measured with accuracy in the experimental zone with negligible error value.
4. All the parameters in the system significantly meaningful and also the good indicators of system’s performance.

3.3. Procedure

This RSM process comprises of the following steps written below [44]:
1. To measure the response of interest with adequacy, a series of experiments are to be designed.
2. Secondly empirical or numerical model has developed to predict the models with best fittings.
3. To find the best set (optimal) of experimental parameters that can produce minimum or maximum values of the responses.
4. Plotting the main and the interaction effects of the process parameters on response in the form of graphical representations using 2D and 3D graphs.

3.4. Experimental Design

After identifying the problem, then develop the design for proposed experimental plan. In general, the majority of engineering investigations are to be experimental and these investigations require more number of experiments. During the period of conducting the trail experiments with design matrix some errors may contains in the collected data as the experimentation is effected by various non-controllable factors. For getting the appropriate conclusions from the experiment as well as to avoid and to minimize the errors the statistical approach is very needful.

Many researchers have studied several types of designs and applied to varieties of problems in all areas mainly in the manufacturing sector. For example, fractional factorial design, Box-Behnken designs, A Optimal, D Optimal, G and V-
Optimal Designs, CCD (central composite design) are the designs used by several researches. The design is chosen depending upon the requirement of the user.

Box and Draper (1987) demonstrated more detailed information about DOE in [46], according to Myers and Montgomery (1995) and Montgomery (2003) [47]. From previous study, it is observed that the CCD is the best experimental design. In the present research work, CCD is applied to plan of the experiments.

3.4.1. Central Composite Design

For best fitting of the second order models, CCD is very suitable and has to be used effectively [47]. Central composite design (Box-Wilson CCD) will have fractional factorial design with centre points and are augmented by the star point groups, which allow the curvature estimations.

If ±1 unit is the distance between the factorial points in the design space centre for each factor, then the distance from the star point to design space centre is ±α with |α| > 1. The accurate value of ‘α’ is depending on design properties and the number of various factors influenced.

Always CCD contains twice as many as star points as factors present in the designs. These star points represent minimum and maximum values for all the factors present in the design. Usually, central composite design consists of a $2^k$ factorial. Where $n_f$ factorial points, $n_a$ star or axial points, and $n_c$ centre points are available. The selected matrix consists of three levels & three-factor design, which contains $L_{27}$, sets of conditions for simulation work and $L_9$ for experimentation.

The design possesses full replication of 24(=16) factorial design points plus star points (8 Nos) and centre points (7 Nos). All the factors are coded and centre point constituted at intermediate level “0” maximum and minimum levels are coded with “-1” and “+1” respectively. Thus, the total chosen experimental trials will allow the estimation of the quadratic, linear and 2-way interactive effects of the control factors on the output responses.

It is essential to produce the good predictions for the second order models. Box and Hunter [48] recommended having 2$^{nd}$ order design a rotatable one. It can be rotatable if the variance of the response (predicted one) at chosen point x will depend on the distance of x from centre point of design.
The term rotatability is the most desirable property in the response surface design selections particularly for quadratic model. Generally, the optimal value is unknown before conduction of the experiments as per the design. Thus the selected design has to provide the provision of estimation in all of the directions. The RSM provides us the advantage of reducing the number of experiments that are required to achieve the desired accuracy. Therefore, for modelling by RSM could be the better option for develop the empirical model.

3.5. Building of Empirical Models

Many cases empirical modelling is done by RSM method. Model for each response is predicted using the data obtained from the experiments and regression analysis is done for the models. Representation of models in quantitative form is possible with RSM and as shown as:

\[ Y = f(X_1, X_2, X_3, \ldots, X_n) \pm \varepsilon \]  \hspace{1cm} (3.2)

In the above equation, \( Y \) = Output response  
\( f \) = Response function  
\( \varepsilon \) = noise or error the experimental data and  
\( X_1, X_2, X_3, \ldots, X_n \) are control variables in the process.

The response surface is obtained by plotting the expected responses. The form of \( f \) is unknown and complicated.

Thus, response surface methodology objective at approximating \( f \) by a suitable lower order polynomial in some region of the control factors. The function shown in (eq. 3.2) is modified and rewritten as:

\[ Y = C_0 + C_1X_1 + C_2X_2 + \ldots + C_nX_n \pm \varepsilon \]  \hspace{1cm} (3.3)

However, the higher order polynomial i.e. quadratic model is applied if curvature presents in the system.

\[ Y = C_0 + \sum_{i=1}^{n} C_iX_i + \sum_{i=1}^{n} d_iX_i^2 \pm \varepsilon \]  \hspace{1cm} (3.4)

The goal of RSM is to examine the output responses over the entire factor space region. Also used for locating the region of interest that may reach its optimal or near optimal values.
3.6. Estimation of Parameters in Regression Models

In a multiple linear regression model, the regression coefficient values are calculated for each model by least squares method.

Let \( y_1, y_2, \ldots, y_n \) = response surface observations

Moreover, each regression variable gives one response corresponding to the response surface variation. The model is represented in the form of matrix notation i.e

\[
y = X\beta + \epsilon
\]

Where \( y = \text{Vector of observations (} n \times 1 \text{ matrix)} \)

\( X = n \times p \text{ matrix} \)

\( \beta = \text{regression coefficients vector for matrix (} p \times 1 \text{)} \)

\( x_{ij} = i^{th} \text{ observation}. \)

\( \epsilon = \text{Random errors vector (} n \times 1 \text{ matrix)} \)

The formula for least squares estimators is

\[
L = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon' \epsilon = (y - X\beta)'(y - X\beta)
\]

And after simplification it becomes

\[
b = (X'X)^{-1}X'y
\]

\((X'X) = \text{symmetric matrix (} p \times p \))

\((X'y) = p \times 1 \text{ column vector.}\)

The \((X'X)\) matrix possesses the special structure. i.e. Diagonal elements is equal to the summation of squares in the X column. The summation of cross products gives the off – diagonal elements.
The regression model after fitting takes the form:

\[ \hat{y} = Xb \]  \hspace{1cm} (3.8)

In scalar notation, the fitted model is written in scalar notation as

\[ \hat{y}_i = \sum_{j=1}^{k} b_j x_{ij}, \quad i = 1, 2, \ldots, n \]  \hspace{1cm} (3.9)

The difference \( e_i = y_i - \hat{y}_i \) is a residual having \( n \times 1 \) vector. Where \( y_i \) is the observation and \( \hat{y}_i \) is the fitted value.

The regression coefficients are calculated using the least squares method for all the models. In this method, the coefficients of regression function are predicted by determining the mathematical values for the control variable that reduce the deviations among the experimental output and the estimated model value. Generally, the predicted coefficients values are more or less equal to experimental value.

**3.7. Checking of Model Adequacy**

Once the models are developed then adequacy tests have to be accomplished before using it further to verify the model suitability for the data. This is done for both the second order, quadratic or first order predicted models. Model validation is done with many of the tools. One such tool is ANOVA analysis including \( R^2 \) analysis, the plots of residuals. The fitted model has to be examined for any violation of assumptions of least squares or approximation with respect to the true system.

In this research few techniques are applied and discussed in detail and are given below.
3.7.1. Properties of Least Square Estimators

This least squares method gives unbiased parameter estimator ($\beta$) for models predicted. The sum of squares for residuals are represented as

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = e'e$$  \hspace{1cm} (3.10)

As $(X'X)b = (X'y)$,

And \[ SS_E = y'y - b'X'y \]  \hspace{1cm} (3.11)

Error or residuals are calculated using the equation 3.11. The $\sigma^2$ (unbiased estimator) is written as

$$\sigma^2 = \frac{SS_E}{n - p} \hspace{1cm} (3.12)$$

Here

$n =$ Observations;  
$p =$ Regression coefficients.

The total sum of squares is shown as

$$SS_T = y'y - \left( \frac{\sum_{i=1}^{n} y_i}{n} \right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$  \hspace{1cm} (3.13)

The $R^2$ (coefficient of multiple determination) is denoted as

$$R^2 = 1 - \frac{SS_E}{SS_T} \hspace{1cm} (3.14)$$

The measure in variability ($R^2$) of $y$ gives the amount of reduction in the predicted model. It is seen from the variance identity equation (Eq. 4.14) that $0 \leq R^2 \leq 1$. The addition of variable to the models always increases $R^2$ irrespective of whether the added term is statistically significant or insignificant. Since $R^2$ every time increases after addition of the terms to the developed model.

That’s why some model builders show interest to use of adjusted $R^2$ and is defined as

$$R_{adj}^2 = 1 - \frac{SS_E / (n - p)}{SS_T / (n - p)} = 1 - \frac{n - 1}{n - p} (1 - R^2) \hspace{1cm} (3.15)$$
The adjusted $R^2$ statistic does not increase always with the addition variable. In fact $R^2_{adj}$ will often decreases if unnecessary terms were added.

For example if non-significant terms have added, it increases the variation between the terms $R^2$ and $R^2_{adj}$. The main purpose of this tests is to predict the values for each regression variable in the regression model.

The model becomes more effective sometimes if additional variables (significant related) are introduced. By addition of variable to predicted model always results sum of squares increase thereby reduces the error. The addition of non-significant variable increases the mean square error reducing the usefulness of the models.

3.7.2. Residual Analysis

The residuals $e_i = y_i - \hat{y}_i$, $i = 1, 2,..., n$, are playing a vital role in the judgement of accuracy of the models. Many response surface analysts prefer scaled residuals in place of ordinary least squares residuals. The reason being it conveys more information. The standardizing process is applied for scaling the residuals. This is by dividing with the average standard value. if the difference in standard deviations are less. Otherwise another method is adopted.

For example, the vector of fitted values $\hat{y}$, with respect to the observed values $y$, is

$$\hat{y} = Xb = X(X'X)^{-1}X'y = Hy$$  \hspace{1cm} (3.16)

Where $n \times n$ matrix $H = X(X'X)^{-1}X'$ is referred as hat matrix that is going to map the vector of that maps the vector of observed values into the vector of fitted values. The properties of hat matrix play a key role in the regression analysis.

Also examine the residuals to check the adequacy of the models.

The methods of normal probability plot of residuals are used to check the adequacy of the models. The residuals form in a straight line in case the model is adequate [47]. To determine the adequacy i.e whether the models describes the experimental data additional check required [48]. This additional check is to determine the values of regression coefficients $R^2$, which lies between 0 and 1.
3.7.3. Analysis of Variance

The ANOVA (Analysis of variance) is a group of numerical models that provides the statistical verification to check the predicted model is significant or not. The ANOVA analysis is preferred to two-sample t-test. This is because conducting multiple two-sample t-tests will increase the chances of type I error. Three different classes of models are available. Fixed effects model, mixed effects model and the random effects modes. The mixed design models describe the situations where the effects of fixed and random are present. Also different types of ANOVA are available and discussed below.

- 1-way ANOVA: This one is used to test the difference between two or more independent groups. Typically, however, this is used to test for difference among the at least 3 groups. Because the two group case will be done by the t-test. The relation between ANOVA and t is given by \( F = t^2 \).
- Factorial ANOVA: Factorial ANOVA is used when the researcher wishes to investigate the influence of two or more variables. The popular type of factorial ANOVA is the \( 2^2 \), here having 2 independent variables has 2 levels of each.
- Repeated measure ANOVA is useful for the designs that used same subjects. It is used to study over effects.
- Mixed-design ANOVA: Applied for testing 2 or more groups subjected to repeated measures i.e mixed effect model.
- Multivariate analysis of variance (MANOVA) applied if more than one dependent variables are present in the system.

3.7.3.1. Test for Significance of the Regression Model

The variance ratio or F-ratio is calculated for the predicted models to check adequacy i.e. significance of the models [49]. The F-ratio is calculated using the formula

\[
F = \frac{\text{Regression mean square}}{\text{Mean square error}}.
\]
3.7.3.2. The F-Test

This F-test is a statistical test that follows F-distribution named after Sir Ronald Fisher. It is used for comparing the statistical models and for checking the adequacy.

The necessity of F-test arises because of the reason the models are fitted using least squares method. F-test in one-way ANOVA computed as:

\[
F = \frac{\text{Variance between group items}}{\text{Variance within group}}
\]  

(3.17)

The variance between group is found using the formula:

\[
\sum n_i(y_i - \bar{y})^2 / (k - 1)
\]  

(3.18)

Here \( \bar{y}_i \) = Sample mean \((i^{th}\) group), 
\( n_i \) = Observations \((i^{th}\) group), and 
\( \bar{y} \) = Overall mean.

The variance within group is found using the derivation shown below:

\[
\sum_j (y_{ij} - \bar{y}_i)^2 / (N - k)
\]  

(3.19)

Where 
\( Y_{ij} \) is the \(j^{th}\) observation \((i^{th}\) among \( K \) groups 
\( N \) = Overall sample size.

This F-statistic will follow the F-distribution.

3.7.3.3. P-Test

The p-value is the probability of obtaining the test statistic at least as extreme as the one that was observed. While null hypothesis is true is the assumption. In multiple testing, E-value (expectation value) is the one that expects to obtain a test statistic at least as extreme as that one observed one.

When the tests are statistically independent, this exception value is the product of p-value and the number of test. The lower the p-value, the less likely the result is if the null hypothesis is true, and also more "significant" the result is, in the sense of
statistical significance. One often accepts the alternative hypothesis, (i.e. rejects a null hypothesis) if the p-value is less than 0.05 or 0.01,

3.7.3.4. Test for Significance on Individual Model Coefficients

This test is the basis for the optimization of empirical models. The p-value is determined in this approach. The effect of the control variables on the responses is decided based on the p-value. The p-value tells the proportion of time that it would expect to get the stated F-value when the control variables are not significant on the output responses.

The calculated “Prob. > F” is compared with desired probability. If the value of the test is within the level of significance, the model is said to be significant. In case of non-significant terms, the model needs to be improved by elimination of non-significant terms. This is done using backward elimination or by using forward addition.

3.7.3.5. Test for Lack-of-Fit

This test for lack of fit is done to know the significance of the replicate error, as the replicate measurements are known. This test divides the residuals of sum of squares error in to two portions.

- First one is on the basis of replicate measurements and
- Second one is lack of fit on the basis of performance of the model.

Lack of Fit = lack-of-fit mean square / pure error mean square

This F-test static is used to determine the significance level of lack of fit error. It is desirable to have the insignificant lack of fit as significant lack of fit indicates the contributions in the repressors-responses relationship that are not accounted for by the model [49].
3.8. Optimization using Genetic Algorithm (GA)

3.8.1. Concepts of Optimization

Optimization techniques are very widely using in many industries for finding optimal control variables. Also the demand for such techniques are increasing keeping in view of the lowering production costs. This demand and competition has made the engineers to search for various decision making methods. Also the methods are tested for its applicability to various engineering problems. Hence research in this area is drastically increasing and many of the engineers are choosing the appropriate method based on their suitability and their requirements. The requirement of any firm is to produce the products both efficiently and economically. This is only possible with the automation of the process or any system. The optimization techniques are applied in wide spectrum of industries includes automotive, chemical, manufacturing & electrical industries. Also usage of computers are increasing, hence it is the prime requirement of the manufacturing engineers to find the appropriate method for each process, thereby it is quite possible to optimize the process and it can be automated. Automation reduces the complexity of the problem. In other words automation helps the engineers to run the process efficiently and economically. Optimization methods are now coupled with the compute based tools to enhance or increase the creative process of detailed and conceptual design of systems in engineering applications [50].

It is the process of finding the feasible solutions (one or more) that corresponds to the extremer values of the output variables. The need to search such optimal solutions of any problem in system is designing the solution with cost effectiveness.

This is the process of finding and comparing the solutions, which are feasible. This iterative process works until no further better solution is found. In terms of the objectives, the solutions are termed as bad or good many of the researchers focused on the single objective optimization problem. If only one objective function is present, it is called as single objective optimization problem. In addition, involvement of more objective functions are termed as multi-objective optimization [50]. The tasks in the real scenario involve either maximizing or minimizing the process. In order to optimize the process, the process needs to build the models and further the model has
been formulated in the form of either single or multi objective function. The design engineer should know the aspects of formulation procedure that help to select the best suited algorithm for the design optimization problem. This requires thorough knowledge on the fundamentals of various optimization methods.

Because the optimization algorithms involve the repetitive application of certain procedures, they must be used with computer. Therefore the algorithms are represented in step-by-step process so as to code them easily.

The optimization algorithms are classified into

(i) Single-variable optimization method (gradient-based algorithms)
(ii) Multi-variable optimization method
(iii) Constrained optimization method
(iv) Specialized optimization method (i.e integer programming, geometric programming)
(v) Non-traditional optimization method (i.e Genetic algorithms, Simulated annealing, PSO Optimization, ANT Colony method, Tabu search & Harmony etc) [50].

Many problems in engineering applications will have multiple optimum solutions out of which maximum solution or minimum solutions are present. These absolute solutions (optimal) are called as global ones and others are treated as the local solutions (Optimum). The main concentration is all about the global optimal solutions because they correspond to the absolute optimum objective function value.

But many of the traditional algorithms are not guaranteed to find the global optimal solutions. For this reason, this research work focuses on the non-traditional optimization algorithms to find the global solutions to the problem undertaken.

3.9. Genetic Algorithms

Genetic algorithm (GA) is one of the optimization technique developed by John Holland [51] & Goldberg [52] is one among the large class of evolutionary algorithms available. GA is able to generate the solutions to the optimization problems. The GA uses the techniques inspired by natural evolution i.e. inheritance, mutation, selection, and crossover. This technique provides very useful results to
various types of optimization problems. In addition, it is used for search problems also.

Genetic Algorithm also called by evolutionary computation does not require any derivatives to solve the problems. It just requires as evaluation function or fitness function. This is considered the major advantage with the GA method. This method is very robust and hence is widely using in many applications or areas to find the optimal solutions.

The GA working is based on the population of designs. The chromosomes (population of strings) encode candidate solutions (individuals) to an optimization problem which evolves towards the better solutions. The solutions are represented in binaries as strings of 0s and 1s.

Initially population will be generated randomly with the help of random generation tools and with this random generated population evolution starts. Then it evaluates the fitness value for each individual in the populations in each generation. Based on the value of fitness, multiple individuals are selected and modified (recombined and possibly randomly mutated) so that it forms a new population. This new population is used for the next iteration. The program terminates if it attains the satisfactory fitness value or if it reaches the specified number of generations.

The genetic algorithm is used for numerous applications i.e in phylogenetics, bioinformatics, economics, engineering, manufacturing, mathematics, chemistry, computational science, physics and other fields.

This GA approach is different to the traditional gradient based optimization techniques in many ways i.e

- Because of the population of points unlike single design point in traditional approach, the algorithm doesn’t get trapped or struck at the local optimum value. The population size varies between $2n$ and $4n$ if the number of design variables is $n$.
- GA does not move from one point to another sequentially, instead in each generation many design points are evaluated (population-type search).
- There is no need of derivatives to be used in this searching process. It only requires the fitness value or objective function value.
- The design points or variables represented as binary a string that corresponds to the chromosomes in the natural genetics. The length of the string varies from one problem to another. It again depends upon the requirements in terms of accuracy.
- Using old generation of the old set of strings, every time new strings are produced by applying random selection and crossover.
- The GA works based on the probabilistic transition rules rather than the deterministic rules.
- This GA technique explores the new combinations with the available knowledge and then gives the new generation with the better fitness value.

The GA algorithm mainly follows the 9 steps as mentioned below
1. Random generation of Population
2. Repeating the process until it meets the termination criterion.
3. Evaluation of fitness values of chromosomes.
4. Prune the population on the basis of fitness
5. Making pairs from best ranking chromosomes
6. Using selected pairs population will be replenished.
7. Perform cross over operation.
8. Apply the mutation operation.
9. Check for the termination criteria (e.g. generations, time, minimum)
10. Loop, if not met termination criterion
The basic flow chart of the genetic algorithm is shown in Fig. 3.1

![Flowchart of genetic algorithm](image_url)

**Fig. 3.1** Flowchart of genetic algorithm
Both the maximization and the minimization optimization problems can be solved using GA. The technique is able to handle both the constrained or unconstrained optimization problems [53]. The following section explains the in detail working of GA Algorithm.

Maximize \( f(x) = x_i^{(L)} \leq x_i \leq x_i^{(U)}, \ i=1,2,\ldots,N \).

### 3.9.1. Individual Representation

The string in general is represented in the binary coded form with 0s and 1s and is given as follows. This representation form is used to describe the parameter for each solution as given in Fig. 3.2.

As per the solution accuracy requirements, the length of the string is defined. For example, in a two variable function problem, four strings are used to code each individual.

i.e \( (x_1^{(L)}, x_2^{(L)})^T, (x_1^{(U)}, x_2^{(U)})^T = (0000 \ 0000) \) and \( (1111 \ 1111) \)  \( (3.20) \)

The substrings 1111 and 0000 have the maximum and the minimum decoded values. As per the fixed mapping rule, eight-bit string also is used for representation. Usually, the following linear mapping rule is used for all the strings:

\[
x_i = x_i^{(L)} + \frac{x_i^{(U)} - x_i^{(L)}}{2^h - 1} \times \text{decoded value} \ (s_i)
\]

(3.21)

In the equation (3.21), the variable \( x_i \) is coded in a substring \( s_i \) of length \( l_i \). for example, the decoded value for (0111) is \( ((1)2^0+(1)2^1+(1)2^2+(1)2^3) \) or 7. The accuracy that can be obtained with a four-bit coding is roughly \( 1/16^{th} \) of the search space. And the accuracy changes if the string length is increased.
3.9.2. Fitness Function

GA works on the principle of the survival-of-the-fittest. In general, this is best suited for maximization problems.

The minimization problems are generally transformed into maximization problems with suitable transformation. The process of GA starts with a population of random strings representing design or responses. The fitness function is obtained from the objective function and will be used in next genetic operations. Reproduction, crossover and mutation are the three important genetic operators and are explained below.

GA mimics the survival-of-the-fittest principle of nature to make a search process. Therefore, GA is naturally suitable for solving maximization problems. Minimization problems are usually transformed into maximization problems by some suitable transformation. The fitness function is derived from the objective function and used in successive genetic operations. The fitness function value of a string is known as the string’s fitness. The brief explanation of genetic operators is given in [54].

3.9.3. Selection / Reproduction

Selection / Reproduction is the first operation applied on a population. The selection scheme determines how individuals are chosen for mating, based on their fitness scores. Too great a bias towards the best individuals can result in premature convergence, so the best selection schemes are designed to maintain a diverse population.

- Rank - pick the best individuals every time
- Roulette wheel (proportionate) - probability of selection is proportional to fitness
- Tournament - initial large number are selected via roulette wheel, then the best ranked are chosen.
- Stochastic - various methods of replenishment of less fit stock (useful) or initial selection (not useful)
• Elite - in combination with other selection schemes, always keep the fittest individual around. The mating pool is formed with the new strings.

It selects the good strings based on fitness calculations. This is also called as the selection. The idea is to create the strings with good fitness values. The selection of chromosomes for the next stage is done here using the formula

\[ F_{average} = \frac{\sum_{i=1}^{n} F_i}{n} \]

for calculating the fitness values.

Fitness is computed for each individual. To help maintain diversity or differentiate between similar individuals, raw objective scores are sometimes scaled to produce the final fitness scores.

3.9.4. Crossover

The crossover operation is the most important one in case of GA. In this operation partial exchange of information takes place among two individuals selected based on fitness values. Off springs are created from the parent strings in this crossover operation. The crossing sites are chosen randomly. By iteratively applying the crossover operator, genes of good chromosomes are expected to appear more frequently in the population, eventually leading to convergence to an overall good solution. The crossover operation is explained in Fig. 3.3. The exchanging of bits takes place on the right side of the crossing site as shown below:

![Image](https://via.placeholder.com/150)

**Fig. 3.3.** Illustration of crossover operator.

To perform this operation, cross over probability is selected in general it lies between 0.6 and 0.9. First from the population 2 strings are selected at random. Then random number is generated between 0 and 1. If the number is less than the \( P_c \) this
operation is performed. Like this child strings are produced. Not all the strings are subjected to cross over operation.

3.9.5. Mutation

Generally, bits are flipped with a small probability. This explores parts of the space that crossover might miss, and helps prevent premature convergence. This operator brings random changes into characteristics of chromosomes. Mutation operator changes 1 to 0 and vice versa with a small mutation probability $p_m$.

Mutation is performed bit wise with the mutation probability set in between 0.02 to 0.06. Fig. 3.4 illustrates the mutation operation at seventh bit position.

![Illustration of mutation operator.](image)

To apply this mutation, random number between 0 and 1 are generated. If random number generated is less than the mutation probability, mutation operation will be performed. Otherwise the strings remain as it is without any changes. Overall with these three operators GA is able to produce good strings so that the final solution will be the optimal one.

3.9.6. Representation of Objective Function and Constraints

Because genetic algorithms are based on the survival-of-the-fittest principle of nature, they try to maximize a function called the fitness function. Thus GA’s are naturally suitable for solving unconstrained maximization problems.

The fitness function, $F(X)$, can be taken to be same as the objective function $f(X)$ of an unconstrained maximization problem so that $F(X) = f(X)$.

A minimization problem can be transformed into a maximization problem before applying the GA’s. Usually the fitness function is chosen to be nonnegative.

The commonly used transformation to convert an unconstrained minimization problem to a fitness function is given by

$$F(X) = \frac{1}{1 + f(X)}$$  \hspace{1cm} (3.22)
It is seen that Eq. (3.22) does not alter the location of the minimum of \( f(X) \), but converts the minimization problem into an equivalent maximization problem. A general constrained minimization problem is stated as

\[
\text{Minimize } f(X)
\]

subject to

\[
g_i(X) \leq 0, \quad i = 1, 2, 3, \ldots, m
\]

and

\[
h_j(X) = 0, \quad j = 1, 2, 3, \ldots, p
\]

This problem is converted into an equivalent unconstrained minimization problem by using the concept of penalty function as

\[
\text{Minimize } \varnothing(X) = f(X) + \sum_{i=1}^{m} r_i (g_i(X))^2 + \sum_{j=1}^{p} R_j (h_j(X))^2
\]

(3.23)

Where, \( r_i \) is the penalty parameter associated with the constraint \( g_i(X) \) and \( R_j \) is the penalty parameter associated with the constraint \( h_j(X) \).

The values of \( r_i \) and \( R_j \) are usually kept constant throughout the solution process.

In Eq. (3.23), the function \( (g_i(X)) \), called the bracket function, is defined as

\[
(g_i(X)) = \begin{cases} 
  g_i(X) & \text{if } g_i(X) > 0 \\
  0 & \text{if } g_i(X) \leq 0
\end{cases}
\]

(3.24)

In most cases, the penalty parameters associated with all the inequality and equality constraints are assumed to be the same constants as

\[
r_i = r, \quad i = 1, 2, \ldots, m \quad \text{and} \quad R_j = R, \quad j = 1, 2, \ldots, p
\]

Where, \( r \) and \( R \) are constants. The fitness function, \( F(X) \), to be maximized in the GA is obtained, similar to Eq. (3.22), as

\[
F(X) = \frac{1}{1 + \varnothing(X)}
\]

(3.25)

Equations (3.23) and (3.24) show that the penalty will be proportional to the square of the amount of violation of the inequality and equality constraints at the design vector \( X \), while there will be no penalty added to \( f(X) \) if all the constraints are satisfied at the design vector \( X \).