NAAC ACCREDITED

तेजस्वि नावधीतमस्तु

ISO 9001:2008 & 14001:2004

FAIRFIELD

Institute of Management & Technology

'A' Grade Institute by DHE, Govt. of NCT Delhi and Approved by the Bar Council of India and NCTE

Reference Material for Three Years

Bachelor in Business Administration (General)

Code : 017

Semester – I

FIMT Campus, Kapashera, New Delhi-110037, Phones : 011-25063208/09/10/11, 25066256/ 57/58/59/60

Fax : 011-250 63212 Mob. : 09312352942, 09811568155 E-mail : fimtoffice@gmail.com Website : www.fimt-ggsipu.org

INDEX

Three Years

Bachelor in Business Administration (General)

Code : 017

Semester – I

Management Process & Organizational Behaviour (101)

Management Process & Organizational Behaviour

Unit 1

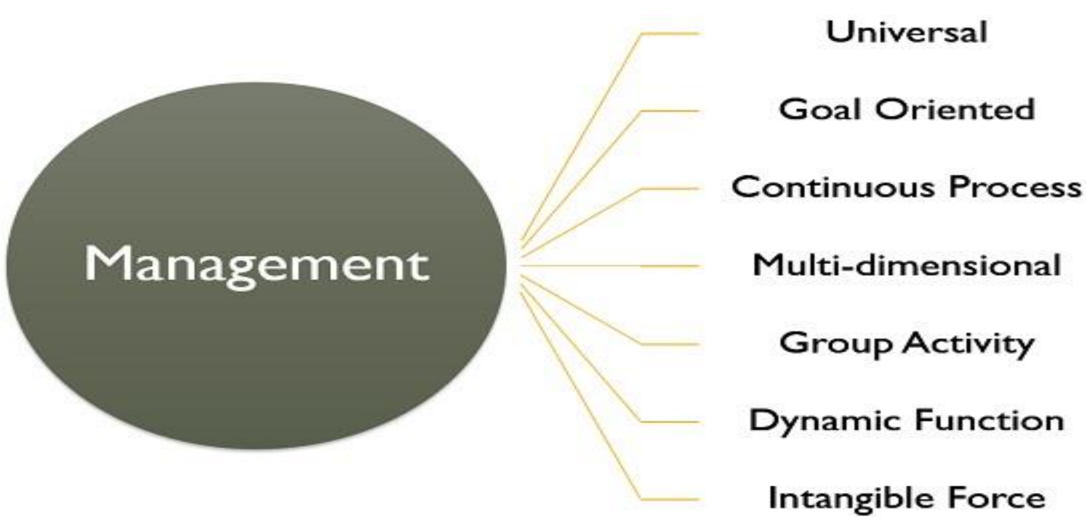
Management

According to Harold Koontz,

‘Management is an art of getting things done through and with the people in formally organized groups. It is an art of creating an environment in which people can perform and individuals and can co-operate towards attainment of group goals.‘

In other words, it is concerned with optimally using 5M’s, i.e. men, machine, material, money and methods and, this is possible only when there proper direction, coordination and integration of the processes and activities, to achieve the desired results.

Characteristics of Management



- Universal: All the organizations, whether it is profit-making or not, they require management, for managing their activities. Hence it is universal in nature.
- Goal Oriented: Every organization is set up with a predetermined objective and management helps in reaching those goals timely, and smoothly.

- Continuous Process: It is an ongoing process which tends to persist as long as the organization exists. It is required in every sphere of the organization whether it is production, human resource, finance or marketing.
- Multi-dimensional: Management is not confined to the administration of people only, but it also manages work, processes and operations, which makes it a multi-disciplinary activity.
- Group activity: An organization consists of various members who have different needs, expectations and beliefs. Every person joins the organization with a different motive, but after becoming a part of the organization they work for achieving the same goal. It requires supervision, teamwork and coordination, and in this way, management comes into the picture.
- Dynamic function: An organization exists in a business environment that has various factors like social, political, legal, technological and economic. A slight change in any of these factors will affect the organization's growth and performance. So, to overcome these changes management formulates strategies and implements them.

Levels of Management

1. Top-Level Management: This is the highest level in the organizational hierarchy, which includes Board of Directors and Chief Executives. They are responsible for defining the objectives, formulating plans, strategies and policies.
2. Middle-Level Management: It is the second and most important level in the corporate ladder, as it creates a link between the top and lower level management. It includes departmental and division heads and managers who are responsible for implementing and controlling plans and strategies which are formulated by the top executives.

3. Lower Level Management: Otherwise called as functional or operational level management. It includes first line managers, foreman, and supervisors. As lower level management directly interacts with the workers, it plays a crucial role in the organization because it helps in reducing wastage and idle time of the workers, improving the quality and quantity of output.

The three management levels form the management hierarchy that represents the position and rank of executives and managers in the chart.

Functions of Management

- Planning: It is the first and foremost function of management, i.e. to decide beforehand what is to be done in future. It encompasses formulating policies, establishing targets, scheduling actions and so forth.

- Organizing: Once the plans are formulated, the next step is to organise the activities and resources, as in identifying the tasks, classifying them, assigning duties to subordinates and allocating the resources.

- Staffing: It involves hiring personnel for carrying out various activities of the organization. It is to ensure that the right person is appointed to the right job.

- Directing: It is the task of the manager to guide, supervise, lead and motivate the subordinates, to ensure that they work in the right direction, so far as the objectives of the organization are concerned.

- Controlling: The controlling function of management involves a number of steps to be taken to make sure that the performance of the employees is as per the plans. It involves establishing performance standards and comparing them with the actual performance. In case of any variations, necessary steps are to be taken for its correction.

Skills of Managers in the Organizational Hierarchy:

- **(i)** Technical Skill: It is knowledge of and proficiency in activities involving methods, processes, and procedures. Thus, it involves working with tools and specific techniques. For examples, mechanics work with tools, and their supervisor should have the ability to teach them how to use these tools. Similarly, accountants apply specific techniques in doing their job.
- **(ii)** Human Skill: It is the ability to work with people; it is cooperative efforts; it is teamwork; it is the creation of an environment in which people feel secure and free express their opinions.
- **(iii)** Conceptual Skill: It is the ability to see the 'big picture' to recognize significant elements in a situation, and to understand the relationships among the elements.
- **(iv)** Design Skill: It is the ability to solve problems in ways that will benefit the enterprise. To be effective, particularly at upper organizational levels, managers must be able to do more than see a problem. They must have, in addition, the skill of a good design engineer in working out a practical solution to a problem.

Role of a Manager:

1. Interpersonal Roles

The interpersonal roles link all managerial work together. The three interpersonal roles are primarily concerned with interpersonal relationships.

Figurehead Role: The manager represents the organization in all matters of formality. The top level manager represents the company legally and socially to those outside of the organization. The supervisor represents the work group to higher management and higher management to the work group.

Liaison Role: The manger interacts with peers and people outside the organization. The top level manager uses the liaison role to gain favors and information, while the supervisor uses it to maintain the routine flow of work.

- The leader Role: It defines the relationships between the manger and employees.

2. Informational Roles

The informational roles ensure that information is provided. The three informational roles are primarily concerned with the information aspects of managerial work.

- Monitor Role: The manager receives and collects information about the operation of an enterprise.

- Disseminator Role: The manager transmits special information into the organization. The top level manager receives and transmits more information from people outside the organization than the supervisor.

- Spokesperson Role: The manager disseminates the organization's information into its environment. Thus, the top level manager is seen as an industry expert, while the supervisor is seen as a unit or departmental expert.

3. Decisional Roles

The decisional roles make significant use of the information and there are four decisional roles.

- Entrepreneur role
- Disturbance handler role
- Resource Allocator role
- Negotiator role

Management vs Administration

Basis	Management	Administration
Meaning	Management is an art of getting things done through others by directing their efforts towards achievement of pre-determined goals.	It is concerned with formulation of broad objectives, plans & policies.
Nature	Management is an executing function.	Administration is a decision-making function.
Process	Management decides who should as it & how should he dot it.	Administration decides what is to be done & when it is to be done.
Function	Management is a doing function because managers get work done under their supervision.	Administration is a thinking function because plans & policies are

		determined under it.
Skills	Technical and Human skills	Conceptual and Human skills
Level	Middle & lower level function	Top level function

"**Co-ordination is the Essence of Management**." The meaning of this sentence implies, Co-ordination affects all the functions of management. In other words, Co-ordination affects Planning, Organizing, Staffing, Directing, Communication, Leading, Motivating and Controlling.

Classical Organisation Theory:

The classical writers viewed organisation as a machine and human beings as components of that machine. They were of the view that efficiency of the organisation can be increased by making human beings efficient. Their emphasis was on specialisation and co-ordination of activities. Most of the writers gave emphasis on efficiency at the top level and few at lower levels of organisation. That is why this theory has given streams; scientific management and administrative management. The scientific management group was mainly concerned with the tasks to be performed at operative levels.

Henry Fayol studied for the first time the principles and functions of management. Some authors like Gullick, Oliver Sheldon, Urwick viewed the problem where identification of activities is necessary for achieving organisation goals. Grouping or departmentation was also considered essential for making the functions effective. Since this theory revolves around structure it is also called ‘structural theory of organisation.”

Pillars of Organisation Theory:

According to classical writers, the organisation theory is built around four key pillars division of work, scalar and functional processes, structure and span of control.

(i) Division of Labour:

Division of labour implies that work must be divided to obtain specialisation with a view to improve the performance of workers. The classical theory rests on the assumption that more a particular job is broken into its simplest component parts, the more specialised a worker can become in carrying out his part of the job.

The specialisation in workers will make the organisation efficient. Various activities of a job are specified and subdivided into different components so that these may be assigned to different persons.

The workers will go on repeating their work under division of labour. The performance of same work will help workers to improve their efficiency and the organisation as a whole is benefitted by this exercise

(ii) Scalar and Functional Process:

The scalar process refers to the growth of chain of command, delegation of authority, unity of command and obligation to report. It is called scalar process because it provides a scale or grading of duties according to the degree of authority and responsibility. It generates superior- subordinate relationship in the organisation. The functional process deals with the division of organisation into specialised parts or departments and regrouping of the parts into compatible units.

(ii) Structure:

It is the framework of formal relationships among various tasks, activities and people in the organisation. The basic structural element in the classical theory is position. Each position is assigned a specific task and authority is delegated for its accomplishment. The efficiency with which these tasks will be accomplished will determine the effectiveness of the organisation. The classical writers emphasised line and staff organisations.

(iv) Span of Control:

The span of control means the number of subordinates a manager can control. Classical thinkers specified numbers at different levels which can be effectively supervised by a superior. A manager cannot exercise proper control if the number of subordinates increases beyond a certain figure, on the other hand if the number is less then his capacity and knowledge cannot be fully utilised.

2. Neo-Classical Organisation Theory:

The classical theory of organisation focussed main attention on physiological and mechanical variables of organisational functioning. The testing of these variables did not show positive results. The Hawthorne Studies conducted by George Elton Mayo and associates discovered that real cause of human behaviour was somewhat more than mere physiological variables. These studies focussed attention on human beings in the organisation.

Neo-classical approach is contained in two points:

(i) Organisational situation should be viewed in social, economic and technical terms, and

(ii) the social process of group behaviour can be understood in terms of clinical method analogous to the doctor's diagnosis of human organism.

This theory views formal and informal forms of organisation as important. The behavioural approach followed in this theory is the other contribution of new-classical thinkers. The pillars of classical theory viz. division of work, departmentation, co-ordination and human behaviour were taken as given but these postulates were regarded as modified by people acting independently or within the context of the informal organisation.

The main propositions of neo-classical theory are given as follows:

1. The organisation in general is a social system composed of numerous interacting parts.
2. Informal organisations exist within the formal organisation. Both are affected by and affect each other.
3. Human being is independent and his behaviour can be predicted in terms of social factors at work.
4. Motivation is a complex process. Many socio- psychological factors operate to motivate human beings at work.
5. A conflict between organisational and individual goals often exists. There is a need to reconcile the goals of the individual with those of the organisation.
6. Team-work is essential for higher productivity.
7. Man’s approach is not always rational. Often, he behaves non- logically in terms of rewards which he seeks from his work.
8. Communication is necessary as it carries information for the functioning of the organisation and the feelings of the people at work.

Systems Approach:

This approach studies the organisation in its totality. The mutually dependent variables are properly analysed. Both internal and external variables are studied in analysing the nature of organisation. Though this theory passes a much higher conceptual level as compared to earlier theories but different writers have given varied views of the system.

Organisation as a system can well be understood by identifying various sub-systems within it. Each sub-system may be identified by certain processes, roles, structures and norms of conduct. Seiler has classified four components in an organisation, human inputs, technological inputs, organisational inputs, and social structure and norms.

Katz and Kahu have identified five sub-systems of organisation:

(i) Technical sub-system concerned with the work that gets done;

(ii) Supportive sub-system of procurement, disposal and institutional relations;

(iii) Maintenance of sub-systems for tying people into their functional roles;

(iv) Adaptive sub-systems concerned with organisational change; and

(v) Managerial sub-systems for direction, adjudication and control of the many sub-systems and the activities of the structure.

Contingency Approach:

Even though systems approach presents a better understanding of organisational and managerial functioning but it does not provide solution for all types of organisational structures. Systems approach offers models which may not suit every type of organisation. A structure suitable for one unit may not be suitable for another. Contingency approach suggests an organisational design which suits a particular unit. A structure will be suitable only if it is tailor made for an enterprise.

The influence of both internal and external factors should be considered while framing a suitable organisational structure. This approach suggests that needs, requirements, situations of a particular concern should be considered while designing an organisational structure.

The factors which influence an organization may be described as:

(i) Environment

(ii) Technology

(iii) Size of operations

(iv) People.

These factors greatly influence a decision for the selection of an appropriate organisation for an enterprise.

Planning

Meaning:

Planning can be defined as "thinking in advance what is to be done, when it is to be done, how it is to be done and by whom it should be done". In simple words we can say, planning bridges the gap between where we are standing today and where we want to reach.

Planning involves setting objectives and deciding in advance the appropriate course of action to achieve these objectives so we can also define planning as setting up of objectives and targets and formulating an action plan to achieve them.

Another important ingredient of planning is time. Plans are always developed for a fixed time period as no business can go on planning endlessly.

Keeping in mind the time dimension we can define planning as "Setting objectives for a given time period, formulating various courses of action to achieve them and then selecting the best possible alternative from the different courses of actions".

Features/Nature/Characteristic of Planning:

1. Planning contributes to Objectives:

Planning starts with the determination of objectives. We cannot think of planning in absence of objective. After setting up of the objectives, planning decides the methods, procedures and steps to be taken for achievement of set objectives. Planners also help and bring changes in the plan if things are not moving in the direction of objectives.

For example, if an organisation has the objective of manufacturing 1500 washing machines and in one month only 80 washing machines are manufactured, then changes are made in the plan to achieve the final objective.

2. Planning is Primary function of management:

Planning is the primary or first function to be performed by every manager. No other function can be executed by the manager without performing planning function because objectives are set up in planning and other functions depend on the objectives only.

3. Pervasive:

Planning is required at all levels of the management. It is not a function restricted to top level managers only but planning is done by managers at every level. Formation of major plan and framing of overall policies is the task of top level managers whereas departmental managers form plan for their respective departments. And lower level managers make plans to support the overall objectives and to carry on day to day activities.

4. Planning is futuristic/Forward looking:

Planning always means looking ahead or planning is a futuristic function. Planning is never done for the past. All the managers try to make predictions and assumptions for future and these predictions are made on the basis of past experiences of the manager and with the regular and intelligent scanning of the general environment.

5. Planning is continuous:

Planning is a never ending or continuous process because after making plans also one has to be in touch with the changes in changing environment and in the selection of one best way.

6. Planning involves decision making:

The planning function is needed only when different alternatives are available and we have to select most suitable alternative. We cannot imagine planning in absence of choice because in planning function managers evaluate various alternatives and select the most appropriate. But if there is one alternative available then there is no requirement of planning.

For example, to import the technology if the licence is only with STC (State Trading Co-operation) then companies have no choice but to import the technology through STC only. But if there is 4-5 import agencies included in this task then the planners have to evaluate terms and conditions of all the agencies and select the most suitable from the company's point of view.

7. Planning is a mental exercise:

It is mental exercise. Planning is a mental process which requires higher thinking that is why it is kept separate from operational activities by Taylor. In planning assumptions and predictions regarding future are made by scanning the environment properly. This activity requires higher level of intelligence. Secondly, in planning various alternatives are evaluated and the most suitable is selected which again requires higher level of intelligence. So, it is right to call planning an intellectual process.

Importance/Significance of Planning:

1. Planning provides Direction:

Planning is concerned with predetermined course of action. It provides the directions to the efforts of employees. Planning makes clear what employees have to do, how to do, etc. By stating in advance how work has to be done, planning provides direction for action. Employees know in advance in which direction they have to work. This leads to Unity of Direction also. If there were no planning, employees would be working in different directions and organisation would not be able to achieve its desired goal.

2. Planning Reduces the risk of uncertainties:

Organisations have to face many uncertainties and unexpected situations every day. Planning helps the manager to face the uncertainty because planners try to foresee the future by making some assumptions regarding future keeping in mind their past experiences and scanning of business environments. The plans are made to overcome such uncertainties. The plans also include unexpected risks such as fire or some other calamities in the organisation. The resources are kept aside in the plan to meet such uncertainties.

3. Planning reduces over lapping and wasteful activities:

The organisational plans are made keeping in mind the requirements of all the departments. The departmental plans are derived from main organisational plan. As a result there will be co-ordination in different departments. On the other hand, if the managers, non-managers and all the employees are following course of action according to plan then there will be integration in the activities. Plans ensure clarity of thoughts and action and work can be carried out smoothly.

4. Planning Promotes innovative ideas:

Planning requires high thinking and it is an intellectual process. So, there is a great scope of finding better ideas, better methods and procedures to perform a particular job. Planning process forces managers to think differently and assume the future conditions. So, it makes the managers innovative and creative.

5. Planning Facilitates Decision Making:

Planning helps the managers to take various decisions. As in planning goals are set in advance and predictions are made for future. These predictions and goals help the manager to take fast decisions.

6. Planning establishes standard for controlling:

Controlling means comparison between planned and actual output and if there is variation between both then find out the reasons for such deviations and taking measures to match the actual output with the planned. But in case there is no planned output then controlling manager will have no base to compare whether the actual output is adequate or not.

For example, if the planned output for a week is 100 units and actual output produced by employee is 80 units then the controlling manager must take measures to bring the 80 unit production upto 100 units but if the planned output, i.e., 100 units is not given by the planners then finding out whether 80 unit production is sufficient or not will be difficult to know. So, the base for comparison in controlling is given by planning function only.

7. Focuses attention on objectives of the company:

Planning function begins with the setting up of the objectives, policies, procedures, methods and rules, etc. which are made in planning to achieve these objectives only. When employees follow the plan they are leading towards the achievement of objectives. Through planning, efforts of all the employees are directed towards the achievement of organisational goals and objectives.

Limitations of Planning:

1. Planning leads to rigidity:

Once plans are made to decide the future course of action the manager may not be in a position to change them. Following predefined plan when circumstances are changed may not bring positive results for organisation. This kind of rigidity in plan may create difficulty.

2. Planning may not work in dynamic environment:

Business environment is very dynamic as there are continuously changes taking place in economic, political and legal environment. It becomes very difficult to forecast these future changes. Plans may fail if the changes are very frequent.

The environment consists of number of segments and it becomes very difficult for a manager to assess future changes in the environment. For example there may be change in economic policy, change in fashion and trend or change in competitor's policy. A manager cannot foresee these changes accurately and plan may fail if many such changes take place in environment.

3. It reduces creativity:

With the planning the managers of the organisation start working rigidly and they become the blind followers of the plan only. The managers do not take any initiative to make changes in the plan according to the changes prevailing in the business environment. They stop giving suggestions and new ideas to bring improvement in working because the guidelines for working are given in planning only.

4. Planning involves huge Cost:

Planning process involves lot of cost because it is an intellectual process and companies need to hire the professional experts to carry on this process. Along with the salary of these experts the company has to spend lot of time and money to collect accurate facts and figures. So, it is a cost-consuming process. If the benefits of planning are not more than its cost then it should not be carried on.

5. It is a time consuming process:

Planning process is a time-consuming process because it takes long time to evaluate the alternatives and select the best one. Lot of time is needed in developing planning premises. So, because of this, the action gets delayed. And whenever there is a need for prompt and immediate decision then we have to avoid planning.

6. Planning does not guarantee success:

Sometimes managers have false sense of security that plans have worked successfully in past so these will be working in future also. There is a tendency in managers to rely on pretested plans.

It is not true that if a plan has worked successfully in past, it will bring success in future also as there are so many unknown factors which may lead to failure of plan in future. Planning only provides a base for analysing future. It is not a solution for future course of action.

7. Lack of accuracy:

In planning we are always thinking in advance and planning is concerned with future only and future is always uncertain. In planning many assumptions are made to decide about future course of action. But these assumptions are not 100% accurate and if these assumptions do not hold true in present situation or in future condition then whole planning will fail.

Planning Process:

1. Setting up of the objectives:

In planning function manager begins with setting up of objectives because all the policies, procedures and methods are framed for achieving objectives only. The managers set up very clearly the objectives of the company keeping in mind the goals of the company and the physical and financial resources of the company. Managers prefer to set up goals which can be achieved quickly and in specific limit of time. After setting up the goals, the clearly defined goals are communicated to all the employees.

2. Developing premises:

Premises refer to making assumptions regarding future. Premises are the base on which plans are made. It is a kind of forecast made keeping in view existing plans and any past information about various policies. There should be total agreement on all the assumptions. The assumptions are made on the basis of forecasting. Forecast is the technique of gathering information. Common forecast are made to find out the demand for a product, change in government or competitor policy, tax rate, etc.

3. Listing the various alternatives for achieving the objectives:

After setting up of objectives the managers make a list of alternatives through which the organisation can achieve its objectives as there can be many ways to achieve the objective and managers must know all the ways to reach the objectives.

For example, if the objective is to increase in sale by 10% then the sale can be increased:

(a) By adding more line of products;

(b) By offering discount;

(c) By increasing expenditure on advertisements;

(d) By increasing the share in the market;

(e) By appointing salesmen for door-to-door sale etc.

So, managers list out all the alternatives.

4. Evaluation of different alternatives:

After making the list of various alternatives along with the assumptions supporting them, the manager starts evaluating each and every alternative and notes down the positive and negative aspects of every alternative. After this the manager starts eliminating the alternatives with more of negative aspect and the one with the maximum positive aspect and with most feasible assumption is selected as best alternative. Alternatives are evaluated in the light of their feasibility.

5. Selecting an alternative:

The best alternative is selected but as such there is no mathematical formula to select the best alternative. Sometimes instead of selecting one alternative, a combination of different alternatives can also be selected. The most ideal plan is most feasible, profitable and with least negative consequences.

After preparing the main plan, the organisation has to make number of small plans to support the main plan. These plans are related to performance of routine jobs in the organisation. These are derived from the major plan. So, they are also known as derivative plans. These plans are must for accomplishing the objective of main plan. The common supportive plans are plans to buy equipment, plan for recruitment and selection of employees, plan to buy raw material, etc.

6. Implement the plan:

The managers prepare or draft the main and supportive plans on paper but there is no use of these plans unless and until these are put in action. For implementing the plans or putting the plans into action, the managers start communicating the plans to all the employees very clearly because the employees actually have to carry on the activities according to specification of plans. After communicating the plan to employees and taking their support the managers start allocating the resources according to the specification of the plans. For example, if the plan is to increase in sale by increasing the expenditure on advertisement, then to put it into action, the managers must allot more funds to advertisement department, select better media, hire advertising agency, etc.

7. Follow-up:

Planning is a continuous process so the manager's job does not get over simply by putting the plan into action. The managers monitor the plan carefully while it is implemented. The monitoring of plan is very important because it helps to verify whether the conditions and predictions assumed in plan are holding true in present situation or not. If these are not coming true then immediately changes are made in the plan.

During follow up many adjustments are made in the plan. For example, if the expenditure planning is done keeping in mind 5% inflation rate but in present situation if the inflation rate rises to 10% then during follow up the managers make changes in the plans according to 10% inflation rate.

Plan:

Plan is a document that outlines how goals are going to be met. It is a specific action proposed to help the organization achieve its objectives. There may be more than one way and means of reaching a particular goal but with the help of logical plans, objectives of an organization could be easily achieved.

Single Use Plans:

Single use plans are one time use plan. These are designed to achieve a particular goal that once achieved will not reoccur in future. These are made to meet the needs of unique situations. The duration or length of single use plan depends upon the activity or goal for which it is made. It may last one day or it may last for weeks or months if the project for which it is made is long.

Standing Plans:

Standing plans are also known as Repeat Use Plans. These plans focus on situations which occur repeatedly. Standing plans are used over and over again. They are made once but retain their value over a period of years. Although some revisions and updates are made in these plans from time to time.

Types of Plans:

Planning is a pervasive function which means it is not the task of top level managers only but managers working at different levels perform planning function. The plans framed by top level manager may differ from the plans formed by middle and lower level managers. The different types of plans or common plans formed by the managers at different levels are:

Objectives – Rules

Strategy – Programmes

Policies – Methods

Procedures – Budgets

1. Objectives:

Objectives are the ends towards which the activities are directed. They are the end result of every activity. An objective:

(a) Should be related to single activity;

(b) Should be related to result and not to activity to be performed;

(c) It should be measurable or must be measured in quantitative term;

(d) It must have a time limit for achievement of objective;

(e) It must be achievable or feasible.

For example, increase in sale by 10% or decrease in rejections by 2%.

2. Strategy:

A strategy is a comprehensive plan to achieve the organisational objectives. The dimensions of strategy are:

(i) Determining long term objectives.

(ii) Adopting a particular course of action.

(iii) Allocating resources for achieving the objectives.

Strategy formulation is the task of top level people and it is must to scan and understand clearly the business environment before framing the strategy. The common decisions in strategy are whether to introduce a new product or not. If to introduce then how, finding out customer for your products making changes in existing products etc. All the strategic decisions are greatly influenced by the business environment. Strategy defines the future decisions regarding the organisation's direction and scope in the long run.

For example, Choice of advertising media, sales promotion techniques, channels of distribution, etc.

3. Policies:

Policy can be defined as organisation's general response to a particular problem or situation. In simple words, it is the organisation's own way of handling the problems. Policies are made at every level because the managers at every level need to decide or predetermine the way of handling a situation and policy acts as a guide to take decisions in unexpected situation.

Policy formation always encourages initiatives of employees because employees have to deal with situations and the way of handling the situation is decided in consultation with the employees. Then they will be able to handle the situation in a much better way. For example, a school may have policy of issuing admission form only to students who secured more than 60% marks.

"No credit sale policy", etc. Introduction of new product in the market.

4. Procedures:

Procedures are required steps established in advance to handle future conditions. The sequence of steps to be followed by employees in different situations must be predetermined so that everyone follows same steps.

The procedure can be defined as the exact manner in which an activity has to be accomplished.

For example, the procedure for admission in a particular school can be:

(a) Set up a file for applicants;

(b) Accept the field forms and put them in a file;

(c) Ask for other certificates to verify score or marks of students;

(d) Put those documents also in the file;

(e) Give the file to admission in-charge.

Procedures are made common for all the departments to co-ordinate their activities. So procedures cut across all the departmental lines. For example, the procedure to handle the order by manufacturing department may involve sales department also.

5. Rules:

Rules spell out special actions or non-actions of the employees. There is no discretion allowed in rules, i.e., they must be followed strictly and if rules are not followed then strict actions can be taken against employees who are disobeying the rules. Rules are spelt out to create the environment of discipline in the organisation. For example, there can be rule of no smoking in the organisation. Rules generally guide the general behaviour of the employees and employees cannot make any changes in them.

6. Programmes:

Programmes are the combination of goals, policies, procedures and rules. All these plans together form a program. The programmes are made to get a systematic working in the organisation. The programmes create relation between policies, procedures and goals. The programmes are also prepared at different levels. A primary programme is prepared by the top level and then to support the primary programme supportive programmes of different levels are prepared for smooth function of the company.

For example, construction of shopping mall, Development of new product.

7. Methods:

Methods can be defined as formalized or systematic way of doing routine or repetitive jobs. The managers decide in advance the common way of doing a job. So, that

(a) There is no doubt in the minds of employees;

(b) There can be uniformity in actions of the employees;

(c) These help in applying the techniques of standardization and simplification;

(d) Act as guide for employees.

If the common way of doing the job is not decided in advance then there will be confusion and comparison will not be possible. For example, for the valuation of stock, the organisation must decide in advance what method has to be adopted (lifoor fifo). So that everyone follows the same method and comparison with the past value of stock can be done, method for calculation of depreciation.

8. Budget:

Budget is the statement of expected result expressed in numerical terms. In budgets the results are always measurable and most of the time these are financial in nature but it does not mean that company prepares only financial budget. Financial budget is also known as profit plan of the company because it includes the expected income and related expenditures with that income and the profit which the company will earn in the coming year.

Along with financial budget capital budget is prepared to find out the expected capital requirement. Operational budget is prepared where instead of finance hourly units are used stating expected hours the employees will be working. Budgets are prepared by managers at every level and lower level managers generally prepare operational budgets.

The most common budget prepared by managers at different levels is cash budget. This budget estimates the expected cash inflow and cash outflow over a period of time. Cash inflow comes from sales and cash outflow is in the form of expenses. Businessmen can find out net cash position by subtracting cash outflow from cash inflow.

Business Forecasting

"Business Forecasting is the research procedure to discover those economic, social and financial influences governing business activity, so as to predict or estimate current and future trends or forces which may have a bearing on company policies or future financial, production and marketing operations."

Types of Business Forecasting:

Various types of Business Forecasting are –

- General Forecast.
- Sales Forecast, and.
- Capital Forecast.

MBO

What is Management by Objective

The process of setting objectives in the organization to give a sense of direction to the employees is called as Management by Objectives.

It refers to the process of setting goals for the employees so that they know what they are supposed to do at the workplace.

Management by Objectives defines roles and responsibilities for the employees and help them chalk out their future course of action in the organization.

Management by objectives guides the employees to deliver their level best and achieve the targets within the stipulated time frame.

Need for Management by Objectives (MBO)

- The Management by Objectives process helps the employees to understand their duties at the workplace.
- KRAs are designed for each employee as per their interest, specialization and educational qualification.
- The employees are clear as to what is expected out of them.
- Management by Objectives process leads to satisfied employees. It avoids job mismatch and unnecessary confusions later on.
- Employees in their own way contribute to the achievement of the goals and objectives of the organization. Every employee has his own role at the workplace. Each one feels indispensable for the organization and eventually develops a feeling of loyalty towards the organization. They tend to stick to the organization for a longer span of time and contribute effectively. They enjoy at the workplace and do not treat work as a burden.
- Management by Objectives ensures effective communication amongst the employees. It leads to a positive ambience at the workplace.
- Management by Objectives leads to well defined hierarchies at the workplace. It ensures transparency at all levels. A supervisor of any organization would never directly interact with the Managing Director in case of queries. He would first meet his reporting boss who would then pass on the message to his senior and so on. Every one is clear about his position in the organization.
- The MBO Process leads to highly motivated and committed employees.
- The MBO Process sets a benchmark for every employee. The superiors set targets for each of the team members. Each employee is given a list of specific tasks.

Benefits of Management by Objectives

1. It facilitates the employees to understand their tasks and duties in a better way.
2. It is helpful in designing Key Result Area (KRA) for each employee, according to their interest, specialization, experience and competency.
3. It eliminates overalpping and confusions in the tasks and duties.

4. Every employee contributes towards the achievement of the objectives by successfully completing the tasks and duties assigned to them by the superior.
5. It creates an open communication enviornment in the organization.

In a nutshell, Management by objectives is nothing but a process wherein the goals, plans and control system of the organization are defined by the management and employees jointly.

Bounded Rationality Model of Decision-Making Definition

There are two primary models or theories for decision-making: the Rational model and the Bounded rationality model. In the former, a decision-maker attempts to optimise the decision by selecting the best possible alternative. In the latter, rationality of individuals is limited by the information they have, cognitive limitations and time constraints.

Meaning of Organising:

Organising is a "process of defining the essential relationships among people, tasks and activities in such a way that all the organisation's resources are integrated and coordinated to accomplish its objectives efficiently and effectively". — Pearce and Robinson

Organising is, thus: (i) A Structure, and

(ii) A Process.

As a structure:

Organising is a set of relationships that defines vertical and horizontal relationships amongst people who perform various tasks and duties. The organisational task is divided into units, people in each unit (departments) are assigned specific tasks and their relationship is defined in a way that maximisesorganisational welfare and individual goals. The relationship amongst people is both vertical and horizontal.

As a process:

As a process, organising consists of two processes:

1) Differentiation,

(2) Integration.

Differentiation means division of work into smaller units and its assignment to individuals according to their skills and abilities. Integration refers to coordination of different activities towards a common goal. It provides unity of action towards **organisational activities**.

It involves:

(i) Identification of work,

(ii) Grouping of work into smaller groups,

(iii) Assigning work to every individual at every level in every department.

(iv) Defining its authority and responsibility, and

(v) Establishing relationships amongst people to make them contribute towards organisational goals in an integrated manner.

Organisation structure and process are not independent concepts. They are complementary to each other. Once the organisation process is defined, organisation structure is the end result or outcome of that process. Organisation structure is the result of organisation process. Organisation is, in fact, a structured, on-going process that defines how to achieve defined goals.

Principles of Organising:

Principles are the guidelines that promote managerial thinking and action. Principles help managers in effectively carrying out the organising function.

These principles are as follows:

(i) Principle of Unity of Objectives:

All organisational activities are geared towards organisational objectives. Objectives are framed for each level (top, middle and low) and each functional area. The objectives must be clearly understood by all. They should support each other at each level to attain objectives at higher levels.

(ii) Organisational Efficiency:

Organisational goals should be achieved efficiently. It means optimum (efficient) use of resources, that is, maximum output should be achieved with minimum inputs. The resources should be spread over activities in various functional areas that collectively result in maximum output through their optimum use.

(iii) Division of Labour:

Division of labour means breaking the main task into smaller units. The major task is broken into sub-tasks. This makes each person concentrate on his part of the job and perform it efficiently thereby, increasing the total output. Work should be divided and assigned to workers according to their skills. This leads to specialisation and contributes to organisational output.

(iv) Authority – Responsibility:

Authority and responsibility must go hand-in-hand. Responsibility means obligation to carry out the assigned task. To carry out this task, authority should be delegated to every person. Conversely, given the authority, the tasks assigned (responsibility) should be within the scope of authority. Authority without responsibility will result in misuse of authority and responsibility without authority will result in poor performance.

(v) Delegation:

The total work load is divided into parts. A part is assigned to subordinates and authority is given to efficiently carry out that task. Top managers delegate part of their duties to lower levels and concentrate on important organisational matters. This speeds up the organisational tasks and enables the organisation to grow in the dynamic, competitive business environment.

(vi) Scalar Chain:

Scalar chain is the line of authority running from top to lower levels. Authority flows from top to bottom in this chain and responsibilities flow from bottom to top. This chain promotes communication amongst people at different levels and facilitates decision making. Every person in the chain knows his superior and subordinate.

(vii) Span of Control:

Span of control means the number of subordinates that a superior can effectively supervise. Exact number of employees that a manager can supervise cannot be determined. It depends upon competence of managers, nature of work, system of control, capacity of subordinates etc.

However, if manager can supervise less number of workers, there will be more levels in the organisation structure and vice-versa. Supervising few subordinates creates tall structures and supervising large number of workers creates flat structures.

(viii) Unity of Command:

One subordinate should have one boss. People should receive orders from their immediate boss only. This brings discipline and order in the organisation. Receiving orders from two or more bosses can create confusion and indiscipline.

(ix) Balance:

There must be balance between different principles of organising. Balance should be maintained between centralisation and decentralisation, narrow and wide span of control etc.

(x) Flexibility:

Organisation should be flexible. Changes in structure should be made according to changes in the environmental factors.

(xi) Continuity:

Organisation should adapt to the environmental changes for its long-run survival, growth and expansion.

(xii) Exception:

Every matter should not be reported to top managers. Only significant deviations should be reported up the hierarchy. Routine matters should be dealt by middle and lower-level managers. It develops lower-level managers as they deal with simple and routine problems.

(xiii) Simplicity:

Organisation structure should be simple that can be understood by everyone. People can work efficiently in a simple structure as they are clear of various jobs and authority/ responsibility associated with those jobs. A simple structure promotes co-operation, coordination and effective communication in the organisation.

(xiv) Departmentation:

It means dividing activities into specialised groups (departments) where each department performs specialisedorganisational task. All activities of similar nature are grouped in one department headed by the departmental manager. Departments can be created on the basis of geographical locations, customers, products etc.

(xv) Decentralisation:

It means delegation of authority to lowest-level managers. It increases the decision-making authority of lower-level managers and increases organisational efficiency.

(xvi) Unity of Direction:

All activities of similar nature are grouped in one unit (production or marketing), headed by the departmental manager. He directs the efforts of departmental members towards a single objective; the departmental objective.

(xvii) Co-Operation:

All individuals and departments should co-operate and help the organisation achieve its goals. Cooperation leads to teamwork and focus on a unified goal.

Types of Organisation

he types are: 1. Line Organization 2. Line and Staff Organization 3.Functional Organization 4.Project Organization 5. Matrix Organization

Type # 1. Line Organisation:

Line organisation is the simplest and oldest form of organisation structure. It is called as military or departmental or scalar type of organization. Under this system, authority flows directly and vertically from the top of the managerial hierarchy ‘down to different levels of managers and subordinates and down to the operative level of workers.

Line organisation clearly identifies authority, responsibility and accountability at each level. The personnel in Line organization are directly involved in achieving the objectives of the organization.

Type # 2. Line and Staff Organization:

This type of organization structure is in large enterprises. The functional specialists are added to the line in line and staff organization. Mere, staff is basically advisory in nature and usually does not possess any command authority over line managers. Allen has defined line and staff organization as follows.

"Line functions are those which have direct responsibility for accomplishing the objectives of the enterprises and staff refers to those elements of the organization that help the line to work most effectively in accomplishing the primary objectives of the enterprises."

Type # 3. Functional Organisation:

The functional organisation was evolved by F.W. Taylor while he was working as a foreman. He suggested eight foremen, four in factory and four in planning division as under.

Factory Division:

(i) The gang boss,

(ii) The speed boss,

(iii) The inspector, and

(iv) The maintenance or repair boss.

Planning Division:

(i) Route Clerk,

(ii) Instruction card clerk,

(iii) Time and cost clerk, and

(iv) The shop disciplinarian.

He evolved his functional organisation system, which consists in "so dividing the work of management that each man, from the assistant superintendent down, shall have as few functions as possible to perform."

According to Terry, "Functional organisation refers to the organisation which is divided into a number of functions such as finance, production, sales, personnel, office and research and development and each of functions are performed by an expert". Line authority, staff authority and functional authority as a third type of authority are in this type of organisation.

Type # 4. Project Organisation:

This organisational structure are temporarily formed for specific projects for a specific period of time, for the project of achieving the goal of developing new product, the specialists from different functional departments such as production, engineering, quality control, marketing research etc., will be drawn to work together. These specialists go back to their respective duties as soon as the project is completed.

Really, the project organisation is set-up with the object of overcoming the major weakness of the functional organisation, such as absence of unity of command, delay in decision-making, and lack of coordination.

Type # 5. Matrix Organisation:

According to Stanley Davis and Paul Lawrence matrix organisation is “any organisation that employs a multiple command system that includes not only the multiple command structure, but also related support mechanism and an associated organisational culture and behaviour pattern.”

A matrix organisation, also referred to as the “multiple command system” has two chains of command. One chain of command is functional in which the flow of authority is vertical. The second chain is horizontal depicted by a project team, which is led by the project, or group manager who is an expert in his team’s assigned area of specialisation.

Authority Vs Responsibility

BASIS FOR COMPARISON	AUTHORITY	RESPONSIBILITY
Meaning	Authority refers to the power or right, attached to a particular job or designation, to give orders, enforce rules, make decisions and exact compliance.	Responsibility denotes duty or obligation to undertake or accomplish a task successfully, assigned by the senior or established by one's own commitment or circumstances.
What is it?	Legal right to issue orders.	Corollary of authority.
Results from	Formal positon in an organization	Superior-subordinate relationship
Task of manager	Delegation of authority	Assumption of responsibility

BASIS FOR COMPARISON	AUTHORITY	RESPONSIBILITY
Requires	Ability to give orders.	Ability to follow orders.
Flow	Downward	Upward
Objective	To make decisions and implement it.	To execute duties, assigned by superior.
Duration	Continues for long period.	Ends, as soon as the task is accomplished.

Delegation Vs Decentralization

BASIS FOR COMPARISON	DELEGATION	DECENTRALIZATION
Meaning	Delegation means handing over an authority from one person of high level to the person of low level.	Decentralization is the final outcome achieved, when the delegation of authority is performed systematically and repeatedly to the lowest level.
What it is?	Technique of management	Philosophy of management.
Accountability	Superiors are accountable for the acts done by subordinates.	Department heads are accountable for the acts of the concerned department.
Requirement	Yes, for all organization delegation of authority is very necessary.	No, it is an optional philosophy which may or may not be adopted by the organization.
Liberty of Work	Subordinates do not have full liberty.	A substantial amount of freedom is there.
Control	The ultimate control is the	The overall control vests with top

BASIS FOR COMPARISON	DELEGATION	DECENTRALIZATION
	hands of superior.	management and delegates authority for day to day control to departmental heads.
Relationship	Creates superior-subordinate relationship.	A step towards creation of semi-autonomous units.

Unit II

Staffing

Definition: Staffing is a managerial function which involves obtaining, utilising and retaining, qualified and competent personnel to fill all positions of an organisation, from top to operative echelon. In finer terms, staffing is placing the right person at the right job.

Importance of Staffing

- It helps in the finding out efficient and effective workforce, to fill different posts in the organisation.
- It improves organisation's performance and productivity by appointing the right person at the right job.
- It facilitates in identifying the staffing requirements of the organisation in future.
- It ensures continuous survival and growth of the organisation, by way of succession planning for executives.
- It develops personnel to take up top managerial positions of the organisation.
- It ensures training and development of the people working in the organisation.
- It assists the organisation in making the optimum use of human resources.

Therefore, staffing increases employee morale and job satisfaction. Further, it helps the top management in ascertaining the manpower requirement resulting from a promotion, transfer, employee turnover, retirement, etc. of the existing employees.

Motivation: Meaning, Definition, Nature and Types

Meaning:

Motivation is an important factor which encourages persons to give their best performance and help in reaching enterprise goals. A strong positive motivation will enable the increased output of employees but a negative motivation will reduce their performance. A key element in personnel management is motivation.

According to Likert, "It is the core of management which shows that every human being gives him a sense of worth in face-to face groups which are most important to him….A supervisor should strive to treat individuals with dignity and a recognition of their personal worth."

Nature of Motivation:

Motivation is a psychological phenomena which generates within an individual. A person feels the lack of certain needs, to satisfy which he feels working more. The need satisfying ego motivates a person to do better than he normally does.

From definitions given earlier the following inferences can be derived:

1. Motivation is an inner feeling which energizes a person to work more.
2. The emotions or desires of a person prompt him for doing a particular work.
3. There are unsatisfied needs of a person which disturb his equilibrium.
4. A person moves to fulfill his unsatisfied needs by conditioning his energies.
5. There are dormant energies in a person which are activated by channelizing them into actions.

Types of Motivation:

When a manager wants to get more work from his subordinates then he will have to motivate them for improving their performance. They will either be offered incentive for more work, or may be in the space of rewards, better reports, recognition etc., or he may instill fear in them or use force for getting desired work.

The following are the types of motivation:

1. Positive Motivation:

Positive motivation or incentive motivation is based on reward. The workers are offered incentives for achieving the desired goals. The incentives may be in the shape of more pay, promotion, recognition of work, etc. The employees are offered the incentives and try to improve their performance willingly.

According to Peter Drucker, the real and positive motivators are responsible for placement, high standard of performance, information adequate for self- control and the participation of the worker as a responsible citizen in the plant community. Positive motivation is achieved by the co-operation of employees and they have a feeling of happiness.

2. Negative Motivation:

Negative or fear motivation is based on force or fear. Fear causes employees to act in a certain way. In case, they do not act accordingly then they may be punished with demotions or lay-offs. The fear acts as a push mechanism. The employees do not willingly co-operate, rather they want to avoid the punishment.

Though employees work up-to a level where punishment is avoided but this type of motivation causes anger and frustration. This type of motivation generally becomes a cause of industrial unrest. In spite of the drawbacks of negative motivation, this method is commonly used to achieve desired results. There may be hardly any management which has not used negative motivation at one or the other time.

Importance of Motivation in an Organisation.

The process of motivation plays a very important role in any organisation, profit or non-profit. The managerial process of direction is driven primarily by the process of motivation as it creates within the mind of an employee the desire to work in the direction determined by the manager. The following aspects may be considered under this head:

- Increases Productivity: Motivation as a process leads to an increase in productivity of the employee. Motivation meets the needs of the employee and thereby creates the drive to work at the best of his abilities. A well-employee will be willing to put in more effort towards the betterment of the organisation than another disheartened employee.

- Ensures Organisational Efficiency: Motivation plays an important role in changing the attitudes of the employees in the organisation. Indifferent attitude is extinguished most efficiently by motivation. The presence of such favorable attitude allows the organisation to thrive and be successful.

- Ensures Loyal Workforce: A well-motivated workforce is a loyal workforce. Motivated employees have high levels of morale and commitment towards the organisation and its goals and objectives. Motivation thus reduces employee turnover and reduces the need for constant induction of new employees.

- Ensures a Reactive Workforce: Adapting to changing business environments is an important feature of any successful business. In order to react to changes easily and to continue smooth functioning, an organisation requires extensive loyalty and commitment of its employees. This reduces resistance to the changes that the organisation intends to make. This in effect makes the organisation efficient in adapting to changing needs.

- Facilitates Direction: Direction is an important managerial function and forms one of its core function. Motivation as already mentioned is a vital part of direction. Direction being a process that involves directing or initiating action according to a plan drawn up requires the employees to work wholeheartedly with commitment and loyalty. The process of direction is thus possible only when the employees proceed in the direction that the manager determines and this requires a motivated workforce.

THEORIES OF MOTIVATION

Motivation is a state-of-mind, filled with energy and enthusiasm, which drives a person to work in a certain way to achieve desired goals. Motivation is a force which pushes a person to work with high level of commitment and focus even if things are against him. Motivation translates into a certain kind of human behaviour.

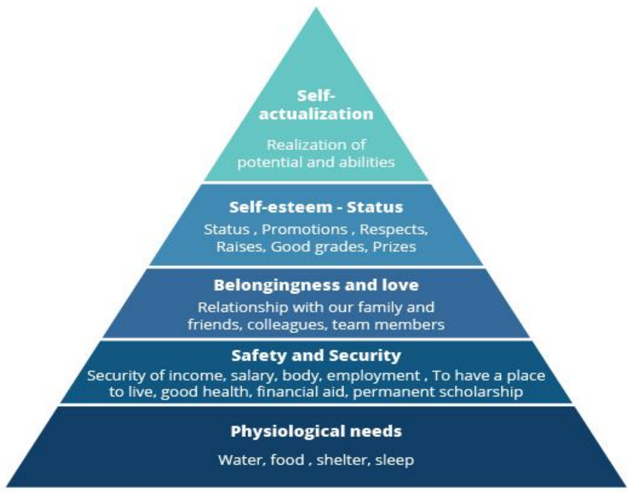
It is important to ensure that every team member in an organization is motivated. Various psychologists have studied human behaviour and have formalized their findings in the form various motivation theories. These motivation theories provide great understanding on how people behave and what motivates them.

Motivation is a huge field of study. There are many theories of motivation. Some of the famous motivation theories include the following:

1. Maslow's hierarchy of needs

Abraham Maslow postulated that a person will be motivated when his needs are fulfilled. The need starts from the lowest level basic needs and keeps moving up as a lower level need is fulfilled. Below is the hierarchy of needs:

- Physiological: Physical survival necessities such as food, water, and shelter.
- Safety: Protection from threats, deprivation, and other dangers.
- Social (belongingness and love): The need for association, affiliation, friendship, and so on.
- Self-esteem: The need for respect and recognition.
- Self-actualization: The opportunity for personal development, learning, and fun/creative/challenging work. Self-actualization is the highest level need to which a human being can aspire.



The leader will have to understand the specific need of every individual in the team and accordingly work to help fulfil their needs.
Herzberg's Motivation-Hygiene Theory

American psychologist and business management expert Frederick Herzberg's theory of motivation was developed in the 1950s-1960s as a way to understand employee motivation and satisfaction. Through his research, Herzberg identified factors repeatedly linked to satisfaction and dissatisfaction (otherwise known as hygiene factors).

The factors for satisfaction are:

- Achievement and recognition
- The work itself

- Responsibility
- Advancement and growth

The hygiene factors are:

- Company policies
- Supervision
- Relationship for supervisor and peers
- Work conditions
- Salary and status
- Security

The hygiene factors are mainly attributed to the workspace environment and what kind of constraints are put around the employees. Through these findings, Herzberg concluded that the most motivation creation occurs not just when hygiene factors are in order, but when hygiene factors are adequately addressed and there is great focus on satisfaction factors such as achievement and recognition. Put more simply, employees perform at their highest level when the work environment is healthy and they feel like they are achieving success and rewards in their job.

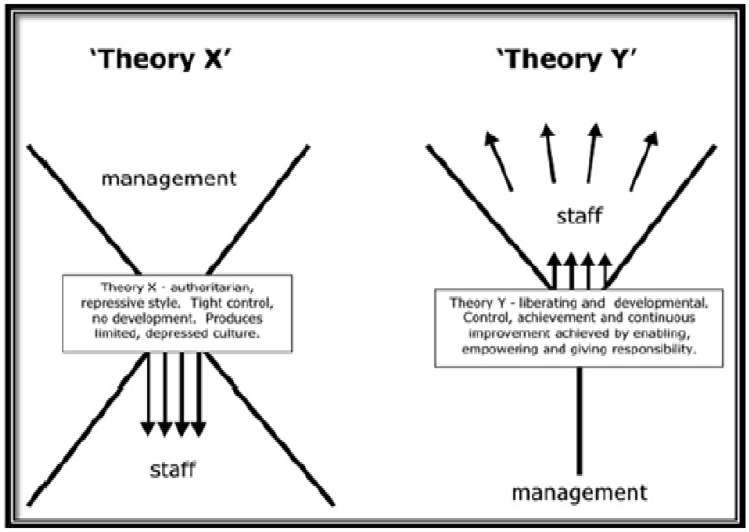
If you connect to this theory of motivation, then you may wish to focus on finding a work environment that satisfies all of these needs as you work toward achieving happiness inside and outside of your career.

Theory X and theory Y

Our management style is firmly influenced by our beliefs and assumptions about what encourages members of our team like: If we believe that our team members dislike work, then we tend towards an authoritarian style of management. However, if we assume that employees take pride in doing a good job, we tend to adopt a more participative style.

Douglas McGregor, the eminent social psychologist, divides management style into two contrasting theories −

- Theory X
- Theory Y



Theory X

This theory believes that employees are naturally unmotivated and dislike working, and this encourages an authoritarian style of management. According to this theory, management must firmly intervene to get things done. This style of management concludes that workers −

- Disfavor working.
- Abstain responsibility and the need to be directed.
- Need to be controlled, forced, and warned to deliver what's needed.
- Demand to be supervised at each and every step, with controls put in place.
- Require to be attracted to produce results, else they have no ambition or incentive to work.

McGregor observed that X-type workers are in fact mostly in minority, and yet in mass organizations, such as large scale production environment, X Theory management may be needed and can be unavoidable.

Theory Y

This theory explains a participative style of management, that is, distributive in nature. It concludes that employees are happy to work, are self-motivated and creative, and enjoy working with greater responsibility. It estimates that workers −

- Take responsibility willingly and are encouraged to fulfill the goals they are given.
- Explore and accept responsibility and do not need much guidance.
- Assume work as a natural part of life and solve work issues imaginatively.

In Y-type organizations, people at lower levels are engaged in decision making and have more responsibility.

Comparing Theory X & Theory Y

Let us now compare both the theories −

Motivation

Theory X considers that people dislike work, they want to avoid it and do not take responsibilities willingly.

While, Theory Y considers that people are self-motivated, and sportingly take responsibilities.

Management Style and Control

In Theory X-type organization, management is authoritarian, and centralized control is maintained.

While in Theory Y-type organization, the management style is participative, employees are involved decision making, but the power retains to implement decisions.

Work Organization

Theory X employees are specialized and the same work cycle continues.

In Theory Y, the work tends to be coordinated around wider areas of skill or knowledge. Employees are also motivated to develop expertise, and make suggestions and improvements.

Rewards and Appraisals

Theory X-type organizations work on a ‘carrot and stick’ basis, and performance assessment is part of the overall mechanism of control and compensation.

Coming to Theory Y-type organizations, appraisal is also regular and crucial, but is usually a separate mechanism from organizational controls. Theory Y-type organizations provide employees frequent opportunities for promotion.

Application

Admitting the fact that Theory X management style is widely accepted as inferior to others, it has its place in large scale production procedure and unskilled production-line work.

Many of the principles of Theory Y are widely accepted by different types of organization that value and motivate active participation.

Theory Y-style management is appropriate for knowledge work and licensed services. Licensed service organizations naturally develop Theory Y-type practices by the nature of their work, even high structure knowledge framework, like call center operations, benefit from its principles to motivate knowledge sharing and continuous improvement.

Definitions of Leadership:

"Leadership is both a process and property. The process of leadership is the use of non coercive influence to direct and coordinate the activities of the members of an organised group towards the accomplishment of group objectivities. As a property, leadership is the set of qualities or characteristics attributed to those who are perceived to successfully employ such influence." Gay and Strake

"Leadership is the activity of influencing people to strive willingly for group objectives." George Terry

Features of Leadership:

(i) Leadership is a process of influencing people.

(ii) It aims at bringing changes in behaviour of people.

(iii) It is a never ending or a continuous process.

(iv) Leadership is practiced to achieve organisational goals.

(v) It explains the relations between leaders and followers.

Importance of Leadership:

Leadership is an important factor in the success of an organisation. A leader not only provides information about organisational goals but also pools the required resources, guides and motivates his followers for achieving the organisational goals.

1.Influences the Behaviour of People:

Leadership involves influencing the behaviour of the people and making them contribute their full efforts for the good of an organisation.

(ii) Handles Conflicts:

A good leader helps in solving the conflicts among his followers and also removes their negative feelings.

(iii) Maintains Coordination:

A leader always maintains personal relations with his followers and provide them the required confidence, support and encouragement. This helps in creating healthy working environment.

(iv) Introduces Changes:

A good leader helps in solving the problem of resistance to changes by inspiring and persuading them to accept the changes without any discontentment.

(v) Provides Training:

A leader also makes arrangement for the training of his subordinates.

Traits of a leader

A leader has got multidimensional traits in him which makes him appealing and effective in behavior. The following are the requisites to be present in a good leader:

1. Physical appearance- A leader must have a pleasing appearance. Physique and health are very important for a good leader.
2. Vision and foresight- A leader cannot maintain influence unless he exhibits that he is forward looking. He has to visualize situations and thereby has to frame logical programmes.
3. Intelligence- A leader should be intelligent enough to examine problems and difficult situations. He should be analytical who weighs pros and cons and then summarizes the situation. Therefore, a positive bent of mind and mature outlook is very important.
4. Communicative skills- A leader must be able to communicate the policies and procedures clearly, precisely and effectively. This can be helpful in persuasion and stimulation.
5. Objective- A leader has to be having a fair outlook which is free from bias and which does not reflects his willingness towards a particular individual. He should develop his own opinion and should base his judgement on facts and logic.
6. Knowledge of work- A leader should be very precisely knowing the nature of work of his subordinates because it is then he can win the trust and confidence of his subordinates.
7. Sense of responsibility- Responsibility and accountability towards an individual's work is very important to bring a sense of influence. A leader must have a sense of responsibility towards organizational goals because only then he can get maximum of capabilities exploited in a real sense. For this, he has to motivate himself and arouse and urge to give best of his abilities. Only then he can motivate the subordinates to the best.
8. Self-confidence and will-power- Confidence in himself is important to earn the confidence of the subordinates. He should be trustworthy and should handle the situations with full will power. (You can read more about Self-Confidence at : Self Confidence - Tips to be Confident and Eliminate Your Apprehensions).

9. Humanist-This trait to be present in a leader is essential because he deals with human beings and is in personal contact with them. He has to handle the personal problems of his subordinates with great care and attention. Therefore, treating the human beings on humanitarian grounds is essential for building a congenial environment.
10. Empathy- It is an old adage “Stepping into the shoes of others”. This is very important because fair judgement and objectivity comes only then. A leader should understand the problems and complaints of employees and should also have a complete view of the needs and aspirations of the employees. This helps in improving human relations and personal contacts with the employees.

From the above qualities present in a leader, one can understand the scope of leadership and it’s importance for scope of business. A leader cannot have all traits at one time. But a few of them helps in achieving effective results.

Likert’s Four Systems of Management

Definition: RensisLikert along with his associates in Michigan University, USA conducted research to study the patterns and styles of managers over three decades, across 200 organizations and developed a four-fold model of the management system that helped in understanding the leadership behavior.

1. Exploitative Authoritative System: In this type of management system, the responsibility lies with the people in higher positions in the hierarchy. Here, the subordinates are not involved in the decision-making process. The superior has no trust and confidence in his subordinate and imposes decisions on him leaving no room for further discussions. In this system, the communication flows downwards, i.e. from the superior to the subordinate and hence there is a lack of communication and teamwork. The management is only concerned with the completion of work; it uses any means or threats to get the work completed through the subordinates.
2. Benevolent Authoritative System: Like exploitative authoritative system, here also the responsibility lies with the people at the upper echelons of the hierarchy and the only difference is that the motivation is based on the rewards, not on fear and threat. The superior has that much trust and confidence in his subordinates which is required in a master-servant relationship. In this system, the subordinates are given rewards for their participation and the communication may flow upwards i.e. from subordinate to superior, but restricted to what the superior wants to hear. Thus, in the benevolent

authoritative system also, the subordinates do not feel free to discuss job-related issues with the superior. This results in the lack of communication and a little teamwork.

3. Consultative System: In this management system, the superior has substantial but not complete, trust and confidence in his subordinates and constructively uses the views and opinions given by them. Here, the motivation is based on rewards and the amount of the individual's involvement in the decision-making process. The consultative system is characterized by a great flow of information both horizontally and vertically. The subordinates feel free to discuss job-related issues with the superiors and hence, the upward flow of communication is more into the consultative system than a benevolent system. But still, the decisions are made by the senior people in the hierarchy.
4. Participative System: In the participative system, the management has full confidence in his subordinates and encourages them to participate actively in the decision-making process. Here, the subordinate feels absolutely free to discuss any issue related to a job with his superior. This system is characterized by a good teamwork and teams are linked with people, who are the members of more than one team and such people are called as "linking pins". The subordinates get motivated through rewards for their participation in the decision-making process.

With these four systems of management, Likert studied seven variables Viz. Leadership, motivation, decision-making process, communication, interaction-influence, control process and goal setting.

Tannenbaum and Schmidt Continuum model

Tannenbaum and Schmidt Continuum model shows the relationship between the levels of freedom that a manager chooses to give to a team, and the level of authority used by the manager. As we already discussed in Situational leadership, number of parameters goes into choosing the managerial style: manager's competence, subordinate's developmental level, the situation.

Tannenbaum& Schmidt defined 7 levels of delegated freedom which moves from manager-oriented to subordinate-oriented. As team develops, level moves from one to the next – the area of freedom increases and the need for manager's intervention decreases. Following levels are self-explanatory and easy to understand:

1. Manager takes decision and announces it – only manager plays the decision-making role; no team involvement

2. Manager decides and then "Sells" his decision to the team – no change in decision; but team may raise some concerns

3. Manager presents decision with background ideas for the decision and invite questions – team knows what options manager considered for his decision; more team involvement

4. Manager suggests provisional decision & invites discussion regarding the decision – team can have a say on manager's decision; it can be changed based on discussion
5. Manager presents the problem or situation, get suggestions, then decides – team is free to come up with options; manager decides on those options
6. Manager explains the situation or problem, defines the parameters and asks team to decide on the solution – manager delegated whole thing to the team; but still manager is accountable for the outcome
7. Manager allows team to develop options and decide on the action, within the manager's received limit – complete freedom level; team does all the work almost as what the manager does at level 1.
The main advantage of this theory: for leaders/managers – it defines the criteria for involvement and delegation & range of choices for the involvement.

Controlling

Definition: Control is a primary goal-oriented function of management in an organisation. It is a process of comparing the actual performance with the set standards of the company to ensure that activities are performed according to the plans and if not then taking corrective action.

Every manager needs to monitor and evaluate the activities of his subordinates. It helps in taking corrective actions by the manager in the given timeline to avoid contingency or company's loss.
Controlling is performed at the lower, middle and upper levels of the management.

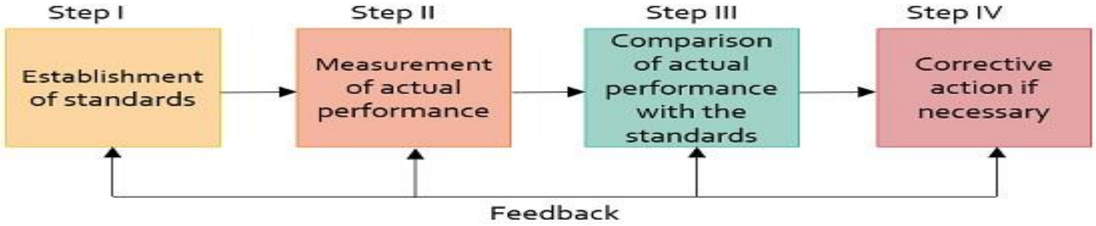
Features of Controlling

- An effective control system has the following features:
- It helps in achieving organizational goals.
- Facilitates optimum utilization of resources.
- It evaluates the accuracy of the standard.
- It also sets discipline and order.
- Motivates the employees and boosts employee morale.
- Ensures future planning by revising standards.
- Improves overall performance of an organization.
- It also minimises errors.

Controlling and planning are interrelated for controlling gives an important input into the next planning cycle. Controlling is a backwards-looking function which brings the management cycle back to the planning function. Planning is a forward-looking process as it deals with the forecasts about the future conditions.

Process of Controlling

Control process involves the following steps as shown in the figure:



- Establishing standards: This means setting up of the target which needs to be achieved to meet organisational goals eventually. Standards indicate the criteria of performance.

  Control standards are categorized as quantitative and qualitative standards. Quantitative standards are expressed in terms of money. Qualitative standards, on the other hand, includes intangible items.

- Measurement of actual performance: The actual performance of the employee is measured against the target. With the increasing levels of management, the measurement of performance becomes difficult.

- Comparison of actual performance with the standard: This compares the degree of difference between the actual performance and the standard.

- Taking corrective actions: It is initiated by the manager who corrects any defects in actual performance.

Controlling process thus regulates companies' activities so that actual performance conforms to the standard plan. An effective control system enables managers to avoid circumstances which cause the company's loss.

Types of control

There are three types of control viz.,

1. Feedback Control: This process involves collecting information about a finished task, assessing that information and improvising the same type of tasks in the future.
2. Concurrent control: It is also called real-time control. It checks any problem and examines it to take action before any loss is incurred. Example: control chart.

3. Predictive/ feedforward control: This type of control helps to foresee problem ahead of occurrence. Therefore action can be taken before such a circumstance arises.

In an ever-changing and complex environment, controlling forms an integral part of the organization.

Advantages of controlling

- Saves time and energy
- Allows managers to concentrate on important tasks. This allows better utilization of the managerial resource.
- Helps in timely corrective action to be taken by the manager.
- Managers can delegate tasks so routinely chores can be completed by subordinates.

On the contrary, controlling suffers from the constraint that the organization has no control over external factors. It can turn out to be a costly affair, especially for small companies.

Techniques of Managerial Control: Traditional and Modern Techniques

1. Traditional Techniques:

Traditional techniques refer to the techniques that have been used by business organisation for longer period of time and are still in use.

Such techniques are:

a. Personal Observation

b. Statistical Reports.

c. Breakeven Analysis.

d. Budgetary Control.

2. Modern Techniques:

Modem techniques are those techniques which are very new in management world. These techniques provide various new aspects for controlling the activities of an organisation.

These techniques are as follows:

(a) Return on Investment.

(b) Ratio Analysis.

(c) Responsibility Accounting.

(d) Management Audit.

(e) PERT and CPM.

(f) Management Information System.

Effective Organizational Control Systems

The management of any organization must develop a control system tailored to its organization's goals and resources. Effective control systems share several common characteristics. These characteristics are:

- A focus on critical points. For example, controls are applied where failure cannot be tolerated or where costs cannot exceed a certain amount. The critical points include all the areas of an organization's operations that directly affect the success of its key operations.
- Integration into established processes. Controls must function harmoniously within these processes and should not bottleneck operations.
- Acceptance by employees. Employee involvement in the design of controls can increase acceptance.
- Availability of information when needed. Deadlines, time needed to complete the project, costs associated with the project, and priority needs are apparent in these criteria. Costs are frequently attributed to time shortcomings or failures.
- Economic feasibility. Effective control systems answer questions such as, “How much does it cost?” “What will it save?” or “What are the returns on the investment?” In short, comparison of the costs to the benefits ensures that the benefits of controls outweigh the costs.
- Accuracy. Effective control systems provide factual information that's useful, reliable, valid, and consistent.
- Comprehensibility. Controls must be simple and easy to understand.

Unit III

Organizational Behavior

Organizational Behavior (OB) can be defined as the understanding, prediction and management of human behavior both individually or in a group that occur within an organization.

Internal and external perspectives are the two theories of how organizational behavior can be viewed from an organization’s point of view. In this tutorial, we will be learning in detail about both the theories.

Importance of OB

While working in an organization, it is very important to understand others behavior as well as make others understand ours. In order to maintain a healthy working environment, we need to adapt to the environment and understand the goals we need to achieve. This can be done easily if we understand the importance of OB.

Following points bring out the importance of OB −

- It helps in explaining the interpersonal relationships employees share with each other as well as with their higher and lower subordinates.
- The prediction of individual behavior can be explained.
- It balances the cordial relationship in an enterprise by maintaining effective communication.
- It assists in marketing.
- It helps managers to encourage their sub-ordinates.
- Any change within the organization can be made easier.
- It helps in predicting human behavior & their application to achieve organizational goals.
- It helps in making the organization more effective.

There are three major factors that affect OB. The working environment being the base for all three factors, they are also known as the determinants of OB. The three determinants are −

- People
- Structure
- Technology

People

An organization consists of people with different traits, personality, skills, qualities, interests, background, beliefs, values and intelligence. In order to maintain a healthy environment, all the employees should be treated equally and be judged according to their work and other aspects that affects the firm.

Example − A company offers campus placement to trainees from different states like Orissa, Haryana, Arunachal Pradesh and many more. However, during and after training, all trainees are examined only on the basis of their performance in the tasks assigned.

Organizational Structure

Structure is the layout design of an organization. It is the construction and arrangement of relationships, strategies according to the organizational goal.

Example − Organizational structure defines the relation of a manager with employees and co-workers.

Technology

Technology can be defined as the implementation of scientific knowledge for practical usage. It also provides the resources required by the people that affect their work and task performance in the right direction.

Example − Introduction of SAP, big data and other software in the market determines individual and organizational performance.

Environment

All companies function within a given internal and external environment. Internal environment can be defined as the conditions, factors, and elements within an enterprise that influences the activities, choices made by the firm, and especially the behavior of the employees. While external environment can be defined as outside factors that affect the company's ability to operate. Some of them can be manipulated by the company's marketing, while others require the company to make adjustments.

Some examples of internal environment include employee morale, culture changes, financial changes or issues, and some examples of external environment include political factors, changes to the economy and the company itself.

The concept of OB is based on two key elements namely −

- Nature of people
- Nature of the organization

Nature of People

In simple words, nature of people is the basic qualities of a person, or the character that personifies an individual they can be similar or unique. Talking at the organizational level, some major factors affecting the nature of people have been highlighted. They are −

- Individual Difference − It is the managerial approach towards each employee individually, that is one-on-one approach and not the statistical approach, that is, avoidance of single rule. Example− Manager should not be biased towards any particular employee rather should treat them equally and try not to judge anyone on any other factor apart from their work.

- Perception − It is a unique ability to observe, listen and conclude something. It is believing in our senses. In short, the way we interpret things and have our point of view is our perception. Example − Aman thinks late night parties spoil youth while Anamika thinks late night parties are a way of making new friends. Here we see both Aman and Anamika have different perception about the same thing.
- A whole person − As we all know that a person's skill or brain cannot be employed we have to employee a whole person. Skill comes from background and knowledge. Our personal life cannot be totally separated from our work life, just like emotional conditions are not separable from physical conditions. So, people function is the functioning of a total human being not a specific feature of human being.
- Motivated behavior − It is the behavior implanted or caused by some motivation from some person, group or even a situation. In an organization, we can see two different types of motivated employees −
  - Positive motivation − Encouraging others to change their behavior or say complete a task by luring them with promotions or any other profits. Example − "If you complete this, you will gain this."
  - Negative motivation − Forcing or warning others to change their behavior else there can be serious consequences. Example − "If you don't complete this, you will be deprived from the office."
- Value of person − Employees want to be valued and appreciated for their skills and abilities followed by opportunities which help them develop themselves.

Nature of Organization

Nature of organization states the motive of the firm. It is the opportunities it provides in the global market. It also defines the employees' standard; in short, it defines the character of the company by acting as a mirror reflection of the company. We can understand the nature of any firm with its social system, the mutual interest it shares and the work ethics.

Let us take a quick look at all these factors −

- Social system − Every organization socializes with other firms, their customers, or simply the outer world, and all of its employees - their own social roles and status. Their behavior is mainly influenced by their group as well as individual drives. Social system are of two types namely −
    - Formal − Groups formed by people working together in a firm or people that belong to the same club is considered as formal social system. Example − A success party after getting a project.
    - Informal − A group of friends, people socializing with others freely, enjoying, partying or chilling. Example − Birthday party.
- Mutual interest − Every organization needs people and people need organizations to survive and prosper. Basically it's a mutual understanding between the organization and the employees that helps both reach their respective objectives. Example − We deposit our money in the bank, in return the bank gives us loan, interest, etc.
- Ethics − They are the moral principles of an individual, group, and organization. In order to attract and keep valuable employees, ethical treatment is necessary and some moral standards need to be set. In fact, companies are now establishing code of ethics training reward for notable ethical behavior.
- In a very broad sense, the scope of OB is the extent to which it can govern or influence the operations of an organization. The scope of OB integrates 3 concepts respectively −
- **Individual Behavior**
- It is the study of individual's personality, learning, attitudes, motivation, and job satisfaction. In this study, we interact with others in order to study about them and make our perception about them.
- Example − The personal interview round is conducted to interact with candidates to check their skills, apart from those mentioned in the resume.
- **Inter-individual Behavior**
- It is the study conducted through communication between the employees among themselves as well as their subordinates, understanding people's leadership qualities, group dynamics, group conflicts, power and politics.

- Example − A meeting to decide list of new board members.
- **Group Behavior**
- Group behavior studies the formation of organization, structure of organization and effectiveness of organization. The group efforts made towards the achievement of organization's goal is group behavior. In short, it is the way how a group behaves.
- Example − Strike, rally etc.

Functions of a manager are the various roles played by the manager in an organization. A manager is accountable for all the happenings in the firm and is answerable to the management. The seven major roles played by the manager are −

- Planning
- Organizing
- Staffing
- Directing/leading
- Coordinating
- Reporting
- Budgeting
- Controlling

Various Challenges of a Manager

We have seen the different roles a manager as to play in order to maintain the workflow balance in an organization. With all these responsibilities, there are some tough challenges a manager has to deal with while trying to balance everything. Following are some challenges a manager has to deal with −

- Managing workforce diversity − Manager shouldn't create or encourage discrimination among employees. Employees from different background, culture, and ethnicity should be treated as equal and rewards should be given only on the basis of work.
- Improving quality and productivity − It is the sole responsibility of the manager to increase the productivity without hampering the quality. It can be done in two ways −
    - Totally quality management − That is constant focus on customer satisfaction by improving organizational process.

    - Process of engineering − Focusing on the manufacturing of the product, so that the quality is not compromised.

- Responding to labor storage − If there is a labor shortage then the manager should quickly respond to solve this problem by arranging for the workforce required so that the product delivery is not delayed.
- Eradication of labor shortage − The manager needs to take quick action, if there is a labor shortage and should assure with backup plans so that there is no labor shortage in future.
- Improving customer service − Manager faces the challenge to constantly improve customer service to survive in an ever-competitive environment.
- Improving ethical behavior − Managers should make sure that the employees behave properly and maintain the decorum of the company. These are few major challenges a manager faces while trying to complete a project. To maintain work-life balance and for the betterment of the organization, the manager should try level best to resolve these challenges.
- Organizational behavior reflects the behavior of the people and management all together, it is considered as field study not just a discipline. A discipline is an accepted science that is based upon theoretical foundation, whereas OB is an inter-disciplinary approach where knowledge from different disciplines like psychology, sociology, anthropology, etc. are included. It is used to solve organizational problems, especially those related to human beings.
- There are four different types of models in OB. We will throw some light on each of these four models.
- **Autocratic Model**
- The root level of this model is power with a managerial orientation of authority. The employees in this model are oriented towards obedience and discipline. They are dependent on their boss. The employee requirement that is met is subsistence. The performance result is less.
- The major drawbacks of this model are people are easily frustrated, insecurity, dependency on the superiors, minimum performance because of minimum wage.

- **Custodial Model**
- The root level of this model is economic resources with a managerial orientation of money. The employees in this model are oriented towards security and benefits provided to them. They are dependent on the organization. The employee requirement that is met is security.
- This model is adapted by firms having high resources as the name suggest. It is dependent on economic resources. This approach directs to depend on firm rather than on manager or boss. They give passive cooperation as they are satisfied but not strongly encouraged.
- **Supportive Model**
- The root level of this model is leadership with a managerial orientation of support. The employees in this model are oriented towards their job performance and participation. The employee requirement that is met is status and recognition. The performance result is awakened drives.
- This model is dependent on leadership strive. It gives a climate to help employees grow and accomplish the job in the interest of the organization. Management job is to assist the employee's job performance. Employees feel a sense of participation.
- **Collegial Model**
- The root level of this model is partnership with a managerial orientation of teamwork. The employees in this model are oriented towards responsible behavior and self-discipline. The employee requirement that is met is self-actualization. The performance result is moderate zeal.
- This is an extension of supportive model. The team work approach is adapted for this model. Self-discipline is maintained. Workers feel an obligation to uphold quality standard for the better image of the company. A sense of "accept" and "respect" is seen.

Individual and interpersonal behaviour

Individual- Knowing yourself and managing behavior

- The ability to make decisions and manage perceptions is essential for achieving social and business success. Cultivating self-awareness is important in navigating career paths. By knowing yourself, your strengths and weaknesses, values, biases, and perceptions, you are best able to put your right foot forward, work in-sync with coworkers, manage and guide direct reports, project success to your superiors, and gauge how you can better perform and ascend the corporate ladder.

- Self-awareness and management also leads to recognition of biases and perception. Perception is a process in which an individual organizes and interprets their sensory impressions to give meaning to their environment (According to Essentials of Organizational Behavior, Perception and Individual Decision Making book). This plays a big role in how we think. Knowing and understanding implicit biases may also help you recognize certain patterns or inclinations. Background and upbringing also factor into our thought-processes, perceptions, and biases. We attribute behavior and outcomes to internal (person or dispositional) and external (situational) causes and motives.
- Interpersonal- Associations between two or more people
- Interpersonal relationships are developed through verbal and nonverbal communication. Strong interpersonal skills are characterized through an ability to influence and persuade (sometimes described as charisma), but also hinges on cultural understanding and global awareness.
- In a business context, the term interpersonal relationship refers to how people communicate and interact within their organization. The ability to influence and communicate is sometimes regarded as innate, but for most of us, it is something that requires hard work to strengthen as we grow. Learning about interpersonal theory, which is the study of characteristics, traits, and patterns, is an effective tool for managing day-to-day interactions.

Personality

The word personality is derived from a Greek word "persona" which means "to speak through." Personality is the combination of characteristics or qualities that forms a person's unique identity. It signifies the role which a person plays in public. Every individual has a unique, personal and major determinant of his behavior that defines his/her personality.

Personality trait is basically influenced by two major features −

- Inherited characteristics
- Learned characteristics

Inherited Characteristics

The features an individual acquires from their parents or forefathers, in other words the gifted features an individual possesses by birth is considered as inherited characteristics. It consists of the following features −

- Color of a person's eye
- Religion/Race of a person
- Shape of the nose
- Shape of earlobes

Learned Characteristics

Nobody learns everything by birth. First, our school is our home, then our society, followed by educational institutes. The characteristics an individual acquires by observing, practicing, and learning from others and the surroundings is known as learned characteristics.

Learned characteristics includes the following features −

- Perception − Result of different senses like feeling, hearing etc.
- Values − Influences perception of a situation, decision making process.
- Personality − Patterns of thinking, feeling, understanding and behaving.
- Attitude − Positive or negative attitude like expressing one's thought.

Traits of Personality

Personality traits are the enduring features that define an individual's behavior. A personality trait is a unique feature in an individual. Psychologists resolved that there are five major personality traits and every individual can be categorized into at least one of them. These five personality traits are −

- Extrovert
- Neurotic
- Open
- Agreeable
- Conscientious

Major Personality Attributes

Following are the five major personality attributes that influence OB −

Locus of Control

Locus of control is the center of control of an individual’s code of conduct. People can be grouped into two categories i.e., internals and externals respectively.

People who consider themselves as the masters of their own fates are known as internals, while, those who affirm that their lives are controlled by outside forces known as externals.

Before making any decision, internals actively search for information, they are achievement driven, and want to command their environment. Thus, internals do well on jobs that craves complex information processing, taking initiative and independent action.

Externals, on the other hand, are more compliant, more willing to follow instructions, so, they do well in structured, routine jobs.

Machiavellianism

Machiavellianism is being practical, emotionally distant, and believing that ends justify means.

Machiavellians are always wanting to win and are great persuaders. Here are the significant features of a high-mach individuals −

- High-Machs prefer precise interactions rather than beating about the bush.
- High-Machs tend to improvise; they do not necessarily abide by rules and regulations all the time.
- High-Machs get distracted by emotional details that are irrelevant to the outcome of a project.

Self-esteem

It is the extent up to which people either like or dislike themselves. Self-Esteem is directly related to the expectations of success and on-the-job satisfaction.

Individuals with high self-esteem think that they have what it takes to succeed. So, they take more challenges while selecting a job.

On the other hand, individuals with low self-esteem are more susceptible to external distractions. So, they are more likely to seek the approval of others and to adapt the beliefs and behaviors of those they respect.

Self-monitoring

Self-monitoring is the capability of regulating one's behavior according to social situations. Individuals with high self-monitoring skill easily adjust their behavior according to external, situational factors. Their impulsive talents allow them to present public personae which are completely different from their private personalities.

However, people with low self-monitoring skills cannot cover themselves. Regardless of any situation, they are always themselves. They have an attitude of, "what you see is what you get."

Risk taking

Generally, managers are reluctant on taking risks. However, individual risk-taking inclination affects the bulk of information required by the managers and how long it takes them to make decisions.

Thus, it is very important to recognize these differences and align risk-taking propensity with precise job demands that can make sense.

Perception

Perception is an intellectual process of transforming sensory stimuli to meaningful information. It is the process of interpreting something that we see or hear in our mind and use it later to judge and give a verdict on a situation, person, group etc.

It can be divided into six types −

- Of sound − The ability to receive sound by identifying vibrations.
- Of speech − The competence of interpreting and understanding the sounds of language heard.
- Touch − Identifying objects through patterns of its surface by touching it.
- Taste − The ability to receive flavor of substances by tasting it through sensory organs known as taste buds.
- Other senses − They approve perception through body, like balance, acceleration, pain, time, sensation felt in throat and lungs etc.
- Of the social world − It permits people to understand other individuals and groups of their social world. Example − Priya goes to a restaurant and likes their customer service, so she will

perceive that it is a good place to hang out and will recommend it to her friends, who may or may not like it. Priya's perception about the restaurant is good.

Perceptual Process

Perceptual process are the different stages of perception we go through. The different stages are −

- Receiving
- Selecting
- Organizing
- Interpreting

Receiving

Receiving is the first and most important stage in the process of perception. It is the initial stage in which a person collects all information and receives the information through the sense organs.

Selecting

Selecting is the second stage in the process. Here a person doesn't receive the data randomly but selectively. A person selects some information out of all in accordance with his interest or needs. The selection of data is dominated by various external and internal factors.

- External factors − The factors that influence the perception of an individual externally are intensity, size, contrast, movement, repetition, familiarity, and novelty.
- Internal factors − The factors that influence the perception of an individual internally are psychological requirements, learning, background, experience, self-acceptance, and interest.

Organizing

Keeping things in order or say in a synchronized way is organizing. In order to make sense of the data received, it is important to organize them.

We can organize the data by −

- Grouping them on the basis of their similarity, proximity, closure, continuity.
- Establishing a figure ground is the basic process in perception. Here by figure we mean what is kept as main focus and by ground we mean background stimuli, which are not given attention.
- Perceptual constancy that is the tendency to stabilize perception so that contextual changes don't affect them.

Interpreting

Finally, we have the process of interpreting which means forming an idea about a particular object depending upon the need or interest. Interpretation means that the information we have sensed and organized, is finally given a meaning by turning it into something that can be categorized. It includes stereotyping, halo effect etc.

Importance of Perception in OB

We need to understand what the role of perception in an organization is. It is very important in establishing different role of perceptions like −

- Understanding the tasks to be performed.
- Understanding associated importance of tasks allotted.
- Understanding preferred behavior to complete respective tasks.
- Clarifying role perceptions.

For example, every member in a group has to be clear regarding the role allotted to them. Programmer writes the code, tester checks it, etc.

Learning

Learning can be defined as the permanent change in behavior due to direct and indirect experience. It means change in behavior, attitude due to education and training, practice and experience. It is completed by acquisition of knowledge and skills, which are relatively permanent.

Nature of Learning

Nature of learning means the characteristic features of learning. Learning involves change; it may or may not guarantee improvement. It should be permanent in nature, that is learning is for lifelong.

The change in behavior is the result of experience, practice and training. Learning is reflected through behavior.

Factors Affecting Learning

Learning is based upon some key factors that decide what changes will be caused by this experience. The key elements or the major factors that affect learning are motivation, practice, environment, and mental group.

Coming back to these factors let us have a look on these factors −

- Motivation − The encouragement, the support one gets to complete a task, to achieve a goal is known as motivation. It is a very important aspect of learning as it acts gives us a positive energy to complete a task. Example − The coach motivated the players to win the match.
- Practice − We all know that ”Practice makes us perfect”. In order to be a perfectionist or at least complete the task, it is very important to practice what we have learnt. Example − We can be a programmer only when we execute the codes we have written.
- Environment − We learn from our surroundings, we learn from the people around us. They are of two types of environment – internal and external. Example − A child when at home learns from the family which is an internal environment, but when sent to school it is an external environment.
- Mental group − It describes our thinking by the group of people we chose to hang out with. In simple words, we make a group of those people with whom we connect. It can be for a social cause where people with the same mentality work in the same direction. Example − A group of readers, travelers, etc.

These are the main factors that influence what a person learns, these are the root level for our behavior and everything we do is connected to what we learn.

How Learning Occurs**?**

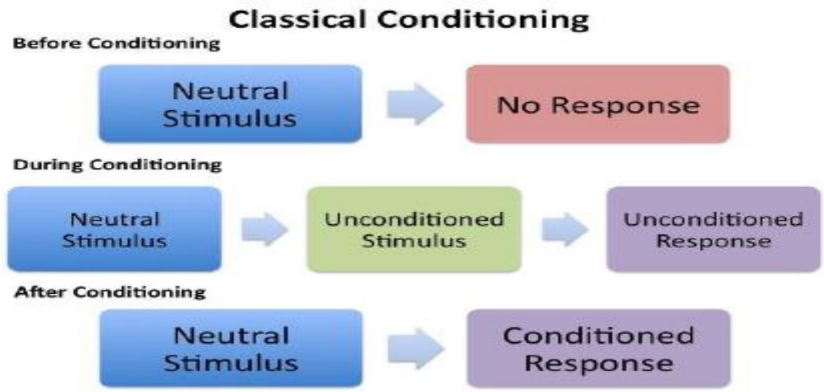
Learning can be understood clearly with the help of some theories that will explain our behavior. Some of the remarkable theories are −

- Classical Conditioning Theory
- Operant Conditioning Theory
- Social Learning Theory
- Cognitive Learning Theory

Classical Conditioning Theory

The classical conditioning occurs when a conditioned stimulus is coupled with an unconditioned stimulus. Usually, the conditioned stimulus (CS) is an impartial stimulus like the sound of a tuning fork, the unconditioned stimulus (US) is biologically effective like the taste of food and the unconditioned response (UR) to the unconditioned stimulus is an unlearned reflex response like salivation or sweating.

After this coupling process is repeated (for example, some learning may already occur after a single coupling), an individual shows a conditioned response (CR) to the conditioned stimulus, when the conditioned stimulus is presented alone. The conditioned response is mostly similar to the unconditioned response, but unlike the unconditioned response, it must be acquired through experience and is nearly impermanent.



Operant Conditioning Theory

Operant conditioning theory is also known as instrumental conditioning. This theory is a learning process in which behavior is sensitive to, or controlled by its outcomes.

Let’s take an example of a child. A child may learn to open a box to get the candy inside, or learn to avoid touching a hot stove. In comparison, the classical conditioning develops a relationship between a stimulus and a behavior. The example can be further elaborated as the child may learn to salivate at the sight of candy, or to tremble at the sight of an angry parent.

In the 20th century, the study of animal learning was commanded by the analysis of these two sorts of learning, and they are still at the core of behavior analysis.

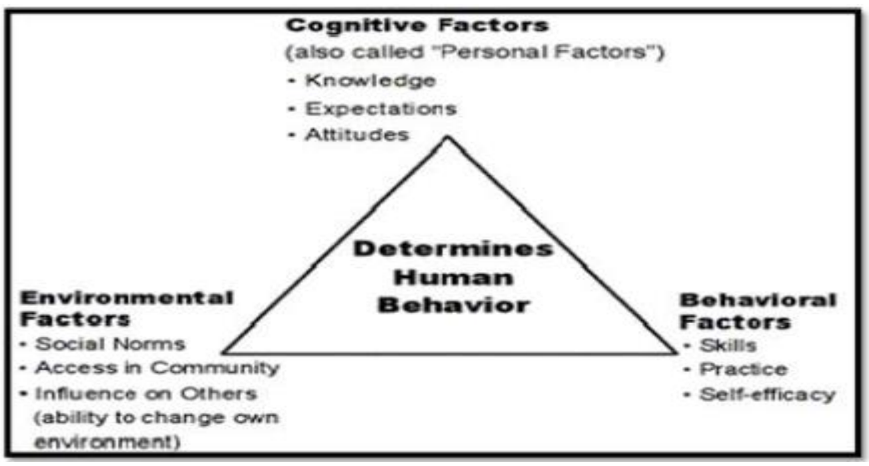
Operant Conditioning

	Reinforcement Increase Behavior	Punishment Decrease Behavior
Positive Stimulus (something added)	Positive Reinforcement Add something to increase behavior	Positive Punishment Add something to decrease behavior
Negative Stimulus (something removed)	Negative Reinforcement Remove something to increase behavior	Negative Punishment Remove something to decrease behavior

Social Learning Theory

The key assumptions of social learning theory are as follows −

- Learning is not exactly behavioral, instead it is a cognitive process that takes place in a social context.
- Learning can occur by observing a behavior and by observing the outcomes of the behavior (known as vicarious reinforcement).
- Learning includes observation, extraction of information from those observations, and making decisions regarding the performance of the behavior (known as observational learning or modeling). Thus, learning can occur beyond an observable change in behavior.
- Reinforcement plays an important role in learning but is not completely responsible for learning.
- The learner is not a passive receiver of information. Understanding, environment, and behavior all mutually influence each other.



Cognitive Learning Theory

Cognition defines a person's ideas, thoughts, knowledge, interpretation, understanding about himself and environment.

This theory considers learning as the outcome of deliberate thinking on a problem or situation based upon known facts and responding in an objective and more oriented manner. It perceives that a person learns the meaning of various objects and events and also learns the response depending upon the meaning assigned to the stimuli.

This theory debates that the learner forms a cognitive structure in memory which stores organized information about the various events that occurs.

Learning & Organizational Behavior

An individual's behavior in an organization is directly or indirectly affected by learning.

Example − Employee skill, manager's attitude are all learned.

Behavior can be improved by following the listed tips −

- Reducing absenteeism by rewarding employees for their fair attendance.
- Improving employee discipline by dealing with employee's undesirable behavior, drinking at workplace, stealing, coming late, etc. by taking appropriate actions like oral reprimands, written warnings and suspension.
- Developing training programs more often so as to grab the trainees' attention, provide required motivational properties etc.

Attitude: Nature, Components and Formation

Following are the salient features which contribute to the meaning of attitudes:

1. Attitudes refer to feelings and beliefs of individuals or groups of individuals. For example "He has a poor attitude", "I like her attitude."
2. The feeling's and beliefs are directed towards other people, objects or ideas. When a person says, "I like my Job". It shows that he has a positive attitude towards his job.
3. Attitudes often result in and affect the behaviour or action of the people. Attitudes can lead to intended behaviour if there are no external interventions.
4. Attitudes constitute a psychological phenomenon which cannot be directly observed. However, an attitude can be observed indirectly by observing its consequences. For example, if a person is very regular in his job, we may infer that he likes his job very much.
5. Attitudes are gradually acquired over a period of time. The process of learning attitude starts right from childhood and continues throughout the life of a person. In the beginning the family members may have a greater impact on the attitude of a child.
6. Attitudes are evaluative statements, either favourable or unfavourable. When a person says he likes or dislikes something or somebody, an attitude is being expressed.
7. All people, irrespective of their status and intelligence hold attitudes.

8. An attitude may be unconsciously held. Most of our attitudes may be about those which we are not clearly aware. Prejudice furnishes a good example.

Components of Attitudes:

Attitudes comprise of three basic components: emotional, informational and behavioural.

These three components are described below:

1. Informational or Cognitive Component:

The informational component consists of beliefs, values, ideas and other information a person has about the object. It makes no difference whether or not this information is empirically correct or real. For example, a person seeking a job may learn from his own sources and other employees working in the company that in a particular company the promotion chances are very favourable. In reality, it may or may not be correct. Yet the information that person is using is the key to his attitude about that job and about that company.

2. Emotional or Affective Component:

The informational component sets the stage for the more critical part of an attitude, its affective component. The emotional components involve the person's feeling or affect-positive, neutral or negative-about an object. This component can be explained by this statement." I like this job because the future prospects in this company are very good".

3. Behavioural Component:

The behavioural component consists of the tendency of a person to behave in a particular manner towards an object. For example, the concerned individual in the above case may decide to take up the job because of good future prospects. Out of the three components of attitudes, only the behavioural component can be directly observed. One cannot see another person's beliefs (the informational component) and his feelings (the emotional component). These two components can only be inferred. But still understanding these two components is essential in the study of organisationalbehaviour or the behavioural component of attitudes.

BASIS FOR COMPARISON	ATTITUDE	BEHAVIOR
Meaning	Attitude refers to a person's mental view, regarding the way he/she thinks or feels about someone or something.	Behavior implies the actions, moves, conduct or functions of an individual or group towards other persons.

BASIS FOR COMPARISON	ATTITUDE	BEHAVIOR
Based on	Experience and observation	Situation
Trait	Human	Inborn
What is it?	A person's mindset.	Outward expression of attitude.
Reflects	What you think or feel?	What you do?
Defined by	Way we perceive things.	Social Norms

The Johari Window Model

The Johari window model is used to enhance the individual's perception on others. This model is based on two ideas- trust can be acquired by revealing information about you to others and learning yourselves from their feedbacks. Each person is represented by the Johari model through four quadrants or window pane. Each four window panes signifies personal information, feelings, motivation and whether that information is known or unknown to oneself or others in four viewpoints.
The Johari Window Model

The Johari Window Model

The method of conveying and accepting feedback is interpreted in this model. A Johari is represented as a common window with four panes. Two of these panes represent self and the other two represent the part unknown to self but to others. The information transfers from one pane to the other as the result of mutual trust which can be achieved through socializing and the feedback got from other members of the group.

1. Open/self-area or arena – Here the information about the person his attitudes, behaviour, emotions, feelings, skills and views will be known by the person as well as by others. This is mainly the area where all the communications occur and the larger the arena becomes the more effectual and dynamic the relationship will be. 'Feedback solicitation' is a process which occurs by understanding and listening to the feedback from another person. Through this way the open area can be increased horizontally decreasing the blind spot. The size of the arena can also be increased downwards and thus by reducing the hidden and unknown areas through revealing one's feelings to other person.

2. Blind self or blind spot – Information about yourselves that others know in a group but you will be unaware of it. Others may interpret yourselves differently than you expect. The blind spot is reduced for an efficient communication through seeking feedback from others.

3. Hidden area or façade – Information that is known to you but will be kept unknown from others. This can be any personal information which you feel reluctant to reveal. This includes feelings, past experiences, fears, secrets etc. we keep some of our feelings and information as private as it affects the relationships and thus the hidden area must be reduced by moving the information to the open areas.

4. Unknown area – The Information which are unaware to yourselves as well as others. This includes the information, feelings, capabilities, talents etc. This can be due to traumatic past experiences or events which can be unknown for a lifetime. The person will be unaware till he discovers his hidden qualities and capabilities or through observation of others. Open communication is also an effective way to decrease the unknown area and thus to communicate effectively.

What Is Transactional Analysis?

Transactional analysis is a social psychology developed by Eric Berne, MD (d.1970). Berne's theory consists of certain key concepts that practitioners use to help clients, students, and systems analyze and change patterns of interaction that interfere with achieving life aspirations. Over the past 40 years, Berne's theory has evolved to include applications in counseling, education, organizational development, and psychotherapy. Research studies have evaluated the effectiveness of transactional analysis in a wide variety of contexts. (See also training and certification in transactional analysis.)

- The counseling specialization is chosen by professionals working in such diverse contexts as social welfare, health care, pastoral work, prevention, mediation, process facilitation, multicultural work, and humanitarian activities, to name a few.
- Educational transactional analysis is used by practitioners working in training centers, preschools, elementary and high schools, universities, and institutions that prepare teachers and trainers as well as in support of learners of all ages to thrive within their families, organizations, and communities.
- Organizational transactional analysts work in, or for, organizations using transactional analysis concepts and techniques to evaluate an organization's developmental processes and challenges as well as its dysfunctional behaviors.
- Psychotherapists utilize transactional analysis to facilitate their clients' capacities for self-actualization and healing by learning to recognize and change old, self-limiting patterns.

Ego States

Definition: The Ego States are an important aspect of transactional analysis that talks about how a person feels, behave or think at any point of time.

let's understand each of these three ego states in detail:

Parent Ego: The parent ego, refers to the behavior and attitude of an emotionally significant individual who acted with quite a maturity when he was a child. He possesses the parental traits of being overprotective, dogmatic, distant, indispensable and upright and behaves very judiciously at any time.

There are two types of a parent ego: critical and nurturing. The critical parent ego is one when an individual shows the critical and evaluative behavior while interacting with the others. Whereas the nurturing parent ego is one, when individual shows the kind and nurturing behavior, not only towards children but towards all with whom he interacts.

Adult Ego: The adult ego shows the logical thinking and reasoning ability of an individual. The person behaving or interacting with adult ego seeks all the information properly, validate it using his reasoning skills and then provide it to the other people. The person possessing the adult ego can be judged through his discussions and the way he thinks about a situation before arriving at the conclusion.

As the individual grows, he updates his parent data to identify what is valid or not valid, similarly the child data is also updated to determine which feeling should be expressed and which should be left unspoken. In this way, the adult ego helps an individual to control his emotional expressions appropriately.

Child Ego: The child ego, refers to the state of an individual when he behaves illogically and takes quick actions to satisfy the immediate needs without thinking much about its consequences. The creativity, depression, conformity, dependence, hate, fear, etc. are some of the main characteristics of this ego state. The child ego represents the childhood state when an individual has not become social and is in its initial stage of development.

The child ego can be natural, adaptive and rebellious. The natural child is sensuous, impulsive, affectionate and does things that come naturally. Whereas the adaptive child is one, who is trained and instructed by parents to behave in a manner taught by them. The rebellious child is one who is not allowed to open up and experiences anger, fear and frustration.

Life Positions

Definition: The Life Positions refers to the specific behavior towards others that an individual learns on the basis of certain assumptions made very early in the life.

The life positions can be categorized as follows:

1. I am O.K., You are O.K.: This life position shows that an individual has several O.K. experiences with others. This means, an individual encountered no severe problems or issues with others in his childhood and had a normal relationship with them. People with such life positions about themselves and others around him can solve any problem very easily and realizes the significance of others being in his life. This position is based on the adult ego.
2. I am O.K., You are not O.K.: This life position is created when an individual was too much ignored when he was a child. Here, an individual believes that he is right, and all the others around him are wrong. These are the individual who possesses the rebellion child ego and put blame on others for anything that goes wrong with them.

3. I am not O.K., you are O.K.: This life position gets created when an individual feels that others do things better than him. He feels inferior to others and believes that others can do many things which he cannot do by himself. These kinds of people always complain about one thing or the other and remain highly dissatisfied with their lives.
4. I am not O.K., you are not O.K.: This kind of life position is created by those who lacks interest in living. They feel life is not worth living and are the ones who have been neglected by their parents in their childhood and were brought up by the servants. Such kind of people commits suicide or homicide to end their lives.

Thus, the life positions talk about the individual developing his identity, sense of worth and perception about others during his childhood and believing it to be true until and unless some major experience changes it.

Applications of TA

TA studies transactions amongst people and understands their interpersonal behaviour. It was developed by Eric Berne, a psychotherapist. He observed there are several ‘people’ inside each person who interact with other people in different ways.

1. Ego States,
2. Life Position and
3. Analysis of Transactions.

1. Ego States:

It represents a person’s way of thinking, feeling and behaving. There are three ego states present in everyone: child, parent and adult. They are related to behaviour of a person and not his age. However, they are present in every person in varying degrees. There may be more of one ego state than another at a specific point of time. When two persons communicate with each other, communication is affected by their ego states. These are;

(a) Child ego:

Child behaviour reflects a person’s response to communicate in the form of joy, sorrow, frustration or curiosity. These are the natural feelings that people learn as children. It reflects immediate action and immediate satisfaction. It reflects childhood experience of a person gained generally up to the age of five years.

A child can be:

(i) Natural child:

He is naturally curious, joyous or scornful. He does what comes his way naturally.

(ii) Adaptive child:

He reacts the way his parents want him to react. He is trained to act.

(iii)Rebellious child:

He has the experience of fear, frustration and anger.

(b) Parent Ego:

Parent behaviour is acquired through external environment. As young children, their parents' behaviour remains embedded in their minds which is reflected as parental ego when they grow up. It usually reflects protection, displeasure, reference to rules and working on the basis of past precedents.

This can be:

(i) Nurturing parent ego:

As nurturing parents, managers praise good performance of the workers. They interact with them and help them during times of distress. They reflect nurturing behaviour towards others.

(ii) Negative or critical parent ego:

As critical parents, managers criticize or ignore poor performance of the workers rather than help them to improve. They have a critical attitude while interacting with others.

(c) Adult ego:

Adult behaviour reflects the ability to analyse the situation and take logical decisions. He overcomes the emotional feelings and takes decisions based on facts and figures. This state is based upon reasoning, thinking, experience, rationality and discussion based on facts.

It updates the parental ego to determine what is right and wrong and child ego to determine what feelings to express and what not to express. These ego states are present in all human beings at some time or the other. People respond to different situations in different ways depending on their ego state.

Unit IV

Groups

A group can be defined as two or more interacting and interdependent individuals who come together to achieve particular objectives. A group behavior can be stated as a course of action a group takes as a family. For example: Strike.

Types of Groups

There are two types of groups an individual forms. They are formal groups and informal groups. Let us know about these two groups.

Formal Groups

These are the type of work groups created by the organization and have designated work assignments and rooted tasks. The behavior of such groups is directed toward achieving organizational goals.

These can be further classified into two sub-groups −

- Command group − It is a group consisting of individuals who report directly to the manager.
- Interest group − It is a group formed by individuals working together to achieve a specific objective. Example − A group of workers working on a project and reporting to the same manager is considered as a command group. A group of friends chilling out together is considered as interest group or say members of a club.

Informal Groups

These groups are formed with friendships and common interests. These can be further classified into two sub-groups −

- Task group − Those working together to finish a job or task is known as task group.
- Friendship group − Those brought together because of their shared interests or common characteristics is known as friendship group.

Why Do People Join Groups

There is no particular reason answering why individuals join groups. Group helps individual to feel stronger, have fewer self-doubts, and be more contrary to threats.

The following points help us understand the need of joining a group by individuals −

- Security mirrors strength in numbers. Status pinpoints a prestige that comes from belonging to a specific group. Inclusion in a group is considered as important because it provides recognition and status.
- Self-esteem transmits people's feelings of self-worth. Membership can sometimes raise feelings of self-esteem like being accepted into a highly valued group.

- Affiliation with groups can meet one's social needs. Work groups significantly contribute to meet the need for friendships and social relations.
- Groups represent power. What mostly cannot be achieved individually becomes possible with group effort. Power might be aimed to protect themselves from unreasonable demands. Informal groups provide options for individuals to practice power.
- People may join a group for goal achievement. Sometimes it takes more than one person to accomplish a particular task.

Group Roles

The concept of roles is applicable to all employees within an organization as well as to their life outside the organization. A role is a set of expected behavior patterns attributed to the one who occupies the position demanded by the social unit.

Individuals play multiple roles at the same time. Employees attempt to understand what kind of behavior is expected from them. An individual when presented by divergent role expectations experiences role conflict. Group roles are divided into three types −

- Task-oriented Roles
- Relationship-oriented Roles
- Individual Roles

Task-oriented Roles

Roles allotted to individuals according to their work and eligibility is known as task-oriented roles. Task-oriented roles can broadly divide individuals into six categories initiator, informer, clarifier, summarizer, reality tester and information seekers or providers respectively.

- Initiator − The one who proposes, suggests, defines.
- Informer − The one who offers facts, expresses feelings, gives opinions.
- Clarifier − The one who interprets, defines, clarifies everything.
- Summarizer − The one who links, restates, concludes, summarizes.
- Reality Tester − The one who provides critical analysis.
- Information seekers or providers − The one who gives information and data.

These roles present the work performed by different individuals according to their marked designation.

Relationship-oriented Roles

Roles that group individuals according to their efforts made to maintain healthy relationship in the group and achieve the goals are known as relationship-oriented roles. There are five categories of individuals in this category namely: harmonizer, gatekeeper, consensus tester, encourager, and compromiser.

- Harmonizers − The one who limits tension and reconciles disagreements.
- Gatekeeper − The one who ensures participation by all.
- Consensus Tester − The one who analyzes the decision-making process.
- Encourager − The one who is warm, responsive, active, shows acceptance.
- Compromiser − The one who admits error, limits conflict.

These roles depict the various roles an individual plays to maintain healthy self as well as group relationships.

Individual Roles

Roles that classify a person according to the measure of individual effort put in the project aimed is known as individual roles. Five types of individuals fall into these roles: aggressor, blocker, dominator, cavalier, and avoidance.

- Aggressor − The one who devalues others, attacks ideas.
- Blocker − The one who disagrees and rebels beyond reason.
- Dominator − The one who insists superiority to manipulate.
- Cavalier − The one who takes part in a group non-productively.
- Avoidance − The one who shows special interest to avoid task.

These are the various roles a person plays in an organization.

Well-Functioning Groups

We know what a group is, why it is important to form a group, and what the group-oriented roles are. Now we need to know how to mark a group as a well-functioning group, what features are necessary for a group to mark it as efficient.

A group is considered effective when it has the following characteristics.

- Relaxed, comfortable, friendly atmosphere.
- Task to be executed are well understood and accepted.
- Members listen well and actively participate in given assignments.
- Assignments are made clear and are accepted.
- Group is acquainted of its operation and function.
- People express their feelings and ideas openly.
- Consensus decision-making process is followed.
- Conflict and disagreement center regarding ideas or method.

Group Behavior – Example

Let us understand group behavior with the help of an example.

To work on a specific project, we make a group of 4 members: Rohit, Raj, Sid, and Rahul. It is not possible for any one of them to complete the project individually as it may be time consuming as well as not all the members as individuals have mastered the skills required to complete the project. This indicates the need to come together as a group.

Moving ahead, let us specify their roles. Rohit is the initiator, as he proposes the idea of the project, Raj collects all the information and resources required for the project and becomes the informer, Sid is the clarifier as he interprets the data and saves refined information, and Rahul is the summarizer as he concludes the result of the project that is what do we achieve by the end of our project. These are the task-oriented roles.

Group Development

Definition: The Group Development means, forming the association of people to work as a group and direct their actions towards the accomplishment of a common goal. The jobs of each group member are interdependent and hence the performance of one will affect the entire group's performance. It is often called as a team building or team development.

The most famous and a well-known model of group development was proposed by Bruce Tuckman, that included four stages Viz, forming, storming, norming, performing and later the fifth stage adjourning was added to this model.

Tuckman's Stages of Group Development.

1. Forming: At this stage, the formation of a new group begins, wherein the members come together and get to know each other through the interactions. Here the individuals are excited and anxious to know about the scope of the task and the ways to approach it. Generally, the individuals come with a desire to get accepted by others and avoid controversy or conflicts.
2. Storming: Once the forming stage is over, the individuals will start interacting with each other in the context of the task to be achieved. The conflict and competition among the group members will be highest at this stage.
3. Norming: Once the role of every member is cleared along with the authority and responsibility of each, the team members start settling in a group. Here, everybody works cohesively towards the target and appreciate each other's experience and skills.
4. Performing: At this stage, synergy gets created between the team members, where everyone works towards the accomplishment of a goal. This stage is characterized by flexibility and interdependence. The team members know each other so well that they can handle any complex problem that comes before the team.

Also, the roles and responsibilities of member changes according to the situation frequently, because at this stage everyone is equally a task-oriented and people-oriented and thus can perform efficiently.

5. Adjourning: This is the last stage of group development, where the group is terminated, and the group members are separated from each other. Every group is created for a purpose, and once the purpose is fulfilled the group is adjourned.

Some authors call this stage as “mourning or deforming,” because, the sense of loss is felt by the group members, at the time of separation from each other.

Thus, the researchers study about the group development to determine the changes that occur within the group.

THEORIES OF GROUP FORMATION

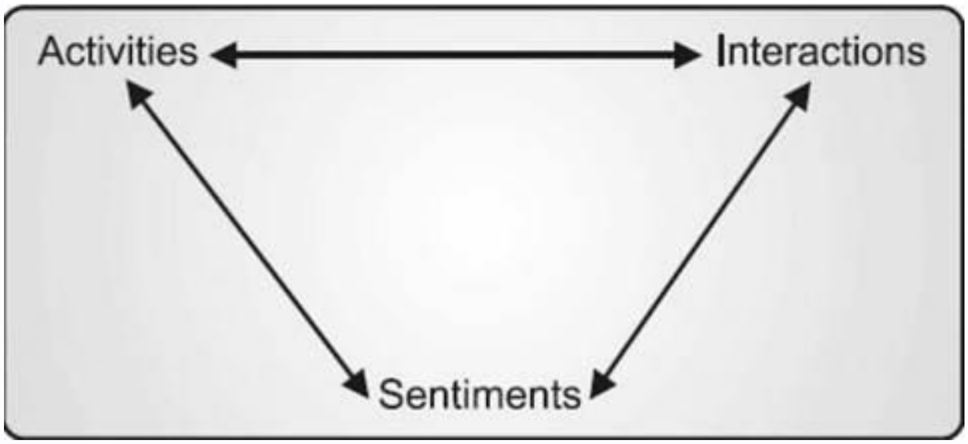
Homan's Theory

Homan's theory of group formation is based on three elements, namely, activities, interaction and sentiments. According to Homan, these three elements are directly related to each other. The required activities are the assigned tasks to people to work. The required interaction takes place when any person's activity takes place or is influenced by the activity of any other person. As regards sentiments, these are the feelings or attitudes of a person towards others, i.e., his likes or dislikes, approval or disapproval.

The key element is interaction because of which they develop common sentiments for one another

This theory explains the basic ideal behind forming groups. Scott observes that interaction is not only helps in attaining goals but also to solve problems, facilitate co-ordination, reduce tension and achieve a balance. Participants interacting in this manner tend to form powerful groups.

The following imaginary example will help understand Homan's theory in a better manner : Suppose, the students of Business Administration are in queue to deposit their examination fees in the State Bank of India. All of them have a common purpose, that is, to deposit fees when there turn comes. They see, a student from Mass Communication Department comes and breaks the queue to go ahead. The student whose turn was dislocated by the queue breaker tells the queue breaker not to do so and pushes him out. This influences all other students standing in the queue to follow the student whose turn was dislocated, in telling him not to break the queue. In this example, they can see activities and interaction taking place among the students. When the students actually disapprove the queue breaker in doing so, it reflects their sentiments towards each other. In sum and substances, all these activities took place because of the sentiments or feelings of the students/people.



Exchange Theory of Reward and Cost Outcome

Thaibaunt and Kelly put forward their theory of group formation, stating the outcome of interaction as the basis of group formation. According to them, the outcome of interaction should result in attraction and affiliation, also called rewards, among the persons of a group. In case, the interaction incurs anxiety or frustration or embarrassment or fatigue to the members of a group, it is, then, called cost for the members rather than a reward. In exchange theory, affiliation, interaction and common attitude play an important role.

Balance Theory

This theory was propounded by Newcomb. According to this theory, groups are formed on the basis of attractions of people towards each other having similar attitudes and values. Rao and Barman form relationship because of their common attitudes and values. They try to maintain a symmetrical relationship between the attraction and common attitudes and values. As and when, their relationship becomes unbalanced, both try to restore balance. However, if balance cannot be restored, then their relationship gets dissolved. Thus, one will appreciate that both attraction and interaction play a significant role in balance theory.

This theory also does not explain the full view of group formation as mere similarity of attitudes does not necessarily lead to group formation. Further, there are other reasons for group formation besides similarity of attitudes.

Team

Success in the workplace depends on your ability to build a team, as well as to interact with others on that team. Together, people are able to accomplish what one person alone can not. This is known as synergy.

Following are the characteristics of a Good/Effective team:

- A clear, elevating goal: This is a goal which has been communicated to all.
- A results-driven structure: The goal has been jointly decided by all the team members. They are fully committed towards achieving it.
- Competent members: Each team member has the required skill set in order to achieve the team objectives.
- Unified commitment: There is nothing happening in silos. With the total commitment from team members, achieving organizational goals becomes easier.
- A collaborative climate: Commitment from team members and a good leadership leads to a collaborative team with a productive work environment.
- Standards of excellence: Quality orientation is vital to the success of any organization.
- External support and recognition: Appreciation as well as appraisal is required to keep the morale of the team high.
- Principled leadership: Leadership defines a team. An able-bodied leadership can chart the team's path to success.
- Each team member participates actively and positively in meetings as well as projects. This shows a person's commitment as well as understanding towards a project.
- Team goals are clearly understood by all: Communication is vital for achieving successful completion of any project.
- Individual members have thought about creative solutions to the team's problem. Thinking out of the box is vital in today's economic scenario.
- Members are listened to carefully as well as given a thoughtful feedback. Listening is an important skill for any team. Each team member is important. The thoughts and ideas of each team member have to be listened to, with respect, no matter how silly they may sound at first.
- Everyone takes the initiative in order to get things done. There is no concept of passing the buck. This is an indication of clear communication leading to understanding of individual responsibilities.
- Each team member trusts the judgment of others: Mutual trust and respect is highly important for the team. This is the only way to achieve the organization goals.
- The team has to be willing to take risks: Risk taking is an attitude which comes with confidence. Confidence on yourself as well as on the team, besides the ability to face all consequences.

- Everyone has to be supportive of the project as well as of others. A team is one unit. Unless these cohesive forces are there, the team will never be able to work efficiently enough.
- There is ample communication between the team members.
- Team decisions are made by using organized as well as logical methods.
- Dissenting opinions are never ignored: In fact, they are always recorded in order to be revisited in case the future situations dictate so.
- Teams are given realistic deadlines: External support as well as aid is vital to the success of any team.

An efficient team needs support from both inside and outside. It needs to meet the individual needs of its members in order to achieve the organization's goals.

Organizational Culture

Every company has its own unique personality, just like people do. The unique personality of an organization is referred to as its culture.

In groups of people who work together, organizational culture is an invisible but powerful force that influences the behavior of the members of that group.

According to Richard Perrin, "Organizational culture is the sum of values and rituals which serve as glue to integrate the members of the organization."

According to Alan Adler, "Organizational culture is civilization in the workplace."

Characteristics of organizational culture are;

- Innovation (Risk Orientation).
- Attention to Detail (Precision Orientation).
- Emphasis on Outcome (Achievement Orientation).
- Emphasis on People (Fairness Orientation).
- Teamwork (Collaboration Orientation).
- Aggressiveness (Competitive Orientation).
- Stability (Rule Orientation).

Conflict Management definition

Conflict management refers to techniques and ideas designed to reduce the negative effects of conflict and enhance the positive outcomes for all parties involved.

The techniques and ideas used depend on the type of conflict that needs managing – researchers differentiate between affective (relational) and substantive (performance, process or task-specific) conflict, as well as interorganisational conflict (between two or more businesses) and intraorganisational (conflict within organisations).

Conflict resolution can be an aim of conflict management but not all conflict management techniques or styles have conflict resolution as the ultimate target as it may not be feasible.

Models of conflict management are very varied – theories that came out of the 1970s and 1980s focused on the intentions of the parties involved in conflict as the key to moving towards positive outcomes.

Khun and Poole (2000) classified approaches as either distributive or integrative, the former focused on distributing a fixed number of positive outcomes between warring parties, and the latter as focused on integrating the opposing needs of the parties to create the best outcome for all involved.

Rahim (2002) identified, among the literature, five common management approaches to conflict resolution: integrating, obliging, dominating, avoiding and compromising.

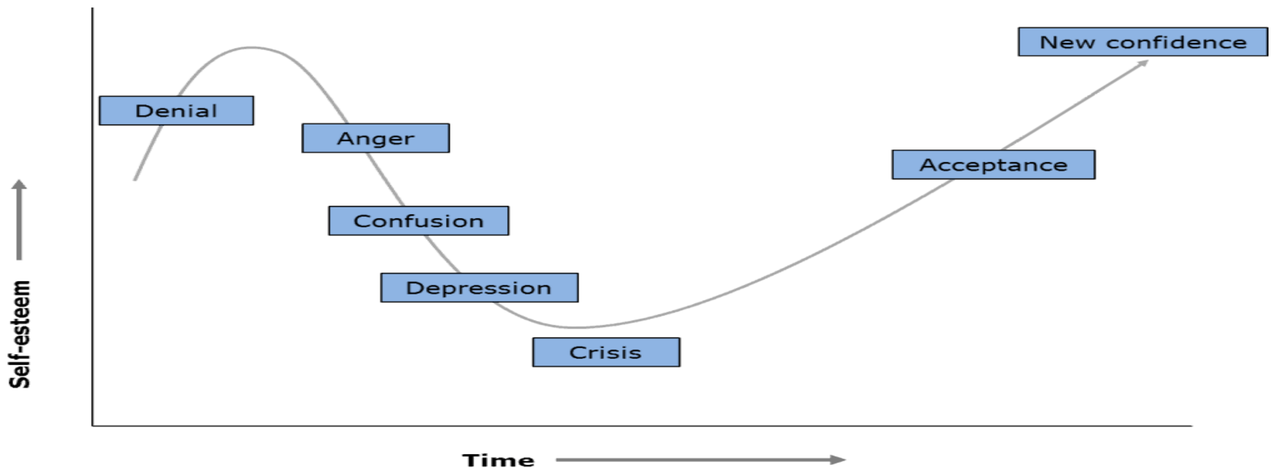
ORGANIZATIONAL CHANGE – REASONS WHY PEOPLE RESIST CHANGE

Expecting resistance to change and planning for it from the start of your change management progamme will allow you to effectively manage objections.

Understanding the most common reasons people object to change gives you the opportunity to plan your change strategy to address these factors.

It’s not possible to be aware of all sources of resistance to change. Expecting that there will be resistance to change and being prepared to manage it is a proactive step. Recognizing behaviors that indicate possible resistance will raise awareness of the need to address the concerns.

Classic psychological reactions to change



At the end of the day all sources of resistance to change need to be acknowledged and people's emotions validated.

It's far better to anticipate objections than to spend your time putting out fires, and knowing how to overcome resistance to change is a vital part of any change management plan.

TOP 12 TYPICAL REASONS FOR RESISTANCE TO CHANGE

1. Misunderstanding about the need for change/when the reason for the change is unclear — If staff do not understand the need for change you can expect resistance. Especially from those who strongly believe the current way of doing things works well…and has done for twenty years!
2. Fear of the unknown — One of the most common reasons for resistance is fear of the unknown. People will only take active steps toward the unknown if they genuinely believe – and perhaps more importantly, feel – that the risks of standing still are greater than those of moving forward in a new direction
3. Lack of competence — This is a fear people will seldom admit. But sometimes, change in organizations necessitates changes in skills, and some people will feel that they won't be able to make the transition very well
4. Connected to the old way — If you ask people in an organization to do things in a new way, as rational as that new way may seem to you, you will be setting yourself up against all that hard wiring, all those emotional connections to those who taught your audience the old way – and that's not trivial

5. Low trust — When people don't believe that they, or the company, can competently manage the change there is likely to be resistance
6. Temporary fad — When people belief that the change initiative is a temporary fad
7. Not being consulted — If people are allowed to be part of the change there is less resistance. People like to know what's going on, especially if their jobs may be affected. Informed employees tend to have higher levels of job satisfaction than uninformed employees
8. Poor communication — It'sself evident isn't it? When it comes to change management there's no such thing as too much communication
9. Changes to routines — When we talk about comfort zones we're really referring to routines. We love them. They make us secure. So there's bound to be resistance whenever change requires us to do things differently
10. Exhaustion/Saturation — Don't mistake compliance for acceptance. People who are overwhelmed by continuous change resign themselves to it and go along with the flow. You have them in body, but you do not have their hearts. Motivation is low
11. Change in the status quo — Resistance can also stem from perceptions of the change that people hold. For example, people who feel they'll be worse off at the end of the change are unlikely to give it their full support. Similarly, if people believe the change favours another group/department/person there may be (unspoken) anger and resentment
12. Benefits and rewards — When the benefits and rewards for making the change are not seen as adequate for the trouble involved

Expecting resistance to change and planning for it from the start of your change management progamme will allow you to effectively manage objections. Not dealing proactively is one pitfall – but there are many other common mistakes.

Cross cultural management

Cross cultural management involves managing work teams in ways that considers the differences in cultures, practices and preferences of consumers in a global or international business context. Many businesses have to learn to modify or adapt their approaches in order to compete on a level in fields no longer bound by physical geography with online interactions more common in business and other situations.

Functions include

- Recruiting candidates that can be effective in cross-cultural environments

- Handling differing regulatory environments for business
- Training employees to handle intercultural communication issues
- Facilitating cross-cultural teams
- Aligning HR policies and procedures across corporate entities in different nations

BUSINESS MATHEMATICS (103)

UNIT – I
Permutation and Combination

INTRODUCTION

Permutation and combination has lately emerged as an important topic for many entrance examinations. This is primary because questions from the topic require analytical skill and a logical bend of mind. Even students who do not have mathematics as a subject can handle them if they have a fairly good understanding of the concepts and their application. Hence anyone who is well-versed in different methods of counting and basic calculations will be able to solve these problems easily

IMPORTANT NOTATION

n! (Read as n factorial)

Product of first n positive integers is called n factorial

$n! = 1 \times 2 \times 3 \times 4 \times 5 \times \ldots n$

$n! = (n - l)!\ n \in N$

In special case $0! = 1$

MEANING OF PERMUTATION AND COMBINATION

Permutation

The arrangement made by taking some or all elements out of a number of things is called a permutation. The number of permutations of n things taking r at a time is denoted by ${}^nP_r$ and it is defined as under:

$${}^n p_r = \llcorner n / \llcorner (n-r)$$

Combination

The group or selection made by taking some or all elements out of a number of things is called a combination.

The number of combinations of n things taking r at a time is denoted by ${}^nC_r$ or and it is defined as under:

$${}^nC_r = \llcorner n / \llcorner r \llcorner\ (n - r)$$

Here $n!$ = Multiple of n natural number

Some Important Results of Permutations

1. ${}^nP_{n-1} = {}^nP_n$

2. ${}^nP_n = n!$
3. ${}^nP_r = n\,({}^{n-1}P_{r-1})$
4. ${}^nP_r = (n - r + 1) \times {}^nP_{r-1}$ 5. ${}^nP_r = {}^{n-1}P_r + r\,({}^{n-1}P_{r-1})$

Types of Permutations

When in a permutation of n things taken r at a time, a particular thing always occurs, then the required number of permutations = $r\,({}^{n-1}P_{r-1})$.

Q. If ${}^nC_{10} = {}^nC_{14}$ then find the value of n

Solution

${}^nC_{10} = {}^nC_{14} \Rightarrow n = (10 + 14) = 24$ $(\because n = p + q)$

Permutations with Repetition

These are the easiest to calculate.

When you have n things to choose from ... you have n choices each time!

When choosing r of them, the permutations are:

$$n \times n \times \ldots \textit{(r times)}$$

(In other words, there are n possibilities for the first choice, AND THEN there are n possibilities for the second choice, and so on, multiplying each time.)

Which is easier to write down using an exponent of r?

$$n \times n \times \ldots \text{(r times)} = n^r$$

Example: in the lock above, there are 10 numbers to choose from (0,1,..9) and you choose 3 of them:

$$10 \times 10 \times \ldots \text{(3 times)} = 10^3 = 1{,}000 \text{ permutations}$$

Permutations without Repetition

In this case, you have to reduce the number of available choices each time.

For example, what order could 16 pool balls be in?

After choosing, say, number "14" you can't choose it again.

So, your first choice would have 16 possibilities, and your next choice would then have 15 possibilities, then 14, 13, etc. And the total permutations would be:

$$16 \times 15 \times 14 \times 13 \times \ldots = 20{,}922{,}789{,}888{,}000$$

But maybe you don't want to choose them all, just 3 of them, so that would be only:

$$16 \times 15 \times 14 = 3{,}360$$

In other words, there are 3,360 different ways that 3 pool balls could be selected out of 16 balls.

But how do we write that mathematically? Answer: we use the "factorial function"

The factorial function (symbol :!) just means to multiply a series of descending natural numbers. Examples:

- $4! = 4 \times 3 \times 2 \times 1 = 24$
- $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5{,}040$
- $1! = 1$

There are also two types of combinations (remember the order does not matter now):

1. Repetition is Allowed: such as coins in your pocket (5,5,5,10,10)
2. No Repetition: such as lottery numbers (2,14,15,27,30,33)

1. Combinations with Repetition

Actually, these are the hardest to explain, so I will come back to this later.

2. Combinations without Repetition

This is how lotteries work. The numbers are drawn one at a time, and if you have the lucky numbers (no matter what order) you win!

The easiest way to explain it is to:

- assume that the order does matter (i.e. permutations),
- then alter it so the order does not matter.

Going back to our pool ball example, let us say that you just want to know which 3 pool balls were chosen, not the order.

We already know that 3 out of 16 gave us 3,360 permutations.

But many of those will be the same to us now, because we don't care what order!

For example, let us say balls 1, 2 and 3 were chosen. These are the possibilities:

Order does matter	Order doesn't matter
1 2 3	
1 3 2	
2 1 3	
2 3 1	1 2 3
3 1 2	
3 2 1	

So, the permutations will have 6 times as many possibilities.

In fact there is an easy way to work out how many ways "1 2 3" could be placed in order, and we have already talked about it. The answer is:

$$3! = 3 \times 2 \times 1 = 6$$

(Another example: 4 things can be placed in $4! = 4 \times 3 \times 2 \times 1 = 24$ different ways, try it for yourself!)

So, all we need to do is adjust our permutations formula to reduce it by how many ways the objects could be in order (because we aren't interested in the order any more):

$$\frac{n!}{(n-r)!} \times \frac{1}{r!} = \frac{n!}{r!(n-r)!}$$

That formula is so important it is often just written in big parentheses like this:

where *n* is the number of things to choose from, and you choose *r* of them

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

(No repetition, order doesn't matter)

It is often called "n choose r" (such as "16 choose 3")

And is also known as the "Binomial Coefficient"

Notation

As well as the "big parentheses", people also use these notations:

$$C(n,r) = {}^{n}C_{r} = {}_{n}C_{r} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example

So, our pool ball example (now without order) is:

16! 16! 20,922,789,888,000

= = = 560

3!(16-3)!   3!×13!   6×6,227,020,800

Or you could do it this way:

$$\frac{16\times15\times14}{3\times2\times1} = \frac{3360}{6} = 560$$

It is interesting to also note how this formula is nice and symmetrical:

In other words choosing 3 balls out of 16, or choosing 13 balls out of 16 have the same number of combinations.

$$\frac{n!}{r!(n-r)!} = \binom{n}{r} = \binom{n}{n-r}$$

$$\frac{16!}{3!\,(16-3)!} = \frac{16!}{13!\,(16-13)!} = \frac{16!}{3!\,\times 13!} = 560$$

Pascal's Triangle

You can also use Pascal's Triangle to find the values. Go down to row "n" (the top row is 0), and then along "r" places and the value there is your answer. Here is an extract showing row 16:

1   14   91   364 ...

1   15   105   455   1365 ...

1 16 120 560 1820 4368 ...

1. Combinations with Repetition

OK, now we can tackle this one ...

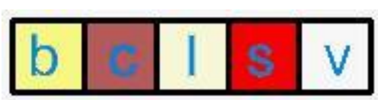
Let us say there are five flavors of ice-cream: banana, chocolate, lemon, strawberry and vanilla. You can have three scoops. How many variations will there be?

Let's use letters for the flavors: {b, c, l, s, v}. Example selections would be

- {c, c, c} (3 scoops of chocolate)
- {b, l, v} (one each of banana, lemon and vanilla)
- {b, v, v} (one of banana, two of vanilla)

(And just to be clear: There are n=5 things to choose from and you choose r=3 of them. Order does not matter, and you can repeat!)

Now, I can't describe directly to you how to calculate this, but I can show you a special technique that lets you work it out.



Think about the ice cream being in boxes, you could say "move past the first box, then take 3 scoops, then move along 3 more boxes to the end" and you will have 3 scoops of chocolate!

So, it is like you are ordering a robot to get your ice cream, but it doesn't change anything, you still get what you want.

Now you could write this down as →○○○→→→(arrow means move, circle means scoop).

In fact the three examples above would be written like this:

{c, c, c} (3 scoops of chocolate):

{b, l, v} (one each of banana, lemon and vanilla):

{b, v, v} (one of banana, two of vanilla):

OK, so instead of worrying about different flavors, we have a *simpler* problem to solve: "how many different ways can you arrange arrows and circles"

Notice that there are always 3 circles (3 scoops of ice cream) and 4 arrows (you need to move 4 times to go from the 1st to 5th container).

So (being general here) there are $r + (n-1)$ positions, and we want to choose r of them to have circles.

This is like saying "we have $r + (n-1)$ pool balls and want to choose r of them". In other words it is now like the pool balls problem, but with slightly changed numbers. And you would write it like this:

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

where n is the number of things to choose from, and you choose r of them (Repetition allowed, order doesn't matter)

Interestingly, we could have looked at the arrows instead of the circles, and we would have then been saying "we have $r + (n-1)$ positions and want to choose $(n-1)$ of them to have arrows", and the answer would be the same...

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1} = \frac{(n+r-1)!}{r!(n-1)!}$$

So, what about our example, what is the answer?

$$= \frac{(5+3-1)!}{3!\,(5-1)!} = \frac{7!}{3! \times 4!} = \frac{5040}{6\times 24} = 35$$

In Conclusion

Phew, that was a lot to absorb, so maybe you could read it again to be sure!

But knowing *how* these formulas work is only half the battle. Figuring out how to interpret a real world situation can be quite hard.

But at least now you know how to calculate all 4 variations of "Order does/does not matter" and "Repeats are/are not allowed".

UNIT-II

DETERMINANTS:

In linear algebra, the determinant is a value associated with a square matrix. It can be computed from the entries of the matrix by a specific arithmetic expression, while other ways to determine its value exist as well. The determinant provides important information about a matrix of coefficients of a system of linear equations, or about a matrix that corresponds to a linear transformation of a vector space. In the first case the system has a unique solution exactly when the determinant is nonzero; when the determinant is zero there are either no solutions or many solutions. In the second case the transformation has an inverse operation exactly when the determinant is nonzero. A geometric interpretation can be given to the value of the determinant of a square matrix with real entries: the absolute value of the determinant gives the scale factor by which area or volume (or a higher dimensional analogue) is multiplied under the associated linear transformation, while its sign indicates whether the transformation preserves orientation. Thus a 2×2 matrix with determinant -2, when applied to a region of the plane with finite area, will transform that region into one with twice the area, while reversing its orientation. The determinant of the matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

is written

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

and has the value

$$aei + bfg + cdh - ceg - bdi - afh.$$

Minor

An element, aij, to the value of the determinant of order n − 1, obtained by deleting the row i and the column j in the matrix is called a minor.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & \boxed{5} & 4 \\ 3 & 6 & 2 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 1 \\ 3 & 2 \end{vmatrix}$$

Cofactor

The cofactor of the element aij is its minor prefixing:

The + sign if i+j is even.

The − sign if i+j is odd.

$$\begin{vmatrix} 1 & 2 & 1 \\ \boxed{2} & 5 & 4 \\ 3 & 6 & 2 \end{vmatrix} \rightarrow - \begin{vmatrix} 2 & 1 \\ 6 & 2 \end{vmatrix}$$

Properties of Determinants:-

The determinant has many properties. Some basic properties of determinants are:

1. $\det(I_n) = 1$ Where I_n is the $n \times n$ identity matrix.
2. $\det(A^{\mathrm{T}}) = \det(A).$
3. $\det(A^{-1}) = \dfrac{1}{\det(A)}.$
4. For square matrices A and B of equal size, $\det(AB) = \det(A)\det(B).$
5. $\det(cA) = c^n \det(A)$ for an $n \times n$ matrix.
6. If A is a triangular matrix, i.e. $a_{i,j} = 0$ whenever $i > j$ or, alternatively, whenever $i < j$, then its determinant equals the product of the diagonal entries:

$$\det(A) = a_{1,1}a_{2,2}\cdots a_{n,n} = \prod_{i=1}^{n} a_{i,i}.$$

This can be deduced from some of the properties below, but it follows most easily directly from the Leibniz formula (or from the Laplace expansion), in which the identity permutation is the only one that gives a non-zero contribution.

A number of additional properties relate to the effects on the determinant of changing particular rows or columns:

7. Viewing an $n \times n$ matrix as being composed of n columns, the determinant is an n-linear function. This means that if one column of a matrix A is written as a sum $v + w$ of two column vectors, and all other columns are left unchanged, then the determinant of A is the sum determinants of the matrices obtained from A by replacing the column by v respectively by w (and a similar relation holds when writing a column as a scalar multiple of a column vector).
8. This n-linear function is an alternating form. This means that whenever two columns of a matrix are identical, or more generally some column can be expressed as a linear combination of the other columns (i.e. the columns of the matrix form a linearly dependent set), its determinant is 0.

MATRICES:

In mathematics, a matrix (plural matrices) is a rectangular array of numbers, symbols, or expressions, arranged in *rows* and *columns*.[1][2] The individual items in a matrix are called its *elements* or *entries*. An example of a matrix with 2 rows and 3 columns is

$$\begin{bmatrix} 1 & 9 & -13 \\ 20 & 5 & -6 \end{bmatrix}.$$

Types of Matrices

A matrix may be classified by types. It is possible for a matrix to belong to more than one type.

A row matrix is a matrix with only one row.

$E = \begin{pmatrix} 4 \end{pmatrix}$

E is a row matrix of order 1 × 1

$B = \begin{pmatrix} 9 & -2 & 5 \end{pmatrix}$

B is a row matrix of order 1 × 3

A column matrix is a matrix with only one column.

$C = \begin{pmatrix} 3 \end{pmatrix}$

C is a column matrix of order 1 × 1

$$D = \begin{pmatrix} -5 \\ 3 \end{pmatrix}$$

D is a column matrix of order 2 × 1

A column matrix of order 2 ×1 is also called a vector matrix.

A zero matrix or a null matrix is a matrix that has all its elements zero.

$$O = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

O is a zero matrix of order 2 × 3

A square matrix is a matrix with an equal number of rows and columns.

$$T = \begin{pmatrix} 6 & 3 \\ 0 & 4 \end{pmatrix}$$

T is a square matrix of order 2 × 2

$$V = \begin{pmatrix} 7 & 1 & 9 \\ 3 & 2 & 5 \\ 2 & 1 & 8 \end{pmatrix}$$

V is a square matrix of order 3 × 3

A diagonal matrix is a square matrix that has all its elements zero except for those in the diagonal from top left to bottom right; which is known as the leading diagonal of the matrix.

$$B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

B is a diagonal matrix

A unit matrix is a diagonal matrix whose elements in the diagonal are all ones.

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

P is a unit matrix.

Matrix Addition and subtraction:-

Two matrices can only be added or subtracted if they have the same size. Matrix addition and subtraction are done entry-wise, which means that each entry in *A*+*B* is the sum of the corresponding entries in *A* and *B*.

Here is an example of matrix addition

$$A = \begin{bmatrix} 7 & 5 & 3 \\ 4 & 0 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 3 & 2 \end{bmatrix}$$

$$A+B=\begin{bmatrix}7+1 & 5+1 & 3+1\\ 4-1 & 0+3 & 5+2\end{bmatrix}=\begin{bmatrix}8 & 6 & 4\\ 3 & 3 & 7\end{bmatrix}$$

And an example of subtraction

$$A=\begin{bmatrix}7 & 5 & 3\\ 4 & 0 & 5\end{bmatrix} \qquad B=\begin{bmatrix}1 & 1 & 1\\ -1 & 3 & 2\end{bmatrix}$$

$$A-B=\begin{bmatrix}7-1 & 5-1 & 3-1\\ 4+1 & 0-3 & 5-2\end{bmatrix}=\begin{bmatrix}6 & 4 & 2\\ 5 & -3 & 3\end{bmatrix}$$

Remember you can not add or subtract two matrices of different sizes.
The following rules apply to sums and scalar multiples of matrices. Let *A*, *B*, and *C* be matrices of the same size, and let *r* and *s* be scalars.

- $A + B = B + A$
- $(A + B) + C = A + (B + C)$
- $A + 0 = A$
- $r(A + B) = rA + rB$
- $(r + s)A = rA + sA$
- $r(sA) = (rs)A$

Multiplication

What is matrix multiplication? You can multiply two matrices if, and only if, the number of columns in the first matrix equals the number of rows in the second matrix.
Otherwise, the product of two matrices is undefined. The product matrix's dimensions are →(rows of first matrix) × (columns of the second matrix)

In above multiplication, the matrices can be multiplied since the number of columns in the 1st one, matrix A, equals the number of rows in the 2nd, matrix B. The Dimensions of the product matrix. Rows of 1st matrix × Columns of 2nd 4 × 3.

If

$$\mathbf{A}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}, \quad \mathbf{B}=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix},$$

their matrix products are:

$$\mathbf{AB}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}=\begin{pmatrix}a\alpha+b\gamma & a\beta+b\delta\\ c\alpha+d\gamma & c\beta+d\delta\end{pmatrix},$$ and

$$\mathbf{BA}=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix}=\begin{pmatrix}\alpha a+\beta c & \alpha b+\beta d\\ \gamma a+\delta c & \gamma b+\delta d\end{pmatrix}.$$

Adjoint of Matrix: - The classical Adjoint of a square matrix A the transpose of the matrix who (i, j) entry is a i j cofactor.

(Adjoint of a Matrix) *Let* A *be an* $n \times n$ *matrix. The matrix* $B = [b_{ij}]$ *with* $b_{ij} = C_{ji}$, *for* $1 \leq i, j \leq n$ *is called the Adjoint of* A, *denoted* $Adj(A)$.

EXAMPLE *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$. *Then* $Adj(A) = \begin{bmatrix} 4 & 2 & -7 \\ -3 & -1 & 5 \\ 1 & 0 & -1 \end{bmatrix}$;

as $C_{11} = (-1)^{1+1}A_{11} = 4, C_{12} = (-1)^{1+2}A_{12} = -3, C_{13} = (-1)^{1+3}A_{13} = 1,$ *and so on.*

Inverse of a Matrix

Definition and Examples

Recall that functions f and g are inverses if

$$f(g(x)) = g(f(x)) = x$$

We will see later that matrices can be considered as functions from R^n to R^m and that matrix multiplication is composition of these functions. With this knowledge, we have the following:

Let A and B be n x n matrices then A and B are *inverses* of each other, then

$$AB = BA = I_n$$

Example

Consider the matrices

$$A = \begin{pmatrix} 2 & 0 & -1 \\ -3 & 0 & 2 \\ -2 & -1 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 1 & 0 \\ -4 & -2 & -1 \\ 3 & 2 & 0 \end{pmatrix}$$

We can check that when we multiply A and B in either order we get the identity matrix. (Check this.)

Not all square matrices have inverses. If a matrix has an inverse, we call it *nonsingular* or *invertible*. Otherwise it is called *singular*. We will see in the next section how to determine if a matrix is singular or nonsingular.

Determinants and Cramer's Rule

The 2 X 2 system

$ax + by = e;$

$cx + dy = f;$

has a unique solution provided $\Delta = ad - bc$ is nonzero, in which case the

Solution is given by

$x = (de – bf)/ (ad - bc)$; $y = (af – ce)/(ad – bc)$

This result, called Cramer's Rule for 2 X2 systems.\

Cofactor Expansion

The special subject of cofactor expansions is used to justify Cramer's rule and to provide an alternative method for computation of determinants. There is no claim that cofactor expansion is e_cient, only that it is possible, and di_erent than Sarrus' rule or the use of the four properties.

The Cayley-Hamilton Theorem:-

Presented here is an adjoint formula $F^{-1} = adj(F)= det(F)$ derivation

for the celebrated Cayley-Hamilton formula

$(-A)^{-n} + p_{n-1}(- A)^{n-1} + \_ \_ \_ + p0I = 0.$

The nxn matrix A is given and I is the identity matrix. The coe_cients pk in above are determined by the characteristic polynomial of matrix A, which is de_ned by the determinant expansion formula

Det (A).

Dependence of Vectors:-

A subset S of a vector space V is called *linearly dependent* if there exist a finite number of distinct vectors $u_1, u_2,..., u_n$ in S and scalars $a_1, a_2,..., a_n$, not all zero, such that

Note that the zero on the right is the zero vectors, not the number zero.

For any vectors $u_1, u_2,..., u_n$ we have that

This is called the trivial representation of *0* as a linear combination of $u_1, u_2,..., u_n$, this motivates a very simple definition of both linear independence and linear dependence, for a set to be linearly dependent, there must exist a non-trivial representation of *0* as a linear combination of vectors in the set.

A subset S of a vector space V is then said to be *linearly independent* if it is not linearly dependent, in other words, a set is linearly independent if the only representation of *0* as a linear combination its vectors are trivial representations.[1]
Note that in both definitions we also say that the vectors in the subset S are linearly dependent or linearly independent.

More generally, let V be a vector space over a field K, and let $\{v_i \mid i \in I\}$ be a family of elements of V. The family is *linearly dependent* over *K* if there exists a family $\{a_j \mid j \in J\}$ of elements of *K*, not all zero, such that where the index set *J* is a nonempty, finite subset of *I*.

Lecture 4: Concept of Limit

Definition: We say that the limit of $f(x)$ is L as x approaches a and write this as

$\lim_{x \to a} f(x) = L$.

An alternative notation that we will occasionally use in denoting limits is $f(x) \to L$ as $x \to a$ without actually letting $x = a$.

This means that the definition says that as x gets closer and closer to $x = a$ from both sides of course then $f(x)$ must be getting closer and closer to L or, as we move in towards $x = a$ then $f(x)$ must be moving in towards L .

Definition: Right-handed limit is denoted by $\lim_{x \to a+} f(x) = L$ and left-handed limit is denoted by $\lim_{x \to a-} f(x) = L$.

Given a function $f(x)$ if, $\lim_{x \to a+} f(x) = L = \lim_{x \to a-} f(x)$ then the limit will exist and $\lim_{x \to a} f(x) = L$

Likewise, if $\lim_{x \to a} f(x) = L$ then, $\lim_{x \to a+} f(x) = L = \lim_{x \to a-} f(x)$

If $\lim_{x \to a+} f(x) \neq \lim_{x \to a-} f(x)$ then the limit does not exist.

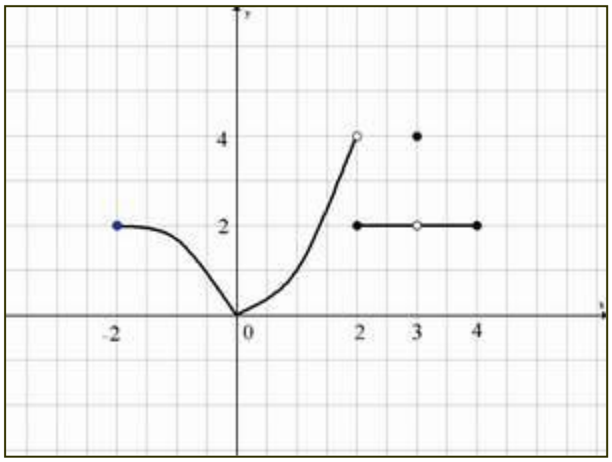
Example 1: Given the following graph, compute each of the following.

1. $f(-2)$ 2. $\lim_{x \to -2+} f(x)$ 3. $f(0)$ 4. $\lim_{x \to 0+} f(x)$ 5. $\lim_{x \to 0-} f(x)$

6. $\lim_{x \to 0} f(x)$ 7. $f(2)$ 8. $\lim_{x \to 2-} f(x)$ 9. $\lim_{x \to 2+} f(x)$ 10. $\lim_{x \to 2} f(x) = 2$

11. $f(-3)$ 12. $\lim_{x \to 3-} f(x)$ 13. $\lim_{x \to 3+} f(x)$ 14. $\lim_{x \to 3} f(x) = 2$ 15. $f(4)$

16. $\lim_{x \to 4-} f(x)$

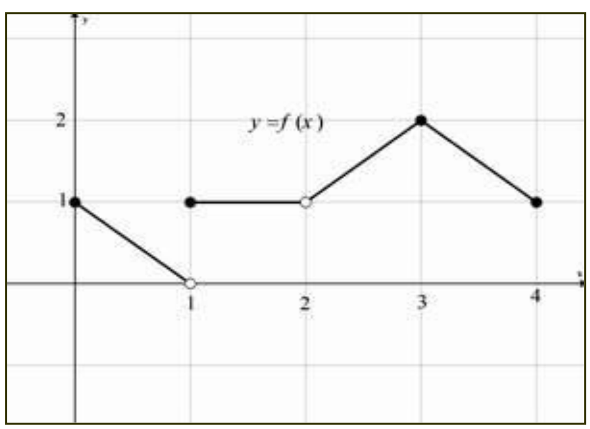


Exercise 1: Given the following graph, compute each of the following.

1. $f(0)$ 2. $\lim_{x \to 0+} f(x)$ 3. $f(1)$ 4. $\lim_{x \to 1+} f(x)$ 5. $\lim_{x \to 1-} f(x)$ 6. $\lim_{x \to 1} f(x)$

7. $f(2)$ 8. $\lim_{x \to 2-} f(x)$ 9. $\lim_{x \to 2+} f(x)$ 10. $\lim_{x \to 2} f(x) = 2$ 11. $f(3)$ 12. $\lim_{x \to 3-} f(x)$

13. $\lim_{x \to 3+} f(x)$ 14. $\lim_{x \to 3} f(x)$ 15. $f(4)$ 16. $\lim_{x \to 4-} f(x)$



Limit Properties

If $\lim_{x \to c} f(x) = L_1$ and $\lim_{x \to c} g(x) = L_2$,then

[Type text]

1. $\lim\limits_{x\to c}(f(x)\pm g(x))=L_1\pm L_2$.

2. $\lim\limits_{x\to c}(f(x).g(x))=L_1.L_2$.

3. $\lim\limits_{x\to c}\dfrac{f(x)}{g(x)}=\dfrac{L_1}{L_2}, L_2\neq 0$.

4. $\lim\limits_{x\to c}(af(x))=aL_1,\ a$ constant.

5. $\lim\limits_{x\to c}(f(x))^n=(\lim\limits_{x\to c}f(x))^n=L_1{}^n,\ n\in N$.

6. $\lim\limits_{x\to c}\sqrt[n]{f(x)}=\sqrt[n]{\lim\limits_{x\to c}f(x)}=\sqrt[n]{L_1},\ n\in N$, and for n even, we assume that $L_1>0$.

7. $\lim\limits_{x\to c}x=c$.

8. $\lim\limits_{x\to c}x^n=c^n$

Fact: If $p_n(x)$ is a polynomial of degree n, then $\lim\limits_{x\to c}p_n(x)=p_n(c)$.

Example 2 $\lim\limits_{x\to 3}\dfrac{x^3-2x^2}{x^2+2}=\dfrac{\lim\limits_{x\to 3}\left(x^3-2x^2\right)}{\lim\limits_{x\to 3}\left(x^2+2\right)}=\dfrac{(3)^3-2(3)^2}{(3)^2+2}=\dfrac{9}{11}$

Example 3:

$$\lim_{x\to -3}\sqrt{\frac{x^2+2x+1}{8+2x}}=\sqrt{\lim_{x\to -3}\left(\frac{x^2+2x+1}{8+2x}\right)}=\sqrt{\frac{\lim\limits_{x\to -3}(x^2+2x+1)}{\lim\limits_{x\to -3}(8+2x)}}=\sqrt{\frac{9-6+1}{8-6}}=\sqrt{\frac{4}{2}}=\sqrt{2}$$

Example 4: Given the function, $f(x)=\begin{Bmatrix}x^3+2x+1, & x<1\\ 3x-1, & x\geq 1\end{Bmatrix}$. Compute the following limits.

1. $\lim\limits_{x\to 2}f(x)$ 2. $\lim\limits_{x\to 1}f(x)$

Lecture 5: Computing Limits

Remark: Avoid common mistakes of the form $\dfrac{0}{0}$. Typically zero in the denominator means it's undefined. However that will only be true if the numerator isn't also zero. Also, zero in the numerator usually means that the fraction is zero, unless the denominator is also zero.

So, there are three cases to compute $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$.

1. $g(a)\neq 0$. In this case $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}=\dfrac{f(a)}{g(a)}$.

2. $g(a)=0$ and $f(a)\neq 0$. In this case $\lim_{x\to a}\frac{f(x)}{g(x)}$ does not exist.

3. $g(a)=0$ and $f(a)=0$. In this case $\lim_{x\to a}\frac{f(x)}{g(x)}$ can be calculated by using algebraic manipulation .

Case 1: $g(a)\neq 0$

Example 1: Evaluate the following limit.

(1) $\lim_{x\to 2}\frac{x^3+x+1}{x^2+2}$ (2) $\lim_{x\to 1}\frac{x^2-1}{x^2+1}$

Solution: (1) $\lim_{x\to 2}\frac{x^3+x+1}{x^2+2}=\frac{(2)^3+2+1}{(2)^2+2}=\frac{11}{6}$ (2) $\lim_{x\to 1}\frac{x^2-1}{x^2+1}=\frac{0}{2}=0$

Case 2: $g(a)=0$ and $f(a)\neq 0$ [Limits that equal infinity]

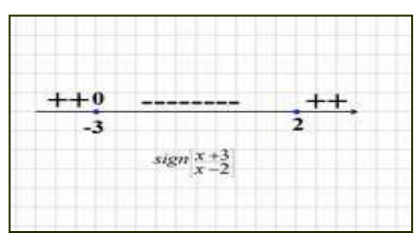
Definition

We say that $\lim_{x\to a}f(x)=\infty$ if we can make $f(x)$ arbitrarily large for all x sufficiently close to $x=a$, from both sides, without actually letting $x=a$. We say that $\lim_{x\to a}f(x)=-\infty$ if we can make $f(x)$ arbitrarily large and negative for all x sufficiently close to $x=a$, from both sides, without actually letting $x=a$.

Remark: Concider the limit $\lim_{x\to a}\frac{f(x)}{g(x)}$

1. If $\lim_{x\to a+}f(x)=\infty$ and $\lim_{x\to a-}f(x)=-\infty$, then the limit doesn't exist.
2. If $\lim_{x\to a+}f(x)=-\infty$ and $\lim_{x\to a-}f(x)=\infty$, then the limit doesn't exist.
3. If $\lim_{x\to a+}f(x)=\infty$ and $\lim_{x\to a-}f(x)=\infty$, then the limit doesn't exist and $\lim_{x\to a}f(x)=\infty$.
4. If $\lim_{x\to a+}f(x)=-\infty$ and $\lim_{x\to a-}f(x)=-\infty$, then the limit doesn't exist and $\lim_{x\to a}f(x)=-\infty$.

Example 2: Evaluate $\lim_{x\to 2}\frac{x+3}{x-2}$



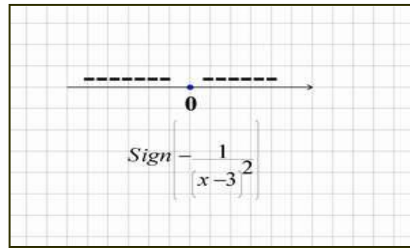
Solution:

$\lim\limits_{x\to 2-}\dfrac{x+3}{x-2}=-\infty$ and $\lim\limits_{x\to 2+}\dfrac{x+3}{x-2}=\infty$. So $\lim\limits_{x\to 2}\dfrac{x+3}{x-2}$ doesn't exist .

Example 3: Evaluate $\lim\limits_{x\to 0}\dfrac{1}{x^2}$

Solution: $\lim\limits_{x\to 0+}\dfrac{1}{x^2}=\infty$ and $\lim\limits_{x\to 0-}\dfrac{1}{x^2}=\infty$. So $\lim\limits_{x\to 0}\dfrac{1}{x^2}=\infty$.

Exercise 1: (1) Evaluate $\lim\limits_{x\to 2-}\dfrac{-x}{\sqrt{4-x^2}}$ (2) $\lim\limits_{x\to 3}\dfrac{-1}{(x-3)^2}$

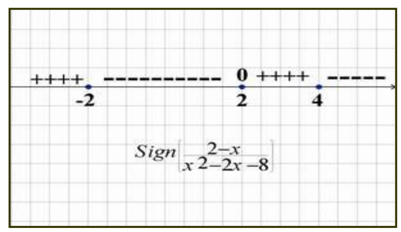


(1) $\lim\limits_{x\to 3+}\dfrac{-1}{(x-3)^2}=-\infty$ and $\lim\limits_{x\to 3-}\dfrac{-1}{(x-3)^2}=-\infty$. So $\lim\limits_{x\to 3}\dfrac{-1}{(x-3)^2}=-\infty$.

(2) $\lim\limits_{x\to 2-}\dfrac{-x}{\sqrt{4-x^2}}=-\infty$

Example 4: Evaluate the following limits.

$\lim\limits_{x\to 4}\dfrac{2-x}{x^2-2x-8}$



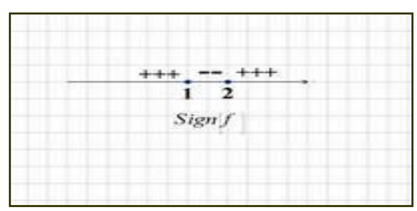
Solution: $\lim\limits_{x\to 4+}\dfrac{2-x}{x^2-2x-8}=-\infty$ $\lim\limits_{x\to 4-}\dfrac{2-x}{x^2-2x-8}=\infty$

So $\lim\limits_{x\to 4}\dfrac{2-x}{x^2-2x-8}$ doesn't exist .

Definition: A line $x=a$ is called a vertical asymptote of the graph of f if either

$\lim\limits_{x\to a+}f(x)=\pm\infty$ or $\lim\limits_{x\to a-}f(x)=\pm\infty$

Example 5: Find the vertical asymptotes of $f(x)=\dfrac{x-1}{x-2}$.

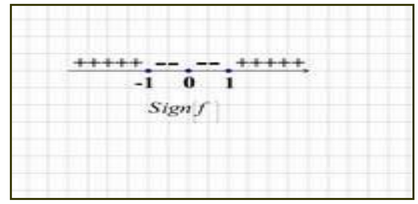


Solution: By analyze the sign of f is in the figure, notice that

$\lim\limits_{x\to 2+}f(x)=\infty$ and $\lim\limits_{x\to 2-}f(x)=-\infty$

So the vertical asymptote is the line $x=2$.

Exercise 2: Find the vertical asymptotes of



(1) $f(x)=\dfrac{x^2}{x^2-1}$ (2) $f(x)=\begin{Bmatrix} \dfrac{x+1}{x-2}, x<0 \\ x^3-3x-1, x\geq 0 \end{Bmatrix}$

Solution: Notice that

(1) $\lim_{x\to 1+} f(x)=\infty$ and $\lim_{x\to 1-} f(x)=-\infty$

$\lim_{x\to -1+} f(x)=-\infty$ and $\lim_{x\to -1-} f(x)=\infty$

So the vertical asymptotes are the lines $x=1$ and $x=-1$.

---

(2) Notice that f has no vertical asymptote because f is defined for all x.

Example 6: Find the vertical asymptotes of (1) $f(x)=xe^{-x}$ (2) $f(x)=\dfrac{x}{\sqrt{4+x^2}}$

(3) $f(x)=4\tan^{-1}x-1$

Solution: f has no vertical asymptote because f is defined for all x.

Exercise 3: Find the vertical asymptotes of (1) $f(x)=3e^{-1/x}$. (2) $f(x)=6x^{1/3}+3x^{4/3}$

Solution: (1) $\lim_{x\to 0+} 3e^{-1/x}=0$ $\lim_{x\to 0-} 3e^{-1/x}=\infty$. So the vertical asymptote is $x=0$ as $x\to 0-$.

(2) Notice that f has no vertical asymptote because f is defined for all x.

Exercise 4: Find the vertical asymptotes of $f(x)=\begin{Bmatrix} \dfrac{4x}{x-4}, x<0 \\ \dfrac{x^2}{x-2}, 0\leq x<4 \\ \dfrac{e^{-x}}{x+1}, x\geq 4 \end{Bmatrix}$.

Solution: $\lim_{x\to 2-} f(x)=-\infty$ and $\lim_{x\to 2+} f(x)=\infty$. So the vertical asymptote is the line $x=2$.

Case 3: $g(a)=0$ and $f(a)=0$

Example 7: Evaluate $\lim\limits_{x\to -2}\dfrac{x^3+8}{x^2-4}$?

Solution: take $x=-2$ we get, $\lim\limits_{x\to -2}\dfrac{x^3+8}{x^2-4}=\left(\dfrac{0}{0}\right)$

$$\lim_{x\to -2}\frac{x^3+8}{x^2-4}=\lim_{x\to -2}\frac{(x+2)(x^2-2x+4)}{(x+2)(x-2)}=\lim_{x\to -2}\frac{(x^2-2x-4)}{(x-2)}$$

$$\lim_{x\to -2}\frac{x^3+8}{x^2-4}=\lim_{x\to -2}\frac{(x^2-2x-4)}{(x-2)}=\frac{(-2)^2-2(-2)-4}{-2-2}=\frac{4+4-4}{-4}=-1$$

Example 8: Evaluate $\lim\limits_{x\to 4}\dfrac{x+\sqrt{x}-6}{\sqrt{x}-2}$.

Solution

$$\lim_{x\to 4}\frac{x+\sqrt{x}-6}{\sqrt{x}-2}\left(\frac{0}{0}\right)=\lim_{x\to 4}\frac{(\sqrt{x}+3)(\sqrt{x}-2)}{(\sqrt{x}-2)}=\lim_{x\to 4}(\sqrt{x}+3)=\sqrt{4}+3=2+3=5$$

Exercise 5: Evaluate the following limit. (1) $\lim\limits_{x\to 3}\dfrac{x^3+4x-39}{x^3-27}$ (2) $\lim\limits_{x\to 3}\dfrac{x^2-x-6}{x^2-2x-3}$

(3) $\lim\limits_{x\to 1}\dfrac{x^{\frac{3}{2}}-x}{x^{\frac{1}{2}}-1}$ (4) $\lim\limits_{x\to 0}\dfrac{(x+2)^2-4}{x}$?

Solution: (1)

$$\lim_{x\to 3}\frac{x^3+4x-39}{x^3-27}\left(\frac{0}{0}\right)=\lim_{x\to 3}\frac{(x-3)(x^2+3x+13)}{(x-3)(x^2+3x+9)}=\lim_{x\to 3}\frac{(x^2+3x+13)}{(x^2+3x+9)}$$

$$=\frac{(3)^2+3(3)+13}{(3)^2+3(3)+9}=\frac{9+9+13}{9+9+9}=\frac{31}{27}$$

(2) $\lim\limits_{x\to 3}\dfrac{x^2-x-6}{x^2-2x-3}\left(\dfrac{0}{0}\right)=\lim\limits_{x\to 3}\dfrac{(x-3)(x+2)}{(x-3)(x+1)}=\lim\limits_{x\to 3}\dfrac{(x+2)}{(x+1)}=\dfrac{3+2}{3+1}=\dfrac{5}{4}$

(3) $\lim\limits_{x\to 1}\dfrac{x^{\frac{3}{2}}-x}{x^{\frac{1}{2}}-1}\left(\dfrac{0}{0}\right)=\lim\limits_{x\to 1}\dfrac{x\,(x^{\frac{1}{2}}-1)}{x^{\frac{1}{2}}-1}=\lim\limits_{x\to 1}x=1$

(4) $=\lim\limits_{x\to 0}\dfrac{x\,(x+4)}{x}=\lim\limits_{x\to 0}(x+4)=4$ $\lim\limits_{x\to 0}\dfrac{(x+2)^2-4}{x}=\lim\limits_{x\to 0}\dfrac{x^2+4x+4-4}{x}=\lim\limits_{x\to 0}\dfrac{x^2+4x}{x}$

Example 8: Evaluate the following limit.

$$\lim_{h\to 0}\frac{(h+1)^2-1}{h}$$

Solution

$$\lim_{h\to 0}\frac{(h+1)^2-1}{h}\left(\frac{0}{0}\right)=\lim_{h\to 0}\frac{h^2+2h+1-1}{h}=\lim_{h\to 0}\frac{h(h+2)}{h}=\lim_{h\to 0}(h+2)=2$$

Example 9: Evaluate the following limit.

$$\lim_{x\to 3}\frac{\sqrt{x+6}-3}{x-3}$$

Solution: $$\lim_{x\to 3}\frac{\sqrt{x+6}-3}{x-3}=\lim_{x\to 3}\frac{\left(\sqrt{x+6}-3\right)\left(\sqrt{x+6}+3\right)}{(x-3)\left(\sqrt{x+6}+3\right)}$$

$$=\lim_{x\to 3}\frac{(x+6)-(9)}{(x-3)\left(\sqrt{x+6}+3\right)}=\lim_{x\to 3}\frac{(x-3)}{(x-3)\left(\sqrt{x+6}+3\right)}=\lim_{x\to 3}\frac{1}{\left(\sqrt{x+6}+3\right)}=\frac{1}{6}$$

Exercise 6: Evaluate the following limit.

(1) $\lim_{x\to 1}\frac{2-\sqrt{x+3}}{x^2+2x-3}$ (2) $\lim_{x\to 2}\frac{\sqrt{5x-1}-3}{8-2x^2}$ (3) $\lim_{x\to 3}\frac{x^2-3x}{x-\sqrt{x+1}-1}$

(4) $\lim_{x\to 4}\frac{3-\sqrt{2x+1}}{\sqrt{x}-2}$ (5) $\lim_{x\to 0}\frac{\sqrt[3]{1+2x}-1}{x}$ (6) $\lim_{x\to 1}\frac{\sqrt[3]{x}-1}{\sqrt[4]{x}-1}$

(7) $\lim_{x\to 1}\frac{(x+1)^3\sqrt{x}-8}{x-1}$ (8) $\lim_{x\to 1}\frac{\frac{5}{2x-3}+5}{4x^2-4}$ (9) $\lim_{x\to 1+}\frac{\sqrt{x^2-1}+\sqrt{x-1}}{\sqrt{x-1}}$

Example 36: Given the function,

$$f(x)=\begin{cases}\dfrac{x^3-1/x^3}{x-1/x}, & x\neq 1\\ 3, & x=1\end{cases}$$

Compute the following limit.

$\lim_{x \to 1} f(x)$

Solution

In doing limits recall that we must always look at what's happening on both sides of the point in question as we move in towards it. As x approaches 1 from the left and from the right is inside the first interval for the function and so there are values of x on both sides of $x = 1$ inside this interval. This means that we can just use the fact to evaluate this limit.

$$\lim_{x \to 1} f(x) = \lim_{x \to 1} \frac{x^3 - 1/x^3}{x - 1/x}\left(\frac{0}{0}\right) = \lim_{x \to 1} \frac{(x - \frac{1}{x})(x^2 + x(\frac{1}{x}) + \frac{1}{x^2})}{(x - \frac{1}{x})} = \lim_{x \to 1}(x^2 + 1 + \frac{1}{x^2}) = 1 + 1 + 1 = 3$$

Example 37: Evaluate the following limit.

$$\lim_{x \to 0} \frac{x^2 + x}{\sqrt{x^4 + 2x^2}}$$

Solution

$$\lim_{x \to 0} \frac{x^2 + x}{\sqrt{x^4 + 2x^2}}\left(\frac{0}{0}\right) = \lim_{x \to 0} \frac{x^2 + x}{\sqrt{x^2\left(x^2 + 2\right)}} = \lim_{x \to 0} \frac{x^2 + x}{\sqrt{x^2}\sqrt{x^2 + 2}} = \lim_{x \to 0} \frac{x^2 + x}{|x|\sqrt{x^2 + 2}}$$

Recall that we must always look at what's happening on both sides of the point in question.

$$\lim_{x \to 0+} \frac{x^2 + x}{|x|\sqrt{x^2 + 2}} = \lim_{x \to 0+} \frac{x(x+1)}{x\sqrt{x^2 + 2}} = \lim_{x \to 0+} \frac{(x+1)}{\sqrt{x^2 + 2}} = \frac{1}{\sqrt{2}}$$

$$\lim_{x \to 0-} \frac{x^2 + x}{|x|\sqrt{x^2 + 2}} = \lim_{x \to 0-} \frac{x(x+1)}{-x\sqrt{x^2 + 2}} = \lim_{x \to 0+} -\frac{(x+1)}{\sqrt{x^2 + 2}} = \frac{-1}{\sqrt{2}}$$

Therefore, $\lim_{x \to 0} \frac{x^2 + x}{\sqrt{x^4 + 2x^2}}$ does not exist.

Example 38: Evaluate the following limit. $\lim_{x \to 0} \frac{2x - |x|}{|3x| - 2x}$

Solution

$$\lim_{x \to 0+} \frac{2x - |x|}{|3x| - 2x} = \lim_{x \to 0+} \frac{2x - x}{3x - 2x} = \lim_{x \to 0+} \frac{x}{x} = \lim_{x \to 0+} 1 = 1$$

$$\lim_{x \to 0-} \frac{2x - |x|}{|3x| - 2x} = \lim_{x \to 0-} \frac{2x - \left(-x\right)}{\left(-3x\right) - 2x} = \lim_{x \to 0-} -\frac{3x}{5x} = \lim_{x \to 0-} -\frac{3}{5} = -\frac{3}{5}$$

We can see that,

$$\lim_{x\to 0+}\frac{2x-|x|}{|3x|-2x}\neq \lim_{x\to 0-}\frac{2x-|x|}{|3x|-2x}. \qquad \text{Therefore, } \lim_{x\to 0}\frac{2x-|x|}{|3x|-2x} \text{ doesn't exist.}$$

Lecture 6: Limits at infinity

Definition:

By limits at infinity we mean one of the following two limits.

$$\lim_{x\to\infty} f(x) \qquad\qquad \lim_{x\to-\infty} f(x)$$

Theorem: For $n>0$ we have

$$\lim_{x\to\infty}\frac{1}{x^n}=0 \qquad\qquad \lim_{x\to-\infty}\frac{1}{x^n}=0$$

This fact should make sense if you think about it. We require $n>0$ to make sure the term stays in the denominator and as we increase x then x^n will also increase. So, what we end up with is a constant divided by an increasingly large number so the quotient of the two will become increasingly small. In the limit we will get zero.

Theorem:

$$\lim_{x\to+\infty}(a_n x^n + a_{n-1}x^{n-1} + \dots\dots\dots + a_1 x + a_0) = \lim_{x\to+\infty} a_n x^n, a_n \neq 0$$
$$\lim_{x\to-\infty}(a_n x^n + a_{n-1}x^{n-1} + \dots\dots\dots + a_1 x + a_0) = \lim_{x\to-\infty} a_n x^n, a_n \neq 0$$

Remark: You can avoid common mistakes by giving careful consideration to the forms $\frac{\infty}{\infty}$ and $\infty-\infty$ during the computations of the limit. Initially, many students incorrectly conclude that $\frac{\infty}{\infty}$ is equal to 1, or that the limit does not exist, or is $+\infty$ or $-\infty$. Many also conclude that $\infty-\infty$ is equal to 0. In fact, the forms $\frac{\infty}{\infty}$ and $\infty-\infty$ are examples of indeterminant forms. This simply means that you have not yet determind an answer. Usually, these indeterminate forms can be circumvented by using algebraic manipulation. Such tools as algebraic simplification and conjugates can easily be used to circumvent the forms $\frac{\infty}{\infty}$ and $\infty-\infty$ so that the limit can be calculated.

Example 1: Compute (1) $\lim_{x\to\infty}\left(3x^3-1000x^2\right)$. (2) $\lim_{x\to-\infty}\left(x^4+5x^2+1\right)$ (3) $\lim_{x\to\infty}\frac{100}{x^2+5}$

Solution: (1) $\lim_{x\to\infty}\left(3x^3-1000x^2\right)=\lim_{x\to\infty}\left(3x^3\right)=\infty$

(2) $\lim_{x\to-\infty}\left(x^4+5x^2+1\right)=\lim_{x\to-\infty}\left(x^4\right)=\infty$

(3) $\lim_{x\to\infty}\dfrac{100}{x^2+5}=\left(\dfrac{100}{\infty}\right)=0$

.

Example 2: Compute (1) $\lim_{x\to-\infty}\dfrac{x+7}{3x+5}$. (2) $\lim_{x\to\infty}\dfrac{7x^2+x-100}{2x^2-5x}$ (3) $\lim_{x\to\infty}\dfrac{x^2-3x+7}{x^3+10x-4}$

Solution: (1) $\lim_{x\to-\infty}\dfrac{x+7}{3x+5}=\left(\dfrac{-\infty}{-\infty}\right)=\lim_{x\to-\infty}\dfrac{\frac{x}{x}+\frac{7}{x}}{\frac{3x}{x}+\frac{5}{x}}=\lim_{x\to-\infty}\dfrac{1+\frac{7}{x}}{3+\frac{5}{x}}=\dfrac{1+0}{3+0}=\dfrac{1}{3}$

(2) $\lim_{x\to\infty}\dfrac{7x^2+x-100}{2x^2-5x}=\left(\dfrac{\infty}{\infty}\right)$ Circumvent it by dividing each term by x^2 .

So $\lim_{x\to\infty}\dfrac{7x^2+x-100}{2x^2-5x}=\lim_{x\to\infty}\dfrac{\frac{7x^2}{x^2}+\frac{x}{x^2}-\frac{100}{x^2}}{\frac{2x^2}{x^2}-\frac{5x}{x^2}}=\lim_{x\to\infty}\dfrac{7+\frac{1}{x}-\frac{100}{x^2}}{2-\frac{5}{x}}=\dfrac{7+0-0}{2-0}=\dfrac{7}{2}$

(3) Circumvent it by dividing each term by x^3 .

So $\lim_{x\to\infty}\dfrac{x^2-3x+7}{x^3+10x-4}=\lim_{x\to\infty}\dfrac{\frac{x^2}{x^3}-\frac{3x}{x^3}+\frac{7}{x^3}}{\frac{x^3}{x^3}+\frac{10x}{x^3}-\frac{4}{x^3}}=\lim_{x\to\infty}\dfrac{\frac{1}{x}-\frac{3}{x^2}+\frac{7}{x^3}}{1+\frac{10}{x^2}-\frac{4}{x^3}}=\dfrac{0-0+0}{1+0-0}=0$

Remark: Dividing by x^2, the highest power of x in the numerator, also leads to the correct answer.

Example 3: Compute (1) $\lim_{x\to\infty}\left(x-\sqrt{x^2+7}\right)$. (2) $\lim_{x\to-\infty}\left(x-\sqrt{x^2+7}\right)$

Solution: (1) $\lim_{x\to\infty}\left(x-\sqrt{x^2+7}\right)=\left(\infty-\infty\right)$

$$\lim_{x\to\infty}\left(x-\sqrt{x^2+7}\right)=\lim_{x\to\infty}\frac{\left(x-\sqrt{x^2+7}\right)\left(x+\sqrt{x^2+7}\right)}{\left(x+\sqrt{x^2+7}\right)}=\lim_{x\to\infty}\frac{\left(x^2-\left(x^2+7\right)\right)}{\left(x+\sqrt{x^2+7}\right)}=\lim_{x\to\infty}\frac{-7}{\left(x+\sqrt{x^2+7}\right)}=\frac{-7}{\infty}=0$$

(2) $\lim_{x\to-\infty}\left(x-\sqrt{x^2+7}\right)=\left(-\infty-\infty\right)$. This is not an indeterminate form. It means $=-\infty$

Exercise 1: Compute (1) $\lim_{x\to-\infty}\frac{7x^2+x+11}{4-x}$ (2) $\lim_{x\to-\infty}\frac{x+3}{\sqrt{9x^2-5x}}$ (3) $\lim_{x\to\infty}\frac{x+3}{\sqrt{9x^2-5x}}$

(4) $\lim_{x\to\infty}\left(\sqrt{5x^2+x+3}-\sqrt{5x^2+4x+7}\right)$ (5) $\lim_{x\to\infty}\left[\frac{\sqrt{5x+9x^2}}{1+3x}+2\right]$

Solution : (1) $\lim_{x\to-\infty}\frac{7x^2+x+11}{4-x}=\left(\frac{\infty}{\infty}\right)$ =

$$\lim_{x\to-\infty}\frac{\frac{7x^2}{x^2}+\frac{x}{x^2}+\frac{11}{x^2}}{\frac{4}{x^2}-\frac{x}{x^2}}=\lim_{x\to-\infty}\frac{7+\frac{1}{x}+\frac{11}{x^2}}{\frac{4}{x^2}-\frac{1}{x}}=\frac{7+0+0}{0-0}=\left(\frac{7}{0}\right)=(\infty)$$

(2)

$$\lim_{x\to-\infty}\frac{x+3}{\sqrt{9x^2-5x}}=\lim_{x\to-\infty}\frac{x+3}{\sqrt{x^2\left(9-\frac{5}{x}\right)}}=\lim_{x\to-\infty}\frac{x+3}{\sqrt{x^2}\sqrt{9-\frac{5}{x}}}=\lim_{x\to-\infty}\frac{x+3}{|x|\sqrt{9-\frac{5}{x}}}=\lim_{x\to-\infty}\frac{x+3}{-x\sqrt{9-\frac{5}{x}}}$$

$$=\lim_{x\to-\infty}\frac{x\left(1+\frac{3}{x}\right)}{-x\sqrt{9-\frac{5}{x}}}=-\lim_{x\to-\infty}\frac{\left(1+\frac{3}{x}\right)}{\sqrt{9-\frac{5}{x}}}=-\frac{1}{\sqrt{9}}=-\frac{1}{3}.$$

(4) $\lim_{x\to\infty}\left(\sqrt{5x^2+x+3}-\sqrt{5x^2+4x+7}\right)\left(\infty-\infty\right)$

$$\lim_{x\to\infty}\left(\sqrt{5x^2+x+3}-\sqrt{5x^2+4x+7}\right)=\lim_{x\to\infty}\frac{\left(\sqrt{5x^2+x+3}-\sqrt{5x^2+4x+7}\right)\left(\sqrt{5x^2+x+3}+\sqrt{5x^2+4x+7}\right)}{\left(\sqrt{5x^2+x+3}+\sqrt{5x^2+4x+7}\right)}$$

$$= \lim_{x\to\infty} \frac{\left(5x^2+x+3-\left(5x^2+4x+7\right)\right)}{\left(\sqrt{5x^2+x+3}+\sqrt{5x^2+4x+7}\right)} = \lim_{x\to\infty} \frac{-(3x+4)}{\left(\sqrt{5x^2+x+3}+\sqrt{5x^2+4x+7}\right)}$$

$$= \lim_{x\to\infty} \frac{-(3x+4)}{\left(|x|\sqrt{5+\frac{1}{x}+\frac{3}{x^2}}+|x|\sqrt{5+\frac{4}{x}+\frac{7}{x^2}}\right)} = \lim_{x\to\infty} \frac{-(3x+4)}{\left(x\sqrt{5+\frac{1}{x}+\frac{3}{x^2}}+x\sqrt{5+\frac{4}{x}+\frac{7}{x^2}}\right)}$$

$$= \lim_{x\to\infty} \frac{-x\left(3+\frac{4}{x}\right)}{x\left(\sqrt{5+\frac{1}{x}+\frac{3}{x^2}}+\sqrt{5+\frac{4}{x}+\frac{7}{x^2}}\right)} = \lim_{x\to\infty} \frac{-\left(3+\frac{4}{x}\right)}{\left(\sqrt{5+\frac{1}{x}+\frac{3}{x^2}}+\sqrt{5+\frac{4}{x}+\frac{7}{x^2}}\right)} = \frac{-3}{2\sqrt{5}}$$

---

(5) First

$$\lim_{x\to\infty} \frac{\sqrt{5x+9x^2}}{1+3x} = \lim_{x\to\infty} \frac{\sqrt{x^2\left(\frac{5}{x}+9\right)}}{1+3x} = \lim_{x\to\infty} \frac{\sqrt{x^2}\sqrt{\frac{5}{x}+9}}{1+3x} = \lim_{x\to\infty} \frac{|x|\sqrt{\frac{5}{x}+9}}{1+3x}$$

$$= \lim_{x\to\infty} \frac{x\sqrt{\frac{5}{x}+9}}{x\left(\frac{1}{x}+3\right)} = \lim_{x\to\infty} \frac{\sqrt{\frac{5}{x}+9}}{\left(\frac{1}{x}+3\right)} = \frac{3}{3} = 1$$

Second: $\lim_{x\to\infty} 2 = 2$. So $\lim_{x\to\infty}\left[\frac{\sqrt{5x+9x^2}}{1+3x}+2\right] = \lim_{x\to\infty} \frac{\sqrt{5x+9x^2}}{1+3x} + \lim_{x\to\infty} 2 = 1+2 = 3$

Example 4: Compute (1) $\lim_{x\to-\infty} \frac{e^x}{4+5e^{3x}}$ (2) $\lim_{x\to\infty} \frac{2^x}{3^x}$ (3) $\lim_{x\to\infty} \frac{5^x}{3^x+2^x}$

(4) $\lim_{x\to\infty} \frac{\ln\left(2+e^{3x}\right)}{\ln\left(1+e^x\right)}$ (5) $\lim_{x\to\infty} e^{2x-1}$

Solution: (1) $\lim_{x\to-\infty} \frac{e^x}{4+5e^{3x}} = \frac{0}{4+0} = 0$

(2) $\lim_{x\to\infty} \frac{2^x}{3^x}\left(\frac{\infty}{\infty}\right) = \lim_{x\to\infty}\left(\frac{2}{3}\right)^x = 0$

(3) $\lim_{x\to\infty} \frac{5^x}{3^x+2^x} = \lim_{x\to\infty} \frac{\frac{5^x}{3^x}}{\frac{3^x}{3^x}+\frac{2^x}{3^x}} = \lim_{x\to\infty} \frac{\left(\frac{5}{3}\right)^x}{1+\left(\frac{2}{3}\right)^x} = \frac{\infty}{0+1} = \infty$

.

(4)
$$\lim_{x\to\infty}\frac{\ln\left(2+e^{3x}\right)}{\ln\left(1+e^{x}\right)}\left(\frac{\infty}{\infty}\right)=\lim_{x\to\infty}\frac{\ln\left(e^{3x}\left(\frac{2}{e^{3x}}+1\right)\right)}{\ln\left(e^{x}\left(\frac{1}{e^{x}}+1\right)\right)}=\lim_{x\to\infty}\frac{\ln e^{3x}+\ln\left(\frac{2}{e^{3x}}+1\right)}{\ln e^{x}+\ln\left(\frac{1}{e^{x}}+1\right)}$$
$$=\lim_{x\to\infty}\frac{3x+\ln\left(\frac{2}{e^{3x}}+1\right)}{x+\ln\left(\frac{1}{e^{x}}+1\right)}=\lim_{x\to\infty}\frac{3x}{x}=\lim_{x\to\infty}3=3$$

(5) Notice that $\lim_{x\to\infty}(2x-1)=\infty$ and $\lim_{x\to\infty}e^{x}=\infty$. Combining these two results getting $\lim_{x\to\infty}e^{2x-1}=\infty$

Exercise 2: Compute (1) $\lim_{x\to\infty}\ln 2x$ (2) $\lim_{x\to 0+}e^{-2/x}$ (3) $\lim_{x\to 0+}\tan^{-1}(\ln x)$ (4) $\lim_{x\to 0+}e^{1/x^2}$

(1) Notice that $\lim_{x\to\infty}2x=\infty$ and $\lim_{x\to\infty}\ln x=\infty$. So $\lim_{x\to\infty}\ln 2x=\infty$

(2) Notices that $\lim_{x\to 0+}-\frac{2}{x}=-\infty$ and $\lim_{x\to-\infty}e^{x}=0$. Thus $\lim_{x\to 0+}e^{-2/x}=0$

(3) Notices that $\lim_{x\to 0+}\ln x=-\infty$ and $\lim_{x\to-\infty}\tan^{-1}x=-\frac{\pi}{2}$. So $\lim_{x\to 0+}\tan^{-1}(\ln x)=-\frac{\pi}{2}$

(4) Notices that $\lim_{x\to 0+}\frac{1}{x^2}=\infty$ and $\lim_{x\to\infty}e^{x}=\infty$. Thus $\lim_{x\to 0+}e^{1/x^2}=\infty$

Exercise 3: If $\lim_{x\to\infty}\frac{f(x)}{g(x)}=0$, find $\lim_{x\to\infty}\frac{f(x)-g(x)}{f(x)+g(x)}$

Solution

$$\lim_{x\to\infty}\frac{f(x)-g(x)}{f(x)+g(x)}=\lim_{x\to\infty}\frac{\frac{f(x)}{g(x)}-\frac{g(x)}{g(x)}}{\frac{f(x)}{g(x)}+\frac{g(x)}{g(x)}}=\lim_{x\to\infty}\frac{\frac{f(x)}{g(x)}-1}{\frac{f(x)}{g(x)}+1}=\frac{0-1}{0+1}=-1$$

Definition: A line $y=L$ is called a horizontal asymptote of the graph of f if either

$\lim_{x\to\infty}f(x)=L$ or $\lim_{x\to-\infty}f(x)=L$

Example 5: Find the horizontal asymptotes of (1) $f(x)=\frac{x-1}{x-2}$ (2) $f(x)=\frac{x^2-1}{x^3}$

(3) $f(x)=\begin{cases}\dfrac{x+1}{x-2}, x<0\\ x^3-3x-1, x\geq 0\end{cases}$ (4) $f(x)=x\ln x^2$ (5) $f(x)=6x^{1/3}+3x^{4/3}$

Solution: (1) $\lim_{x\to\infty} f(x)=1$ and $\lim_{x\to-\infty} f(x)=1$. So the horizontal asymptote is the line $y=1$.

(2) $\lim_{x\to\infty} f(x)=1$ and $\lim_{x\to-\infty} f(x)=1$. So the horizontal asymptote is the line $y=1$.

(3) $\lim_{x\to-\infty} f(x)=\lim_{x\to-\infty}\dfrac{x+1}{x-2}=1$ $\qquad \lim_{x\to\infty} f(x)=\lim_{x\to\infty}\left(x^3-3x-1\right)=\infty$.

So the horizontal asymptote is the line $y=1$ as $x\to-\infty$.

(4) $\lim_{x\to\infty} f(x)=\infty$ and $\lim_{x\to-\infty} f(x)=-\infty$. So f has no horizontal asymptote.

(5) $\lim_{x\to\infty}\left[6x^{1/3}+3x^{4/3}\right]=\infty$ $\qquad \lim_{x\to-\infty}\left[6x^{1/3}+3x^{4/3}\right]=\infty$. So f has no horizontal asymptote.

Exercise 4: Find the horizontal asymptotes of (1) $f(x)=x^2\ln x$ (2) $f(x)=\dfrac{x}{\sqrt{4+x^2}}$

(3) $f(x)=4\tan^{-1}x-1$ (4) $f(x)=3e^{-1/x}$ (5) $f(x)=\begin{cases}\dfrac{4x}{x-4}, x<0\\ \dfrac{x^2}{x-2}, 0\leq x<4\\ \dfrac{e^{-x}}{x+1}, x\geq 4\end{cases}$

Solution: (1) $\lim_{x\to\infty} f(x)=\infty$. So f has no horizontal asymptote.

(2)
$$\lim_{x\to\infty}\frac{x}{\sqrt{4+x^2}}=\lim_{x\to\infty}\frac{x}{|x|\sqrt{\frac{4}{x^2}+1}}=\lim_{x\to\infty}\frac{x}{x\sqrt{\frac{4}{x^2}+1}}=\lim_{x\to\infty}\frac{1}{\sqrt{\frac{4}{x^2}+1}}=\frac{1}{\sqrt{1}}=1$$
$$\lim_{x\to-\infty}\frac{x}{\sqrt{4+x^2}}=\lim_{x\to-\infty}\frac{x}{|x|\sqrt{\frac{4}{x^2}+1}}=\lim_{x\to-\infty}\frac{x}{-x\sqrt{\frac{4}{x^2}+1}}=\lim_{x\to\infty}-\frac{1}{\sqrt{\frac{4}{x^2}+1}}=-\frac{1}{\sqrt{1}}=-1$$

So the horizontal asymptotes are $y=1$ as $x\to\infty$ and $y=-1$ as $x\to-\infty$

(3) $\lim_{x\to\infty} 4\tan^{-1}x-1=4\left(\dfrac{\pi}{2}\right)-1=2\pi-1$ $\qquad \lim_{x\to-\infty} 4\tan^{-1}x-1=4\left(-\dfrac{\pi}{2}\right)-1=-2\pi-1$

So the horizontal asymptotes are $y=2\pi-1$ as $x\to\infty$ and $y=-2\pi-1$ as $x\to-\infty$.

(4) $\lim_{x\to\infty} 3e^{-1/x} = 3$ $\qquad$ $\lim_{x\to-\infty} 3e^{-1/x} = 3$. So the horizontal asymptote is the line $y = 3$.

(5) $\lim_{x\to-\infty} f(x) = \lim_{x\to-\infty} \frac{4x}{x-4} = 4$ $\qquad$ $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{e^{-x}}{x+1} = 0$

So the horizontal asymptotes are the lines $y = 0$ as $x \to \infty$ and $y = 4$ as $x \to -\infty$.

Lecture 7: Limits of Trigonometric Functions:

Theorem

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

Example 1: Evaluate the following limit.

$$\lim_{x\to 0}\left(\frac{1-\cos x}{x}\right)$$

Solution

$$\lim_{x\to 0} \frac{1-\cos x}{x}\left(\frac{0}{0}\right) = \lim_{x\to 0} \frac{1-\cos x}{x} \cdot \frac{1+\cos x}{1+\cos x} = \lim_{x\to 0} \frac{1-\cos^2 x}{x(1+\cos x)} = \lim_{x\to 0} \frac{\sin^2 x}{x(1+\cos x)}$$

$$= \lim_{x\to 0} \frac{\sin x}{x} \cdot \lim_{x\to 0} \frac{\sin x}{(1+\cos x)} = (1)(\frac{0}{1+1}) = (1)(0) = 0$$

Example 2: show that.

$$1. \lim_{x\to 0} \frac{\sin ax}{bx} = \lim_{x\to 0} \frac{ax}{\sin bx} = \lim_{x\to 0} \frac{\sin ax}{\sin bx} = \frac{a}{b}$$

$$2. \lim_{x\to 0} \frac{\tan ax}{bx} = \lim_{x\to 0} \frac{ax}{\tan bx} = \lim_{x\to 0} \frac{\tan ax}{\tan bx} = \frac{a}{b}$$

$$3. \lim_{x\to 0} \frac{\sin ax}{\tan bx} = \lim_{x\to 0} \frac{\tan ax}{\sin bx} = \frac{a}{b}$$

Solution

We will show that

$$\lim_{x\to 0} \frac{\sin ax}{bx} = \frac{a}{b} \qquad \lim_{x\to 0} \frac{\tan ax}{bx} = \frac{a}{b} \qquad \lim_{x\to 0} \frac{\sin ax}{\tan bx} = \frac{a}{b}$$

and left the rest as an exercise.

Solution 1

$$\lim_{x\to 0}\frac{\sin ax}{bx}=\lim_{x\to 0}\frac{\frac{\sin ax}{ax}}{\frac{bx}{ax}}=\lim_{x\to 0}\frac{\frac{\sin ax}{ax}}{\frac{b}{a}}=\lim_{x\to 0}\frac{a}{b}\frac{\sin ax}{ax}$$

let $y=ax$, as $x\to 0, y\to 0$. So

$$\lim_{x\to 0}\frac{\sin ax}{bx}=\frac{a}{b}\lim_{x\to 0}\frac{\sin ax}{ax}=\frac{a}{b}\lim_{y\to 0}\frac{\sin y}{y}=\frac{a}{b}(1)=\frac{a}{b}$$

Solution 2

$$\lim_{x\to 0}\frac{\tan ax}{bx}=\lim_{x\to 0}\left(\frac{\frac{\sin ax}{\cos bx}}{bx}\right)=\lim_{x\to 0}\left(\frac{1}{\cos bx}.\frac{\sin ax}{bx}\right)=\lim_{x\to 0}\left(\frac{1}{\cos bx}\right).\lim_{x\to 0}\left(\frac{\sin ax}{bx}\right)=1.\frac{a}{b}=\frac{a}{b}$$

Solution 3

$$\lim_{x\to 0}\frac{\sin ax}{\tan bx}=\lim_{x\to 0}\frac{\frac{\sin ax}{ax}}{\frac{\tan bx}{ax}}=\frac{\lim_{x\to 0}\frac{\sin ax}{ax}}{\lim_{x\to 0}\frac{\tan bx}{ax}}=\frac{1}{\frac{b}{a}}=\frac{a}{b}$$

Example 3: Evaluate the following limit

$$\lim_{x\to 0}\frac{4x}{\tan 3x+\sin 2x}$$

Solution

$$\lim_{x\to 0}\frac{4x}{\tan 3x+\sin 2x}\left(\frac{0}{0}\right)=\lim_{x\to 0}\frac{\frac{4x}{x}}{\frac{\tan 3x+\sin 2x}{x}}=\lim_{x\to 0}\frac{4}{\frac{\tan 3x}{x}+\frac{\sin 2x}{x}}=\frac{4}{\frac{3}{1}+\frac{2}{1}}=\frac{4}{5}$$

Example 4: Evaluate the following limit

$$\lim_{x\to 0}\frac{5\sin 3x+\tan 7x}{3x+x^2}$$

Solution

$$\lim_{x\to 0}\frac{5\sin 3x+\tan 7x}{3x+x^2}\left(\frac{0}{0}\right)=\lim_{x\to 0}\frac{\frac{5\sin 3x}{x}+\frac{\tan 7x}{x}}{\frac{3x}{x}+\frac{x^2}{x}}=\lim_{x\to 0}\frac{\frac{5\sin 3x}{x}+\frac{\tan 7x}{x}}{3+x}=\frac{5\left(\frac{3}{1}\right)+\left(\frac{7}{1}\right)}{3+0}=\frac{22}{3}$$

Example 5: Evaluate the following limit

$$\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta\sin\theta}$$

Solution

$$\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta\sin\theta}\left(\frac{0}{0}\right)=\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta\sin\theta}.\frac{1+\cos\theta}{1+\cos\theta}=\lim_{\theta\to 0}\frac{1-\cos^2\theta}{\theta\sin\theta(1+\cos\theta)}=\lim_{\theta\to 0}\frac{\sin^2\theta}{\theta\sin\theta(1+\cos\theta)}$$

$$=\lim_{\theta\to 0}\frac{\sin\theta}{\theta(1+\cos\theta)}=\lim_{\theta\to 0}\frac{\sin\theta}{\theta}.\lim_{\theta\to 0}\frac{1}{(1+\cos\theta)}=(1)(\frac{1}{1+1})=\frac{1}{2}$$

Example 6: Evaluate the following limit

$$\lim_{t\to\pi}\frac{\sin t}{\pi-t}$$

Solution

$\lim_{t\to\pi}\frac{\sin t}{\pi-t}=\left(\frac{0}{0}\right)$. Let $y=\pi-t$, as $t\to\pi$, $y\to 0$. So

$$\lim_{t\to\pi}\frac{\sin t}{\pi-t}=\lim_{y\to 0}\frac{\sin(\pi-y)}{y}=\lim_{y\to 0}\frac{\sin\pi\cos y-\cos\pi\sin y}{y}=\lim_{y\to 0}\frac{0-(-1)\sin y}{y}=\lim_{y\to 0}\frac{\sin y}{y}=1$$

Example 7: Evaluate the following limit

$$\lim_{x\to 0}\frac{x\tan x+\cos 2x-1}{x^2}$$

Solution

$$\lim_{x\to 0}\left[\frac{x\tan x+\cos 2x-1}{x^2}\right]\left(\frac{0}{0}\right)=\lim_{x\to 0}\left[\frac{x\tan x+(1-2\sin^2 x)-1}{x^2}\right]\quad \text{use the identity } \left[\cos(2x)=1-2\sin^2 x\right]$$

$$=\lim_{x\to 0}\left[\frac{x\tan x-2\sin^2 x}{x^2}\right]=\lim_{x\to 0}\left[\frac{x\tan x}{x^2}-\frac{2\sin^2 x}{x^2}\right]=\lim_{x\to 0}\left[\frac{\tan x}{x}-2.\frac{\sin x}{x}.\frac{\sin x}{x}\right]$$

$$=\lim_{x\to 0}\left[\frac{\tan x}{x}\right]-2\lim_{x\to 0}\left[\frac{\sin x}{x}\right].\lim_{x\to 0}\left[\frac{\sin x}{x}\right]=1-(2)(1)(1)=-1$$

Example 8: Evaluate the following limit.

$$\lim_{x \to \pi/2} \frac{\cos(x)}{\cos(-x)}$$

Solution

$$\lim_{x \to \pi/2} \frac{\cos(x)}{\cos(-x)}\left(\frac{0}{0}\right) = \lim_{x \to \pi/2} \frac{\cos(x)}{\cos(x)} = \lim_{x \to \pi/2} 1 = 1$$

Example 9: Evaluate the following limit

$$\lim_{x \to 1} \frac{\sin(3x-3)}{1-x^3}$$

Solution

$$\lim_{x \to 1}\left[\frac{\sin(3x-3)}{1-x^3}\right]\left(\frac{0}{0}\right) = \lim_{x \to 1} \frac{\sin[3(x-1)]}{-(x-1)(1+x+x^2)} = \lim_{x \to 1}\left[\frac{-\sin[3(x-1)]}{(x-1)}\right].\lim_{x \to 1}\left[\frac{1}{1+x+x^2}\right]$$

To evaluate $\lim_{x \to 1} \frac{\sin 3(x-1)}{(x-1)}$

Let $y = x-1$, as $x \to 1, y \to 0$.So

$$\lim_{x \to 1} \frac{\sin 3(x-1)}{(x-1)} = \lim_{y \to 0} \frac{\sin 3y}{y} = 3$$

Therefore,

$$\lim_{x \to 1}\left[\frac{\sin(3x-3)}{1-x^3}\right]\left(\frac{0}{0}\right) = \lim_{x \to 1} \frac{\sin[3(x-1)]}{-(x-1)(1+x+x^2)} = 3\left(-\frac{1}{3}\right) = -1$$

Example 10: Evaluate the following limit

$$\lim_{x \to 0} \frac{\tan 2x}{\sqrt{3x+1}-1}$$

Solution

$$\lim_{x\to 0}\frac{\tan 2x}{\sqrt{3x+1}-1}\left(\frac{0}{0}\right)=\lim_{x\to 0}\frac{\tan 2x}{\sqrt{3x+1}-1}\cdot\frac{\sqrt{3x+1}+1}{\sqrt{3x+1}+1}=\lim_{x\to 0}\frac{\tan 2x\,(\sqrt{3x+1}+1)}{(3x+1)-1}$$

$$=\lim_{x\to 0}\frac{\tan 2x\,(\sqrt{3x+1}+1)}{3x}=\lim_{x\to 0}\frac{\tan 2x}{3x}\cdot\lim_{x\to 0}(\sqrt{3x+1}+1)=(\frac{2}{3})(2)=\frac{4}{3}$$

Example 11: Evaluate the following limit

$$\lim_{x\to 1}\frac{\sin^2(\pi x)}{1-2x+x^2}$$

Solution

$$\lim_{x\to 1}\frac{(\sin(\pi x))^2}{1-2x+x^2}\left(\frac{0}{0}\right)=\lim_{x\to 1}\frac{(\sin(\pi-\pi x))^2}{(1-x)^2}=\lim_{x\to 1}\frac{(\sin(\pi(1-x))^2}{(1-x)^2}=\lim_{x\to 1}\frac{\sin\pi(1-x)}{(1-x)}\cdot\lim_{x\to 1}\frac{\sin\pi(1-x)}{(1-x)}$$ Let $y=x-1$, as $x\to 1, y\to 0$. So

$$\lim_{x\to 1}\frac{\sin\pi(1-x)}{(1-x)}\cdot\lim_{x\to 1}\frac{\sin\pi(1-x)}{(1-x)}=\lim_{y\to 0}\frac{\sin\pi y}{y}\cdot\lim_{y\to 0}\frac{\sin\pi y}{y}=(\pi)(\pi)=\pi^2$$

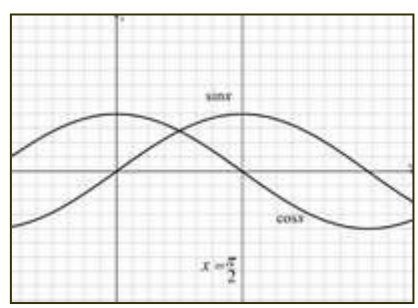
Example 12: Find all values of k if $\lim_{x\to 0}\frac{\sin^2(kx)}{x^2}=4$.

Solution

$$\lim_{x\to 0}\frac{\sin^2(kx)}{x^2}=4\Rightarrow\lim_{x\to 0}\left[\frac{\sin(kx)}{x}\cdot\frac{\sin(kx)}{x}\right]=4\Rightarrow k^2=4\Rightarrow k=-2,2$$

Example 13: Evaluate the following limit.

$$\lim_{x\to\frac{\pi}{2}}\tan x$$



Solution

Notice that

$$\lim_{x\to\frac{\pi}{2}}\tan x=\lim_{x\to\frac{\pi}{2}}\frac{\sin x}{\cos x}$$

The limit of the numerator is 1, and the limit of the denominator is 0. So the limit of the ratio does not exist. To be more specific than this, we need to analyze the sign of the ratio.

Let's take a look at the left-handed limit first. In this case we are going to be assuming that whatever x is it will be less than $\frac{\pi}{2}$. Therefore, as x gets closer and closer to $x = \frac{\pi}{2}$ the numerator is getting closer and closer to 1 while the denominator is getting closer and closer to 0 and will always be positive since we know that whatever x is it must satisfy $x < \frac{\pi}{2}$ (see the figure).

So, as we get closer and closer to $x = \frac{\pi}{2}$ (from the left) we have a positive, finite number in the numerator divided by an increasingly smaller positive number. This will result in increasing large and positive numbers. In other words,

$$\lim_{x \to \frac{\pi}{2}^-} \tan x = \infty$$

The right-handed limit is similar. As we move in towards $x = \frac{\pi}{2}$ from the right we will always have $x > \frac{\pi}{2}$ and so we will have a positive, finite number in the numerator divided by a increasingly smaller negative number and so the whole thing should be getting larger and larger. In this case the right-handed limit is,

$$\lim_{x \to \frac{\pi}{2}^+} \frac{\sin x}{\cos x} = -\infty$$

So $\lim_{x \to \frac{\pi}{2}} \frac{\sin x}{\cos x}$ doesn't exist .

Theorem: Squeeze Theorem

Suppose that for all x on $[a,b]$ we have,

$$g(x) \le f(x) \le h(x)$$

Also suppose that,

$$\lim_{x \to c} g(x) = \lim_{x \to c} h(x) = L$$

for some $a \le c \le b$ .Then,

$$\lim_{x \to c} f(x) = L$$

The Squeeze theorem is also known as the Sandwich Theorem and the Pinching Theorem.

So, how do we use this theorem to help us with limits? Let's take a look at the following example to see the theorem in action.

Example 14: Evaluate the following limit.

$$\lim_{x \to 0} x^2 \cos\left(\frac{1}{x}\right)$$

Solution: In this example none of the previous examples can help us. There's no factoring or simplifying to do. We can't rationalize and one-sided limits won't work. There's even a question as to whether this limit will exist since we have division by zero inside the cosine at $x = 0$.

We know the following about cosine.

$$-1 \le \cos x \le 1$$

We don't just have an x in the cosine, but as long as we avoid $x = 0$ we can say the same thing for our cosine.

$$-1 \le \cos\left(\frac{1}{x}\right) \le 1 \quad [\text{for all } x \ne 0]$$

Its okay for us to ignore $x = 0$ since we are taking a limit and we know that limits don't care about what's actually going on at $x = 0$ in this case.

Now if we have the above inequality for our cosine we can just multiply everything by an x^2 and get the following.

$$-x^2 \le x^2 \cos\left(\frac{1}{x}\right) \le x^2$$

In other words we've managed to squeeze the function that we where interested in between two other function that are very easy to deal with. So, the limits of the two outer functions are.

$$\lim_{x \to 0} x^2 = 0 = \lim_{x \to 0}\left(-x^2\right)$$

These are the same and so by the Squeeze theorem we must also have,

$$\lim_{x \to 0} x^2 \cos\left(\frac{1}{x}\right) = 0$$

Remark: the squeezing theorem also holds for one sided limits and limits at ∞ and $-\infty$.

Example 15: Evaluate the following limit.

$$\lim_{x \to \infty}\left(e^{-3x} \cos 2x\right)$$

Solution: We know the following about cosine.

$$-1 \le \cos x \le 1$$

We can say the same thing for our cosine.

$$-1 \le \cos(2x) \le 1$$

Now if we have the above inequality for our cosine we can just multiply everything by an e^{-x} and get the following.

$$-e^{-x} \le e^{-x}\cos(2x) \le e^{-x}$$

The limits of the two outer functions are.

$$\lim_{x\to\infty} -e^{-x} = 0 = \lim_{x\to\infty} e^{-x}$$

These are the same and so by the Squeeze theorem we must also have,

$$\lim_{x\to\infty}\left(e^{-x}\cos(2x)\right) = 0$$

Example 16: Evaluate the following limit.

$$\lim_{x\to 0+}\sqrt{x}\sin\left(x+\frac{1}{x}\right)$$

Solution: We know the following about sine.

$$-1 \le \sin x \le 1$$

We can say the same thing for our sine.

$$-1 \le \sin\left(x+\frac{1}{x}\right) \le 1 \qquad [\text{for all } x \ne 0]$$

If $x > 0$, then $\sqrt{x} > 0$. Multiply through by $\sqrt{x}$, we get

$$-\sqrt{x} \le \sqrt{x}\sin\left(x+\frac{1}{x}\right) \le \sqrt{x}$$

In other words we've managed to squeeze the function that we where interested in between two other function that are very easy to deal with. So, the limits of the two outer functions are.

$$\lim_{x\to 0+} -\sqrt{x} = 0 = \lim_{x\to 0+}\sqrt{x}$$

These are the same and so by the Squeeze theorem we must also have,

$$\lim_{x \to 0+} \sqrt{x} \cos\left(x + \frac{1}{x}\right) = 0$$

Example 17: Evaluate the following limit.

$$\lim_{x \to 0+} x^{1/3} \cos\left(2 + \frac{1}{x}\right)$$

Solution: We know the following about cosine.

$$-1 \le \cos x \le 1$$

We can say the same thing for our cosine.

$$-1 \le \cos\left(2 + \frac{1}{x}\right) \le 1 \qquad [\text{for all } x \ne 0]$$

If $x > 0$, then $x^{1/3} > 0$ and so

$$-x^{1/3} \le x^{1/3} \cos\left(2 + \frac{1}{x}\right) \le x^{1/3}$$

The limits of the two outer functions are.

$$\lim_{x \to 0+} -x^{1/3} = 0 = \lim_{x \to 0+} x^{1/3}$$

These are the same and so by the Squeeze theorem we must also have,

$$\lim_{x \to 0+} x^{1/3} \cos\left(2 + \frac{1}{x}\right) = 0$$

Continuity

Definition

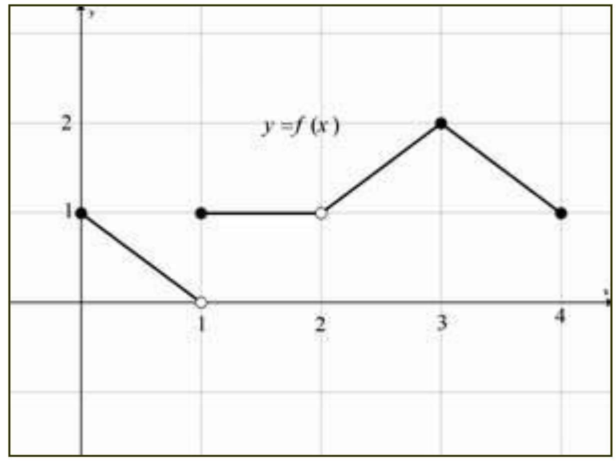
A function $f(x)$ is said to be continuous at $x = c$ if

1. $f(c)$ is defined. 2. $\lim_{x \to c} f(x)$ is exist. 3. $\lim_{x \to c} f(x) = f(c)$

A function is said to be continuous on the interval $[a,b]$ if it is continuous at each point in the interval.

Now, this definition justifies how we've been computing limits for awhile now and so it's good in that sense. However, it doesn't really tell us just what it means for a function to be continuous. Let's take a look at the following example to help us understand just what it means for a function to be continuous.

Example 1: Given the graph of $f(x)$ shown below, determine if $f(x)$ is continuous at $x=1$, $x=2$, and $x=3$.



Solution: To answer the question for each point we'll need to get both the limit at that point and the function value at that point. If they are equal the function is continuous at that point and if they aren't equal the function isn't continuous at that point.

First $x=1$.

$f(1)=1 \qquad \lim_{x\to 1} f(x)$ doesn't exist

The function value and the limit aren't the same and so the function is not continuous at this point.

Now $x=2$.

$f(2)$: undefined $\qquad \lim_{x\to 1} f(x)=2$

The function value and the limit aren't the same and so the function is not continuous at this point.

Finally $x=3$.

$f(3)=2 \qquad \lim_{x\to 3} f(x)=2$

The function is continuous at this point since the function and limit have the same value.

Example 2: discuss the continuity at $x=3$ for the following functions.

$$f(x)=\frac{x^2-9}{x-3} \qquad g(x)=\begin{cases}\dfrac{x^2-9}{x-3}, x\neq 3\\ 4, x=3\end{cases} \qquad h(x)=\begin{cases}\dfrac{x^2-9}{x-3}, x\neq 3\\ 6, x=3\end{cases}$$

Solution

1. the functions is undefined at $x=3$, and hence is not continuous at that point.

2.

$$f(3)=4$$

$$\lim_{x\to 3} f(x)=\lim_{x\to 3}\frac{x^2-9}{x-3}=\lim_{x\to 3}\frac{(x-3)(x+3)}{x-3}=\lim_{x\to 3}(x+3)=6$$

The function value and the limit aren't the same and so the function is discontinuous at $x=3$.

3.

$f(3) = 6$ $\quad\quad\quad$ $\lim\limits_{x \to 3} f(x) = 6$

The function is continuous at $x = 3$ since the function and limit have the same value.

Example 3: Find values of a,b where the function

$$f(x) = \begin{cases} ax^2 + 1, & x > 2 \\ -11, & x = 2 \\ x^3 + b, & x < 2 \end{cases}.$$

Is continuous.

Solution: f is continuous every where. In particular f is continuous at $x = 2$. This implies that:

1. $\lim\limits_{x \to 2} f(x)$ exist. This implies that:

$$\lim_{x \to 2+} f(x) = \lim_{x \to 2-} f(x)$$
$$\Rightarrow \lim_{x \to 2+} \left(ax^2 + 1\right) = \lim_{x \to 2-} \left(x^3 + b\right)$$
$$\Rightarrow 4a + 1 = 8 + b \qquad (1)$$

2. $\lim\limits_{x \to 2} f(x) = f(2)$. this implies that

$$\lim_{x \to 2} f(x) = -11$$
$$\Rightarrow 4a + 1 = -11 \qquad (\text{or } 8 + b = -11)$$
$$\Rightarrow a = -3$$

Substitute $a = -3$ in (1), getting

$b = -19$

Example 4: Find all values of the constant k that make

$$f(x) = \begin{cases} \dfrac{2x}{\tan kx} + 1, & x < 0 \\ 3x + k, & x \geq 0 \end{cases}.$$

Is continuous at $x = 0$.

Solution: To make f continuous at $x = 0$, we must have $\lim\limits_{x \to 0} f(x)$ exist.

i.e

$$\lim_{x\to 0+} f(x) = \lim_{x\to 0-} f(x)$$

$$\Rightarrow \lim_{x\to 0+}(3x+k) = \lim_{x\to 0-}\left(\frac{2x}{\tan kx}+1\right) \Rightarrow k = \left(\frac{2}{k}\right)+1, k\neq 0$$

$$\Rightarrow k^2 = 2+k \Rightarrow k^2 - k - 2 = 0 \Rightarrow (k-2)(k+1) = 0 \Rightarrow k = 2 \text{ and } k = -1$$

Example 5: Find value(s) for the constant k so that

$$f(x) = \begin{cases} \dfrac{\sin kx}{x}, & x<0 \\ 3x+2k^2, & x\geq 0 \end{cases}.$$

will be continuous at $x=0$.

Solution: To make f continuous at $x=0$, we must have $\lim_{x\to 0} f(x)$ exist.

i.e

$$\lim_{x\to 0+} f(x) = \lim_{x\to 0-} f(x)$$

$$\Rightarrow \lim_{x\to 0+} 3x+2k^2 = \lim_{x\to 0-}\frac{\sin kx}{x} \Rightarrow 2k^2 = k \Rightarrow 2k^2 - k = 0 \Rightarrow k(2k-1) = 0 \Rightarrow k = 0, k = \frac{1}{2}$$

Theorem: Polynomials are continuous every where

Theorem: A rational function is continuous everywhere except at the point where the denominator is 0.

Example 6: discuss the points of discontinuity for the function.

$$f(x) = \frac{x^2-9}{x^2-5x+6}$$

Solution: A rational function is continuous everywhere except at the point where the denominator is 0.Solving this equation

$$x^2-5x+6=0$$

Yields two points of discontinuity $x=2$ and $x=3$.

Theorem: If $\lim_{x\to c} g(x) = L$ and if the function $f(x)$ is continuous at L, then

$$\lim_{x\to c} f(g(x)) = f(\lim_{x\to c} g(x))$$

Remark: the above theorem also holds for one sided limits and limits at ∞ and $-\infty$.

With this fact we can now do limits like the following examples

Example 7: Evaluate the following limit.

$$\lim_{x \to 0} e^{\sin x}$$

Solution: Since we know that exponentials are continuous we can use the fact above.

$$\lim_{x \to 0} e^{\sin x} = e^{\lim_{x \to 0} \sin x} = e^0 = 1$$

Example 8: Compute $\lim_{x \to \infty} \log_{10}\left(\frac{x^6 - 500}{x^6 + 500}\right)$.

Solution

$$\lim_{x \to \infty} \log_{10}\left(\frac{x^6 - 500}{x^6 + 500}\right) = \log_{10}\left[\lim_{x \to \infty}\left(\frac{x^6 - 500}{x^6 + 500}\right)\right]$$

The previous step is valid because of the continuity of the logarithm function. Note also that the expression $\frac{x^6 - 500}{x^6 + 500}$ leads to the indeterminate form $\frac{\infty}{\infty}$. Circumvent it by dividing each term by x^6, the highest power of x.

$$= \log_{10}\left[\lim_{x \to \infty}\left(\frac{\frac{x^6}{x^6} - \frac{500}{x^6}}{\frac{x^6}{x^6} + \frac{500}{x^6}}\right)\right] = \log_{10}\left[\lim_{x \to \infty}\left(\frac{1 - \frac{500}{x^6}}{1 + \frac{500}{x^6}}\right)\right] = \log_{10}\left(\frac{1-0}{1+0}\right) = \log_{10} 1 = 0$$

The term $\frac{500}{x^6}$ approaches 0 as x approaches ∞.

Example 9: Evaluate the following limit.

$$\lim_{x \to 3} \sqrt{\frac{3x - 9}{2x^2 - 18}}$$

Solution

$$\lim_{x \to 3} \sqrt{\frac{3x - 9}{2x^2 - 18}} = \sqrt{\lim_{x \to 3} \frac{3x - 9}{2x^2 - 18}\left(\frac{0}{0}\right)}$$

previous step is valid because of the continuity of the square root function.

$$= \sqrt{\lim_{x\to 3} \frac{3(x-3)}{2(x^2-9)}} = \sqrt{\lim_{x\to 3} \frac{3(x-3)}{2(x-3)(x+3)}}$$

$$= \sqrt{\lim_{x\to 3} \frac{3}{2(x+3)}} = \sqrt{\frac{3}{2(3+2)}} = \sqrt{\frac{3}{12}} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

Example 10: Compute $\lim_{x\to\infty} \sqrt{\frac{x^3+7x}{4x^3+5}}$.

Solution

$$\lim_{x\to\infty} \sqrt{\frac{x^3+7x}{4x^3+5}} = \sqrt{\lim_{x\to\infty} \frac{x^3+7x}{4x^3+5} \left(\frac{\infty}{\infty}\right)}$$

The previous step is valid because of the continuity of the square root function.

Inside the square root sign lies an indeterminate form. Circumvent it by dividing each term by x^3, the highest power of x inside the square root sign.

$$= \sqrt{\lim_{x\to\infty} \frac{\frac{x^3}{x^3}+\frac{7x}{x^3}}{\frac{4x^3}{x^3}+\frac{5}{x^3}}} = \lim_{x\to\infty} \sqrt{\lim_{x\to\infty} \frac{1+\frac{7}{x^2}}{4+\frac{5}{x^3}}} = \sqrt{\frac{1+0}{4+0}} = \frac{1}{2}$$

Each of the two expressions $\frac{7}{x^2}$ and $\frac{5}{x^3}$ approaches 0 as x approaches ∞.

Another very nice consequence of continuity is the Intermediate Value Theorem.

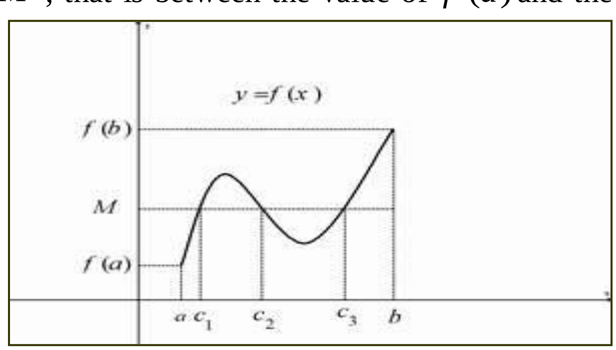
Theorem: Intermediate Value Theorem

Suppose that $f(x)$ is continuous on $[a,b]$ and let M be any number between $f(a)$ and $f(b)$.

Then there exists a number $c \in (a,b)$ such that, $f(c) = M$.

Remark: All the Intermediate Value Theorem is really saying is that a continuous function will take on all values between $f(a)$ and $f(b)$. Below is a graph of a continuous function that illustrates the Intermediate Value Theorem.

As we can see from this image if we pick any value, M, that is between the value of $f(a)$ and the value of $f(b)$ and draw a line straight out from this point the line will hit the graph in at least one point. In other words somewhere between a and b the function will take on the value of M. Note that the figure also shows that it may take on the value more than one place.



It's also important to note that the Intermediate value theorem only says that the function will take on the value of M somewhere between a and b. It doesn't say just what that value will be. It only says that it exists. It also does not tell us how many times the function may take on this value. It only tells us that it takes the value at least once.

A nice use of the Intermediate Value Theorem is to prove the existence of roots of equations.

Example 11: Show that $f(x) = x^3 + 2x - 15$ has a root somewhere in the interval $[2,3]$.

Solution

What we're really asking here is whether or not the function will take on the value

$f(x) = 0$

somewhere between 2 and 3. In other words, we're using $M = 0$ in the Intermediate value theorem. All we need to show is that 0 is between $f(2)$ and $f(3)$ and we'll be done.

$f(2) = -3$ $\qquad\qquad$ $f(3) = 18$

So by the Intermediate Value Theorem there must be a number c somewhere between 2 and 3 such that

$f(c) = 0$

Therefore the function does have a root between 2 and 3.

Example 12: If f and g are continuous on $[a,b]$ and $f(a) > g(a)$, $f(b) < g(b)$, then there is at least one solution of the equation $f(x) = g(x)$ in (a,b).

Solution

Let $h(x) = f(x) - g(x)$.Since both $f(x)$ and $g(x)$ are both continuous on $[a,b]$, then $h(x)$ is continuous on $[a,b]$.

Notice that $h(a) = f(a) - g(a) > 0$ and $h(b) = f(b) - g(b) < 0$

So by the Intermediate Value Theorem there must be a number c somewhere between a and b such that

$h(c) = 0$

This implies that $f(c) - g(c) = 0$ or $f(c) = g(c)$.

Therefore the equation $f(x) = g(x)$ has a solution in (a,b).

UNIT-III
Techniques of Differentiation

I. Notations for the Derivative

The derivative of $y = f(x)$ may be written in any of the following ways:

$$f'(x), \qquad y', \qquad \frac{dy}{dx}, \qquad \frac{d}{dx}\left[f(x)\right], \quad \text{or} \quad D_x\left[f(x)\right].$$

II. Basic Differentiation Rules

A. Suppose c and n are constants, and f *and* g are differentiable functions.

(1) $f(x) = cg(x)$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{cg(b) - cg(x)}{b - x} = c \lim_{b \to x} \frac{g(b) - g(x)}{b - x} = cg'(x)$$

(2) $f(x) = g(x) \pm k(x)$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{[g(b) \pm k(b)] - [g(x) \pm k(x)]}{b - x} =$$

$$\lim_{b \to x} \frac{g(b) - g(x)}{b - x} \pm \lim_{b \to x} \frac{k(b) - k(x)}{b - x} = g'(x) \pm k'(x)$$

(3) $f(x) = g(x)k(x)$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{g(b)k(b) - g(x)k(x)}{b - x} =$$

$$\lim_{b \to x} \frac{g(b)k(b) - g(b)k(x) + g(b)k(x) - g(x)k(x)}{b - x} =$$

$$\left[\lim_{b \to x} g(b)\right]\left[\lim_{b \to x} \frac{k(b) - k(x)}{b - x}\right] + \left[\lim_{b \to x} k(x)\right]\left[\lim_{b \to x} \frac{g(b) - g(x)}{b - x}\right] =$$

$$g(x)k'(x) + k(x)g'(x) \quad \text{(Product Rule)}$$

(4) $f(x) = \dfrac{g(x)}{k(x)} \Rightarrow f(x)k(x) = g(x) \Rightarrow g'(x) = f(x)k'(x) + k(x)f'(x) \Rightarrow$

$$f'(x) = \frac{g'(x) - f(x)k'(x)}{k(x)} = \frac{g'(x) - \left[\dfrac{g(x)}{k(x)}\right]k'(x)}{k(x)} = \frac{k(x)g'(x) - g(x)k'(x)}{[k(x)]^2}.$$

This derivative rule is called the Quotient Rule.

(5) $f(x) = c$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{c - c}{b - x} = \lim_{b \to x} \frac{0}{b - x} = \lim_{b \to x} 0 = 0$$

(6) $f(x) = x$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{b - x}{b - x} = \lim_{b \to x} 1 = 1$$

(7) $f(x) = x^n$

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

$$\lim_{h \to 0} \frac{\left[x^n + nx^{n-1}h + \dfrac{n(n-1)}{2}x^{n-2}h^2 + \ldots\right] - x^n}{h} =$$

$$\lim_{h \to 0} \left[\frac{nx^{n-1}h + h^2 \left(\frac{n(n-1)}{2} x^{n-2} + \ldots \right)}{h} \right] =$$

$$\lim_{h \to 0} \left[nx^{n-1} + h \left(\frac{n(n-1)}{2} x^{n-2} + \ldots \right) \right] = nx^{n-1} \quad \text{(Power Rule)}$$

Example 1: Suppose f and g are differentiable functions such that $f(1) = 3$, $g(1) = 7$, $f'(1) = -2$, and $g'(1) = 4$. Find (i) $(f+g)'(1)$, (ii) $(g-f)'(1)$, (iii) $(fg)'(1)$, (iv) $\left(\frac{g}{f}\right)'(1)$, and $\left(\frac{f}{g}\right)'(1)$.

(i) $(f+g)'(1) = f'(1) + g'(1) = -2 + 4 = 2$

(ii) $(g-f)'(1) = g'(1) - f'(1) = 4 - (-2) = 6$

(iii) $(fg)'(1) = f(1)g'(1) + g(1)f'(1) = 3(4) + 7(-2) = 12 + (-14) = -2$

(iv) $\left(\frac{g}{f}\right)'(1) = \frac{f(1)g'(1) - g(1)f'(1)}{[f(1)]^2} = \frac{3(4) - 7(-2)}{3^2} = \frac{12+14}{9} = \frac{26}{9}$

(v) $\left(\frac{f}{g}\right)'(1) = \frac{g(1)f'(1) - f(1)g'(1)}{[g(1)]^2} = \frac{7(-2) - 3(4)}{7^2} = \frac{-14-12}{49} = \frac{-26}{49}$

Example 2: If $f(x) = x^4 - 3x^3 + 5x^2 - 7x + 11$, find $f'(x)$.

$$f'(x) = 4x^3 - 3(3x^2) + 5(2x) - 7(1) + 0 = 4x^3 - 9x^2 + 10x - 7$$

Example 3: If $f(x) = 4\sqrt{x} - \frac{3}{\sqrt[3]{x^2}} + \frac{5}{x} - \frac{7}{x^5}$, then find $f'(x)$.

$$f(x)=4\sqrt{x}-\frac{3}{\sqrt[3]{x^2}}+\frac{5}{x}-\frac{7}{x^5}=4x^{1/2}-3x^{-2/3}+5x^{-1}-7x^{-5}\Rightarrow$$

$$f'(x)=4\left(\tfrac{1}{2}x^{-1/2}\right)-3\left(-\tfrac{2}{3}x^{-5/3}\right)+5\left(-1x^{-2}\right)-7\left(-5x^{-6}\right)=$$

$$2x^{-1/2}+2x^{-5/3}-5x^{-2}+35x^{-6}=\frac{2}{\sqrt{x}}+\frac{2}{\sqrt[3]{x^5}}-\frac{5}{x^2}+\frac{35}{x^6}$$

Example 4: If $f(x)=\dfrac{x^2+2x-3}{3x-4}$, then find $f'(1)$.

$$f'(x)=\frac{(3x-4)(2x+2)-(x^2+2x-3)(3)}{(3x-4)^2}=\frac{6x^2-2x-8-3x^2-6x+9}{(3x-4)^2}=$$

3

$$\frac{3x^2-8x+1}{(3x-4)^2}\Rightarrow f'(1)=\frac{3(1)^2-8(1)+1}{[3(1)-4]^2}=\frac{-4}{1}=-4 \text{ or}$$

$$f'(1)=\frac{[3(1)-4][2(1)+2]-[1^2+2(1)-3](3)}{[3(1)-4]^2}=\frac{(-1)(4)-(0)(3)}{(-1)^2}=\frac{-4}{1}=-4$$

B. Trigonometric functions

(1) $f(x)=\sin x$

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{\sin(x+h)-\sin x}{h}=$$

$$\lim_{h\to 0}\frac{\sin x\cosh+\cos x\sinh-\sin x}{h}=\lim_{h\to 0}\frac{\sin x(\cosh-1)+\cos x\sinh}{h}=$$

$$(\sin x)\left[\lim_{h\to 0}\frac{\cos h-1}{h}\right]+(\cos x)\left[\lim_{h\to 0}\frac{\sin h}{h}\right]=(\sin x)(0)+(\cos x)(1)=$$

$$\cos x$$

(2) $f(x)=\cos x$

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{\cos(x+h)-\cos x}{h}=$$

$$\lim_{h\to 0}\frac{\cos x\cos h-\sin x\sin h-\cos x}{h}=\lim_{h\to 0}\frac{\cos x(\cos h-1)-\sin x\sin h}{h}=$$

$$(\cos x)\left[\lim_{h\to 0}\frac{\cos h-1}{h}\right]-(\sin x)\left[\lim_{h\to 0}\frac{\sin h}{h}\right]=(\cos x)(0)-(\sin x)(1)=$$

$$-\sin x$$

(3) $f(x)=\tan x=\dfrac{\sin x}{\cos x}$

$$f'(x)=\frac{(\cos x)(\cos x)-(\sin x)(-\sin x)}{(\cos x)^2}=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}=\frac{1}{\cos^2 x}=\sec^2 x$$

(4) $f(x)=\sec x=\dfrac{1}{\cos x}$

$$f'(x)=\frac{(\cos x)(0)-1(-\sin x)}{(\cos x)^2}=\frac{\sin x}{\cos^2 x}=\frac{1}{\cos x}\cdot\frac{\sin x}{\cos x}=\sec x\tan x$$

(5) $f(x)=\csc x=\dfrac{1}{\sin x}$

$$f'(x) = \frac{(\sin x)(0) - 1(\cos x)}{(\sin x)^2} = \frac{-\cos x}{\sin^2 x} = \frac{-1}{\sin x} \cdot \frac{\cos x}{\sin x} = -\csc x \cot x$$

(6) $f(x) = \cot x = \dfrac{\cos x}{\sin x}$

$$f'(x) = \frac{(\sin x)(\sin x) - (\cos x)(\cos x)}{(\sin x)^2} = \frac{-\cos^2 x - \sin^2 x}{\sin^2 x} = \frac{-1}{\sin^2 x} = -\csc^2 x$$

C. Composition and the generalized derivative rules

(1) $f(x) = (g \circ k)(x) = g(k(x))$

$$f'(x) = \lim_{b \to x} \frac{f(b) - f(x)}{b - x} = \lim_{b \to x} \frac{g(k(b)) - g(k(x))}{b - x} = \lim_{b \to x} \frac{g(k(b)) - g(k(x))}{b - x} \cdot$$

$$\frac{k(b) - k(x)}{k(b) - k(x)} = \lim_{b \to x} \frac{g(k(b)) - g(k(x))}{k(b) - k(x)} \cdot \lim_{b \to x} \frac{k(b) - k(x)}{b - x} =$$

$$\lim_{k(b) \to k(x)} \frac{g(k(b)) - g(k(x))}{k(b) - k(x)} \cdot \lim_{b \to x} \frac{k(b) - k(x)}{b - x} = g'(k(x)) \cdot k'(x).$$

This derivative rule for the composition of functions is called the Chain Rule.

(2) Suppose that $f(x) = g(k(x))$ where $g(x) = x^n$. Then $f(x) = [k(x)]^n$.

$g(x) = x^n \Rightarrow g'(x) = nx^{n-1} \Rightarrow g'(k(x)) = n[k(x)]^{n-1}$. Thus, $f'(x) =$ $g'(k(x)) \cdot k'(x) = n[k(x)]^{n-1} \cdot k'(x)$. This derivative rule for the power of a function is called the Generalized Power Rule.

(3) Suppose that $f(x) = g(k(x))$ where $g(x) = \sin x$. Then $f(x) = \sin[k(x)]$.

$g(x) = \sin x \Rightarrow g'(x) = \cos x \Rightarrow g'(k(x)) = \cos[k(x)]$. Thus, $f'(x) =$

$g'(k(x)) \cdot k'(x) = \cos[k(x)] \cdot k'(x)$.

(4) Similarly, if $f(x) = \cos[k(x)]$, then $f'(x) = -\sin[k(x)] \cdot k'(x)$.

(5) If $f(x) = \tan[k(x)]$, then $f'(x) = \sec^2[k(x)] \cdot k'(x)$.

(6) If $f(x) = \sec[k(x)]$, then $f'(x) = \sec[k(x)]\tan[k(x)] \cdot k'(x)$.

(7) If $f(x) = \cot[k(x)]$, then $f'(x) = -\csc^2[k(x)] \cdot k'(x)$.

(8) If $f(x) = \csc[k(x)]$, then $f'(x) = -\csc[k(x)]\cot[k(x)] \cdot k'(x)$.

Example 1: Suppose f and g are differentiable functions such that:

$f(1) = 9$	$f(2) = -5$	$g(1) = 2$	$g(9) = 3$
$f'(1) = -2$	$f'(2) = -6$	$g'(1) = 4$	$g'(9) = 7$

Find each of the following: (i) $(f \circ g)'(1)$; (ii) $(g \circ f)'(1)$; (iii) $h'(1)$ if $h(x) = \sqrt{f(x)}$; (iv) $j'(1)$ if $j(x) = [g(x)]^5$; (v) $l'(1)$ if $l(x) = \dfrac{3}{[f(x)]^2}$; (vi) $s'(1)$ if $s(x) = \sin[f(x)]$; and (vii) $m'(1)$ if $m(x) = \sec[g(x)]$.

(i) $(f \circ g)'(1) = f'(g(1)) \cdot g'(1) = f'(2) \cdot g'(1) = (-6)(4) = -24$

(ii) $(g \circ f)'(1) = g'(f(1)) \cdot f'(1) = g'(9) \cdot f'(1) = 7(-2) = -14$

(iii) $h(x) = \sqrt{f(x)} = [f(x)]^{1/2} \Rightarrow h'(x) = \frac{1}{2}[f(x)]^{-1/2} \cdot f'(x) = \dfrac{f'(x)}{2\sqrt{f(x)}} \Rightarrow$

$$h'(1) = \frac{f'(1)}{2\sqrt{f(1)}} = \frac{-2}{2\sqrt{9}} = -\frac{1}{3}$$

(iv) $j(x) = [g(x)]^5 \Rightarrow j'(x) = 5[g(x)]^4 \cdot g'(x) \Rightarrow j'(1) = 5[g(1)]^4 \cdot g'(1) =$

$5(2)^4(4) = 320$

(v) $l(x) = \dfrac{3}{[f(x)]^2} = 3[f(x)]^{-2} \Rightarrow l'(x) = -6[f(x)]^{-3} \cdot f'(x) \Rightarrow l'(1) =$

$\dfrac{-6f'(1)}{[f(1)]^3} = \dfrac{-6(-2)}{9^3} = \dfrac{12}{729} = \dfrac{4}{243}$

(vi) $s'(x) = \cos[f(x)] \cdot f'(x) \Rightarrow s'(1) = \cos[f(1)] \cdot f'(1) = \cos(9) \cdot (-2) = -2\cos 9$

(vii) $m'(x) = \sec[g(x)]\tan[g(x)] \cdot g'(x) \Rightarrow m'(1) = \sec[g(1)]\tan[g(1)] \cdot g'(1) =$

$\sec(2)\tan(2) \cdot 4 = 4\sec 2 \tan 2$

Example 2: If $f(x) = \sqrt[3]{2x^4 - x^2 + 5x + 2}$, then find $f'(1)$.

$f(x) = \sqrt[3]{2x^4 - x^2 + 5x + 2} = (2x^4 - x^2 + 5x + 2)^{1/3} \Rightarrow f'(x) =$

$\frac{1}{3}(2x^4 - x^2 + 5x + 2)^{-2/3}(8x^3 - 2x + 5) = \dfrac{8x^3 - 2x + 5}{3\sqrt[3]{(2x^4 - x^2 + 5x + 2)^2}} \Rightarrow$

$f'(1) = \dfrac{8 - 2 + 5}{3\sqrt[3]{(2 - 1 + 5 + 2)^2}} = \dfrac{11}{3\sqrt[3]{64}} = \dfrac{11}{12}$

Example 3: If $g(x) = \dfrac{4}{(x^3 + 4)^8}$, then find $g'(x)$.

$g(x) = \dfrac{4}{(x^3 + 4)^8} = 4(x^3 + 4)^{-8} \Rightarrow g'(x) = -32(x^3 + 4)^{-9}(3x^2) = \dfrac{-96x^2}{(x^3 + 4)^9}$

Example 4: If $h(x) = \sin(\cos x)$, then find $h'(x)$.

$h'(x) = \cos(\cos x) \cdot (-\sin x)$

Example 5: If $j(x) = \tan(2x^2 - 3x + 1)$, then find $j'(x)$.

$j'(x) = \sec^2(2x^2 - 3x + 1) \cdot (4x - 3)$

Example 6: If $k(x) = x^2\sqrt{3x + 4}$, then find $k'(x)$.

$k(x) = x^2\sqrt{3x + 4} = x^2(3x + 4)^{1/2} \Rightarrow k'(x) = x^2\left[\frac{1}{2}(3x + 4)^{-1/2}(3)\right] +$

$$(3x+4)^{1/2}(2x) = \frac{3x^2}{2(3x+4)^{1/2}} + \frac{2x(3x+4)^{1/2}}{1} = \frac{3x^2+4x(3x+4)}{2(3x+4)^{1/2}} =$$

$$\frac{15x^2+16x}{2(3x+4)^{1/2}}$$

Example 7: If $l(x) = \left(\frac{2x-1}{3x+4}\right)^4$, then find $l'(x)$.

$$l'(x) = 4\left(\frac{2x-1}{3x+4}\right)^3\left[\frac{(3x+4)(2)-(2x-1)(3)}{(3x+4)^2}\right] = \frac{4(2x-1)^3}{(3x+4)^3}\left[\frac{11}{(3x+4)^2}\right] =$$

$$\frac{44(2x-1)^3}{(3x+4)^5}.$$

Example 8: If $k(x) = \frac{\sin x}{1+\cos x}$, then find $k'(x)$.

$$k'(x) = \frac{(1+\cos x)(\cos x)-(\sin x)(-\sin x)}{(1+\cos x)^2} = \frac{\cos x+\cos^2 x+\sin^2 x}{(1+\cos x)^2} =$$

$$\frac{\cos x+1}{(1+\cos x)^2} = \frac{1}{1+\cos x}.$$

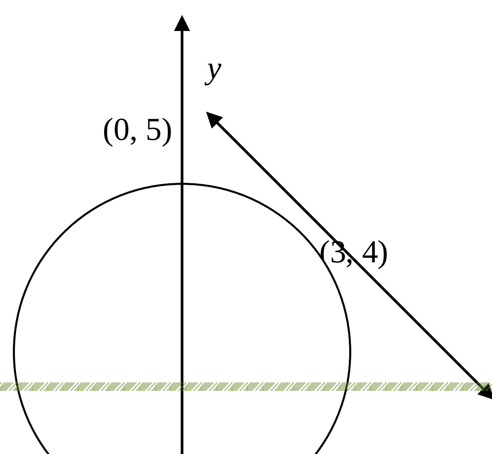
Example 9: If $s(x) = \sin^3(x^2-1)$, then find $s'(x)$.

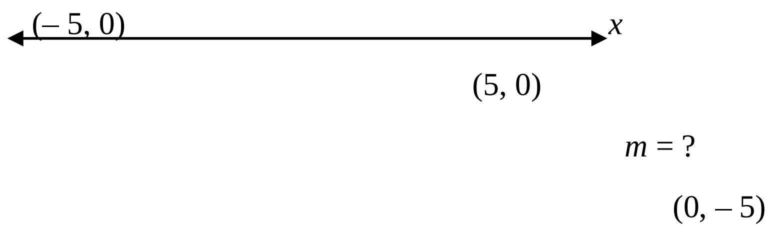
$$s(x) = \sin^3(x^2-1) = [\sin(x^2-1)]^3 \Rightarrow s'(x) = 3[\sin(x^2-1)]^2 \cdot \cos(x^2-1) \cdot 2x =$$

$$6x\sin^2(x^2-1)\cos(x^2-1).$$

III. Implicit Differentiation

Example 1: Find the slope of the tangent line to the circle $x^2+y^2=25$ at the point (3, 4).





Solution 1 : A circle is not a function. However, $x^2 + y^2 = 25 \Rightarrow y^2 =$

$25 - x^2 \Rightarrow y = \pm\sqrt{25 - x^2} \Rightarrow y = \sqrt{25 - x^2}$ is the equation of the upper half circle and $y = -\sqrt{25 - x^2}$ is the equation of the lower half circle.

Since the point (3, 4) is on the upper half circle, use the function $f(x) =$

$$\sqrt{25 - x^2} = \left(25 - x^2\right)^{1/2} \Rightarrow f'(x) = \frac{1}{2}\left(25 - x^2\right)^{-1/2}(-2x) = \frac{-x}{\sqrt{25 - x^2}} \Rightarrow$$

$$m = f'(3) = \frac{-3}{\sqrt{25 - 3^3}} = \frac{-3}{\sqrt{25 - 9}} = \frac{-3}{\sqrt{16}} = -\frac{3}{4}.$$

Sometimes, an equation [$x^2 + y^2 = 25$] in two variables, say x and y, is given, but it is not in the form of $y = f(x)$. In this case, for each value of one of the variables, one or more values of the other variable may exist. Thus, such an equation may describe one or more functions [$y = \sqrt{25 - x^2}$ and $y = -\sqrt{25 - x^2}$]. Any function defined in this manner is said to be defined implicitly. For such equations, we may not be able to solve for y explicitly in terms of x [in the example, I was able to solve for y explicitly in terms of x]. In fact, there are applications where it is not essential to obtain a formula for y in terms of x. Instead, the value of the derivative at certain points must be obtained. It is possible to accomplish this goal by using a technique called implicit differentiation. Suppose an equation in two variables, say x and y, is given and we are told that this equation defines a differentiable function f with $y = f(x)$. Use the following steps to differentiate implicitly:

(1) Simplify the equation if possible. That is, get rid of parentheses by multiplying using the distributive property or by redefining subtraction, and

clear fractions by multiplying every term of the equation by a common denominator for all the fractions; simplify and combine like terms.

(2) Differentiate both sides of the equation with respect to x. Use all the relevant differentiation rules, being careful to use the Chain Rule when differentiating expressions involving y.

(3) Solve for $\frac{dy}{dx}$.

Note: It might be helpful to substitute $f(x)$ into the equation for y before differentiating with respect to x. This will remind you when you must use the generalized forms of the Chain Rule. Since $f'(x) = \frac{dy}{dx}$, you differentiate with respect to x and substitute y for $f(x)$ and $\frac{dy}{dx}$ for $f'(x)$. Then you can solve for $\frac{dy}{dx}$.

Solution 2: $x^2 + y^2 = 25 \Rightarrow x^2 + [f(x)]^2 = 25 \Rightarrow \frac{d}{dx}\left(x^2 + [f(x)]^2 = 25\right) \Rightarrow$

$$2x + 2[f(x)]f'(x) = 0 \Rightarrow f'(x) = \frac{-2x}{2[f(x)]} \Rightarrow \frac{dy}{dx} = \frac{-x}{y} \Rightarrow \left.\frac{dy}{dx}\right|_{\substack{x=3 \\ y=4}} = -\frac{3}{4}.$$

Example 2: Suppose that the equation $\frac{2}{x} + \frac{3}{y} = x$ defines a function f with $y = f(x)$. Find $\frac{dy}{dx}$ and the slope of the tangent line at the point (2, 3).

Solution 1: Solve for y. $xy\left(\frac{2}{x} + \frac{3}{y}\right) = xy(x) \Rightarrow 2y + 3x = x^2 y \Rightarrow y = \frac{3x}{x^2 - 2} \Rightarrow$

$$\frac{dy}{dx}=\frac{(x^2-2)(3)-3x(2x)}{(x^2-2)^2}=\frac{-3x^2-6}{(x^2-2)^2}\Rightarrow \left.\frac{dy}{dx}\right|_{x=2}=\frac{-18}{4}=-\frac{9}{2}$$

Solution 2: Clear fractions $\Rightarrow 2y+3x=x^2y \Rightarrow \frac{d}{dx}\left(2y+3x=x^2y\right)\Rightarrow$

$$2\frac{dy}{dx}+3=x^2\frac{dy}{dx}+2xy \Rightarrow \frac{dy}{dx}=\frac{3-2xy}{x^2-2}\Rightarrow \left.\frac{dy}{dx}\right|_{\substack{x=2\\y=3}}=\frac{3-12}{2}=-\frac{9}{2}$$

Solution 3: $\frac{d}{dx}\left(\frac{2}{x}+\frac{3}{y}=x\right)\Rightarrow \frac{d}{dx}\left(2x^{-1}+3y^{-1}=x\right)\Rightarrow -2x^{-2}-3y^{-2}\frac{dy}{dx}=1\Rightarrow$

$$\frac{-2}{x^2}-\frac{3}{y^2}\frac{dy}{dx}=1\Rightarrow -2y^2-3x^2\frac{dy}{dx}=x^2y^2 \Rightarrow \frac{dy}{dx}=\frac{-2y^2-x^2y^2}{3x^2}\Rightarrow$$

$$\left.\frac{dy}{dx}\right|_{\substack{x=2\\y=3}}=\frac{-18-36}{12}=\frac{-54}{12}=-\frac{9}{2}$$

Example 3: If $\cos(xy)=y$, then find $\frac{dy}{dx}$.

$$\frac{d}{dx}\left(\cos(xy)=y\right)\Rightarrow -\sin(xy)\left[x\frac{dy}{dx}+y(1)\right]=\frac{dy}{dx}\Rightarrow -x\sin(xy)\frac{dy}{dx}-y\sin(xy)=$$

$$\frac{dy}{dx}\Rightarrow -y\sin(xy)=\frac{dy}{dx}\left(1+x\sin(xy)\right)\Rightarrow \frac{dy}{dx}=\frac{-y\sin(xy)}{1+x\sin(xy)}$$

IV. Higher Order Derivatives

A. Notation

(1) 1st derivative (derivative of the original function $y=f(x)$): $\frac{dy}{dx}=f'(x)$

(2) 2nd derivative (derivative of the 1st derivative): $\frac{d^2y}{dx^2}=f''(x)$

(3) 3rd derivative (derivative of the 2nd derivative): $\dfrac{d^3y}{dx^3} = f'''(x)$

B. Distance functions

Suppose $s(t)$ is a distance function with respect to time t. Then $s'(t) = v(t)$ is an instantaneous velocity (or velocity) function with respect to time t, and $s''(t) = v'(t) = a(t)$ is an acceleration function with respect to time t.

Example 1: If $f(x) = x^2 \sin x$, then find $f'(x)$ and $f''(x)$.

$$f'(x) = x^2 \cos x + 2x \sin x$$

$$f''(x) = x^2(-\sin x) + 2x\cos x + 2x\cos x + 2\sin x = -x^2 \sin x + 4x\cos x + 2\sin x$$

Example 2: If $g(x) = \dfrac{2x+3}{4x-5}$, then find $g'(x)$ and $g''(x)$.

$$g'(x) = \frac{(4x-5)(2)-(2x+3)(4)}{(4x-5)^2} = \frac{8x-10-8x-12}{(4x-5)^2} = \frac{-22}{(4x-5)^2} = -22(4x-5)^{-2}$$

$$g''(x) = 44(4x-5)^{-3}(4) = 176(4x-5)^{-3} = \frac{176}{(4x-5)^3}$$

Example 3: If $x^2 + y^2 = 25$, then find $\dfrac{dy}{dx}$ and $\dfrac{d^2y}{dx^2}$.

$$\frac{d}{dx}\left(x^2 + y^2 = 25\right) \Rightarrow 2x + 2y\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = \frac{-2x}{2y} = \frac{-x}{y}$$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{-x}{y}\right) = \frac{y(-1)-(-x)\left(\frac{dy}{dx}\right)}{y^2} = \frac{-y+x\left(\frac{-x}{y}\right)}{y^2} = \frac{-y^2-x^2}{y^3} =$$

$$\frac{-(x^2+y^2)}{y^3}=\frac{-25}{y^3}$$

Practice Sheet – Techniques of Differentiation

I. Find the derivative of each function defined as follows; there is no need to simplify your answers.

(1) $f(x)=x^4-5x^3+9x^2-7x+5$

(2) $y=\frac{9}{x^2}-\frac{8}{x^3}+\frac{2}{x^4}$

(3) $g(x)=8\sqrt{x}-\frac{6}{\sqrt[3]{x^2}}$

(4) $y=\frac{3x^2-6x}{x^3}$

(5) $h(x)=\frac{3x+2}{x^2+1}$

(6) $y=x^2\cos x$

(7) $f(x)=\frac{\sin x}{x}$

(8) $y=\sqrt[3]{x^2-3x+4}$

(9) $g(x)=\sin(\sqrt{x})$

(10) $y=\cos^3 x$

(11) $h(x)=\sqrt{\frac{2x+1}{3x-4}}$

(12) $y=\frac{\sec x}{1+\tan x}$

(13) $k(x)=x\sqrt{9-x^2}$

(14) $y=\sin(3x)\cos(4x)$

(15) $f(x)=\tan^4\left(x^3\right)$

(16) $y=\frac{\sqrt{x}+1}{\sqrt{x}-1}$

(17) $g(x) = x\sec\left(\frac{1}{x}\right)$

(18) $y = \sqrt{1+\sin 2x}$

II. Find $\frac{dy}{dx}$ by implicit differentiation for each of the following:

(1) $-3xy - 4y^2 = 2$

(2) $8x^2 = 2y^3 + 3xy^2$

(3) $\frac{3}{2x} + \frac{1}{y} = y$

(4) $3x^2 = \frac{2-y}{2+y}$

12

(5) $x = \tan y$

(6) $y = \cos(x-y)$

(7) $x\sin y + y\sin x = 1$

(8) $x = \sec^3(y^2 - 1)$

III. Find the slope of the tangent line at the given point on each curve defined by the given equation:

(1) $x^2 + 3y^2 = 21$; $(3, -2)$

(2) $x^3 + \sqrt[3]{y} = 3$; $(1, 8)$

(3) $\sqrt{xy} - y = -2$; $(1, 4)$

(4) $3xy - 2x^4 = y^3 - 23$; $(2, -3)$

(5) $x = \cos y$; $\left(\frac{1}{2}, \frac{-\pi}{3}\right)$

(6) $\sin(xy) = x$; $\left(1, \frac{\pi}{2}\right)$

IV. For each of the following functions $f(x)$, find $f'(x)$ and $f''(x)$.

(1) $f(x) = 3x^4 - 4x^2 + 7x - 11$

(2) $f(x) = \dfrac{3x+1}{2x-1}$

(3) $f(x) = x^3 \cos(4x)$

(4) $f(x) = \sin^4 x$

V. Suppose the distance (in feet) that an object travels in t seconds is given by the formula $s(t) = 2t^3 + 4t - 5$. Find $s(2)$, $v(2)$, and $a(2)$.

Solution Key for Techniques of Differentiation

I. (1) $f'(x) = 4x^3 - 15x^2 - 18x - 7$

(2) $y = 9x^{-2} - 8x^{-3} + 2x^{-4} \Rightarrow \dfrac{dy}{dx} = -18x^{-3} + 24x^{-4} - 8x^{-5}$

(3) $g(x) = 8x^{1/2} - 6x^{-2/3} \Rightarrow g'(x) = 4x^{-1/2} + 4x^{-5/3}$

(4) $y = 3x^{-1} - 6x^{-2} \Rightarrow \dfrac{dy}{dx} = -3x^{-2} + 12x^{-3}$

(5) $h'(x) = \dfrac{(x^2+1)(3) - (3x+2)(2x)}{(x^2+1)^2}$

(6) $\dfrac{dy}{dx} = x^2(-\sin x) + (\cos x)(2x)$

(7) $f'(x)=\dfrac{x(\cos x)-(\sin\ x)(1)}{x^2}$

(8) $y=\left(x^2-3x+4\right)^{1/3}\Rightarrow\dfrac{dy}{dx}=\frac{1}{3}\left(x^2-3x+4\right)^{-2/3}(2x-3)$

(9) $g'(x)=\cos\left(\sqrt{x}\right)\cdot\left(\frac{1}{2}x^{-1/2}\right)$

(10) $y=(\cos x)^3\Rightarrow\dfrac{dy}{dx}=3(\cos x)^2(-\sin\ x)$

(11) $h(x)=\left(\dfrac{2x+1}{3x-4}\right)^{1/2}\Rightarrow h'(x)=\frac{1}{2}\left(\dfrac{2x+1}{3x-4}\right)^{-1/2}\left[\dfrac{(3x-4)(2)-(2x+1)(3)}{(3x-4)^2}\right]$

(12) $\dfrac{dy}{dx}=\dfrac{(1+\tan x)(\sec x\tan x)-\sec x(\sec^2 x)}{(1+\tan x)^2}$

(13) $k(x)=x\left(9-x^2\right)^{1/2}\Rightarrow k'(x)=x\left[\frac{1}{2}\left(9-x^2\right)^{-1/2}(-2x)\right]+\left(9-x^2\right)^{1/2}(1)$

(14) $\dfrac{dy}{dx}=\sin(3x)[-\sin(4x)(4)]+\cos(4x)[\cos(3x)(3)]$

(15) $f(x)=\left[\tan\left(x^3\right)\right]^4\Rightarrow f'(x)=4\left[\tan\left(x^3\right)\right]^3\left[\sec^2\left(x^3\right)\right]\left(3x^2\right)$

(16) $\dfrac{dy}{dx}=\dfrac{\left(\sqrt{x}-1\right)\left(\dfrac{1}{2\sqrt{x}}\right)-\left(\sqrt{x}+1\right)\left(\dfrac{1}{2\sqrt{x}}\right)}{\left(\sqrt{x}-1\right)^2}$ $\quad\left[\dfrac{d}{dx}\left(\sqrt{x}\right)=\dfrac{1}{2\sqrt{x}}\right]$

(17) $g'(x)=x\left[\sec\left(\dfrac{1}{x}\right)\tan\left(\dfrac{1}{x}\right)\left(-\dfrac{1}{x^2}\right)\right]+\sec\left(\dfrac{1}{x}\right)(1)$

(18) $y=(1+\sin\ 2x)^{1/2}\Rightarrow\dfrac{dy}{dx}=\frac{1}{2}(1+\sin\ 2x)^{-1/2}(\cos 2x)(2)$

II. (1) $-3x\frac{dy}{dx}-3y-8y\frac{dy}{dx}=0 \Rightarrow \frac{dy}{dx}=\frac{-3y}{3x+8y}$

(2) $16x=6y^2\frac{dy}{dx}+6xy\frac{dy}{dx}+3y^2 \Rightarrow \frac{dy}{dx}=\frac{16x-3y^2}{6y^2+6xy}$

(3) $\frac{3}{2x}+\frac{1}{y}=y \Rightarrow 3y+2x=2xy^2 \Rightarrow 3\frac{dy}{dx}+2=4xy\frac{dy}{dx}+2y^2 \Rightarrow \frac{dy}{dx}=\frac{2-2y^2}{4xy-3}$

(4) $6x^2+3x^2y=2-y \Rightarrow 12x+3x^2\frac{dy}{dx}+6xy=-\frac{dy}{dx} \Rightarrow \frac{dy}{dx}=\frac{-12x-6xy}{3x^2+1}$

(5) $1=\left(\sec^2 y\right)\frac{dy}{dx} \Rightarrow \frac{dy}{dx}=\frac{1}{\sec^2 y}=\cos^2 y$ or $\frac{dy}{dx}=\frac{1}{1+\tan^2 y}=\frac{1}{1+x^2}$

(6) $\frac{dy}{dx}=-\sin(x-y)\left[1-\frac{dy}{dx}\right] \Rightarrow \frac{dy}{dx}=\frac{-\sin(x-y)}{1-\sin(x-y)}$

(7) $x(\cos y)\frac{dy}{dx}+\sin y+y\cos x+(\sin x)\frac{dy}{dx}=0 \Rightarrow \frac{dy}{dx}=\frac{-\sin y-y\cos x}{x\cos y+\sin x}$

(8) $1=3\sec^2(y^2-1)\left[\sec(y^2-1)\tan(y^2-1)\right]\left(2y\frac{dy}{dx}\right) \Rightarrow \frac{dy}{dx}=$

$$\frac{1}{6y\sec^3(y^2-1)\tan(y^2-1)}=\frac{1}{6xy\tan(y^2-1)}$$

III. (1) $2x+6y\frac{dy}{dx}=0 \Rightarrow 2(3)+6(-2)\frac{dy}{dx}=0 \Rightarrow 6-12\frac{dy}{dx}=0 \Rightarrow \frac{dy}{dx}=\frac{1}{2}$

(2) $3x^2 + \left(\frac{1}{3} y^{-\frac{2}{3}}\right)\frac{dy}{dx} = 0 \Rightarrow 3x^2 + \left(\frac{1}{3\sqrt[3]{y^2}}\right)\frac{dy}{dx} = 0 \Rightarrow 3(1)^2 + \left(\frac{1}{3\sqrt[3]{8^2}}\right)\frac{dy}{dx} = 0 \Rightarrow$

$3 + \frac{1}{12}\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = -36$

(3) $xy = (y-2)^2 \Rightarrow x\frac{dy}{dx} + y = 2(y-2)\frac{dy}{dx} \Rightarrow 1\frac{dy}{dx} + 4 = 4\frac{dy}{dx} \Rightarrow \frac{dy}{dx} = \frac{4}{3}$

(4) $3x\frac{dy}{dx} + 3y - 8x^3 = 3y^2\frac{dy}{dx} \Rightarrow 3(2)\frac{dy}{dx} + 3(-3) - 8(2^3) = 3(-3)^2\frac{dy}{dx} \Rightarrow$

$6\frac{dy}{dx} - 9 - 64 = 27\frac{dy}{dx} \Rightarrow \frac{dy}{dx} = \frac{-73}{21}$

(5) $1 = (-\sin\, y)\frac{dy}{dx} \Rightarrow 1 = \left[-\sin\left(\frac{-\pi}{3}\right)\right]\frac{dy}{dx} \Rightarrow 1 = \frac{\sqrt{3}}{2}\frac{dy}{dx} \Rightarrow \frac{dy}{dx} = \frac{2}{\sqrt{3}}$

(6) $\cos(xy)\left[x\frac{dy}{dx} + y\right] = 1 \Rightarrow \frac{dy}{dx} = \frac{1 - y\cos(xy)}{x\cos(xy)} \Rightarrow \frac{dy}{dx}\cdot_{\substack{x=1\\ y=\frac{\pi}{2}}} = \frac{1 - \frac{\pi}{2}\cos\left(\frac{\pi}{2}\right)}{1\cos\left(\frac{\pi}{2}\right)} =$

$\frac{1}{0} \Rightarrow \frac{dy}{dx}$ does not exist.

IV. (1) $f'(x) = 12x^3 - 8x + 7$ and $f''(x) = 36x^2 - 8$

(2) $f'(x) = \frac{(2x-1)(3) - (3x+1)(2)}{(2x-1)^2} = \frac{-5}{(2x-1)^2} = -5(2x-1)^{-2}$ and

$$f''(x) = 10(2x-1)^{-3}(2) = \frac{20}{(2x-1)^3}$$

(3) $f'(x) = x^3[-4\sin(4x)] + 3x^2\cos(4x) = -4x^3\sin(4x) + 3x^2\cos(4x)$ and

$$f''(x) = -4x^3[4\cos(4x)] - 12x^2\sin(4x) + 3x^2[-4\sin(4x)] + 6x\cos(4x) =$$

$$-16x^3\cos(4x) - 24x^2\sin(4x) + 6x\cos(4x)$$

(4) $f'(x) = 4(\sin x)^3\cos x$ and $f''(x) = 4(\sin x)^3(-\sin x) + (\cos x)\left(12(\sin x)^2(\cos x)\right) =$

$$-4\sin^4 x + 12\sin^2 x\cos^2 x$$

V. $s(2) = 2\left(2^3\right) + 4(2) - 5 = 16 + 8 - 5 = 19\ ft$

$v(t) = s'(t) = 6t^2 + 4 \Rightarrow v(2) = 6\left(2^2\right) + 4 = 28\ ft/sec$

$a(t) = v'(t) = s''(t) = 12t \Rightarrow a(2) = 12(2) = 24\ \ ft/\sec^2$

This is one of a series of worksheets designed to help you increase your confidence in handling Mathematics. This worksheet contains both theory and exercises which cover:-

1. Exponential functions
2. Logarithmic functions
3. Implicit Differentiation
4. Logarithmic Differentiation
5. Parametric Equations

1. Exponential Functions

It can be shown that e^x is the function such that $\frac{d\left(e^x\right)}{dx} = e^x$

Examples

Differentiate the following (i) $y = e^{2x}$ (ii) $y = e^{f(x)}$ (iii) $y = \dfrac{1+x}{e^{2x}}$

(i) $y = e^{2x}$ is a function of a function

Writing $y = e^u$ where $u = 2x \Rightarrow \dfrac{dy}{du} = e^u$ and $\dfrac{du}{dx} = 2$

giving $\dfrac{dy}{dx} = \dfrac{dy}{du} \times \dfrac{du}{dx} = e^u \times 2 = 2e^{2x}$

(ii) $y = e^{f(x)}$ is a function of a function

Writing $y = e^u$ where $u = f(x) \Rightarrow \dfrac{dy}{du} = e^u$ and $\dfrac{du}{dx} = f'(x)$

giving $\dfrac{dy}{dx} = \dfrac{dy}{du} \times \dfrac{du}{dx} = e^u \times f'(x) = f'(x)e^{f(x)}$

Remembering $\boxed{\dfrac{d\left(e^{f(x)}\right)}{dx} = f'(x)e^{f(x)}}$ will enable you to differentiate most exponential functions!

(iii) $y = \dfrac{1+x}{e^{2x}}$ is a quotient

writing $u = 1+x,\ v = e^{2x}$ gives $u' = 1,\ v' = 2e^{2x}$

using $\dfrac{dy}{dx} = \dfrac{u'v - uv'}{v^2}$

gives $\dfrac{dy}{dx} = \dfrac{1 \times e^{2x} - (1+x) \times 2e^{2x}}{\left(e^{2x}\right)^2}$

$$= \frac{(1-2-2x)e^{2x}}{e^{4x}} = \frac{-(1+2x)}{e^{2x}}$$

Exercise 1

Differentiate the following

1. e^{7x} 2. $e^{\cos x}$ 3. $2e^{2x}+3e^{x^2}-e$ 4. $e^{(\sin x+\cos x)}$

5. x^2e^x 6. $\dfrac{e^x+1}{x^3}$ 7. $\dfrac{(e^x-1)}{e^x}$ 8. $\dfrac{e^{x^2}}{x}$

2. Logarithmic functions

The inverse function of e^x is $\log_e x$ which is usually written as $\ln x$ (shorthand for natural or Napierian logarithms after Napier who developed them). For more information see the logs booklet.

Given $y=\ln x$ then, from the definition of logarithms,

$$x=e^y \text{ which gives } \frac{dx}{dy}=e^y \Rightarrow \frac{dy}{dx}=\frac{1}{e^y}=\frac{1}{x}$$

hence $$\boxed{\frac{d(\ln x)}{dx}=\frac{1}{x}}$$

Extending this to differentiate $y=\ln[f(x)]$ which is a function ($\ln$) of the function $f(x)$.

write $y=\ln u$ where $u=f(x)\Rightarrow \dfrac{dy}{du}=\dfrac{1}{u}$ and $\dfrac{du}{dx}=f'(x)$

using the chain rule $\dfrac{dy}{dx}=\dfrac{dy}{du}\times\dfrac{du}{dx}$

giving $$\frac{dy}{dx}=\frac{1}{f(x)}\times f'(x)=\frac{f'(x)}{f(x)}$$

Another important result to learn $$\boxed{\frac{d}{dx}[\ln(f(x))]=\frac{f'(x)}{f(x)}}$$

Examples

Differentiate the following functions

(i) $y=\ln(5x^2-6)$ (ii) $y=\ln\left(\dfrac{x+2}{x+3}\right)$ (iii) $y=\ln\left(\dfrac{\sin^2 x}{\cos x}\right)$

(i) Using the above

$$y = \ln\left(5x^2 - 6\right) \text{ gives } \frac{dy}{dx} = \frac{10x}{5x^2 - 6}$$

(ii) Simplifying the expression gives

$$y = \ln\left(\frac{x+2}{x+3}\right) = \ln(x+2) - \ln(x+3)$$

$$\text{Hence} \quad \frac{dy}{dx} = \frac{1}{x+2} - \frac{1}{x+3}$$

$$= \frac{(x+3)-(x+2)}{(x+2)(x+3)} = \frac{1}{(x+2)(x+3)}$$

Note you could do this without simplifying but it is more difficult!

(iii) Simplifying the expression gives

$$y = \ln\left(\frac{\sin^2 x}{\cos x}\right) = \ln\left(\sin^2 x\right) - \ln(\cos x)$$

$$= 2\ln(\sin x) - \ln(\cos x)$$

$$\frac{dy}{dx} = 2\frac{\cos x}{\sin x} - \frac{-\sin x}{\cos x}$$

$$= \frac{2\cos^2 x + \sin^2 x}{\sin x \cos x} = \frac{1+\cos^2 x}{\sin x \cos x} \left(\text{using } \cos^2 x + \sin^2 x = 1\right)$$

Exercise 2

Differentiate the following

1. $\ln 7x$
2. $-\ln 6x$
3. $\ln\left(x^2+3\right)$
4. $\ln\left(x^2+2x-1\right)$
5. $\ln\left(\frac{5x}{2x-3}\right)$
6. $\ln\left(\frac{x^2}{1+x}\right)$
7. $\ln\left(\frac{\sin x}{x\cos x}\right)$
8. $\ln\left(\frac{(x-1)^2}{\sqrt{x+1}}\right)$
9. $\ln\left[\left(\frac{x+4}{\sqrt[3]{5x+2}}\right)^2\right]$

3. Implicit Functions

A function such as $y = x^5 + 3x^3 - x + 5$ is called an explicit function as y is explicitly given in terms of x.

A function such as $x^5 + 3x^3y - xy^2 + 5y - 3x = 15$ is called an implicit function as y is not given explicitly in terms of x nor x in terms of y.

An implicit function can be differentiated with respect to x as it stands.

Consider $\quad x^2 + 3y - y^2 + xy - 3x = 15$

Differentiating each term with respect to x we get:

$$\frac{d(x^2)}{dx} + \frac{d(3y)}{dx} - \frac{d(y^2)}{dx} + \frac{d(xy)}{dx} - \frac{d(3x)}{dx} = \frac{d(15)}{dx}$$

To differentiate a function of y with respect to x we need

to use the chain rule $\dfrac{d[f(y)]}{dx} = \dfrac{d[f(y)]}{dy} \times \dfrac{dy}{dx}$

giving $\dfrac{d(3y)}{dx} = \dfrac{3d(y)}{dy} \times \dfrac{dy}{dx} = 3\dfrac{dy}{dx}$ and $\dfrac{d(y^2)}{dx} = \dfrac{d(y^2)}{dy} \times \dfrac{dy}{dx} = 2y\dfrac{dy}{dx}$

using the product formula $\dfrac{d(xy)}{dx} = \dfrac{d(x)}{dx} \times \dfrac{y}{1} + \dfrac{x}{1} \times \dfrac{dy}{dx} = y + x\dfrac{dy}{dx}$

Putting these together we have:

$$\frac{d(x^2)}{dx} + \frac{d(3y)}{dx} - \frac{d(y^2)}{dx} + \frac{d(xy)}{dx} - \frac{d(3x)}{dx} = \frac{d(15)}{dx}$$

$$2x + 3\frac{dy}{dx} - 2y\frac{dy}{dx} + y + x\frac{dy}{dx} - 3 = 0$$

$$(3 - 2y + x)\frac{dy}{dx} = 3 - y - 2x$$

$$\frac{dy}{dx} = \frac{3 - y - 2x}{3 - 2y + x}$$

Example

Find the gradient of the curve $x^2 + y^2 + 2xy - 5x + 3y = 10$ at the points where $x = 1$

First we need to find the values of y when $x = 1$

Putting $x = 1$ we get $1 + y^2 + 2y - 5 + 3y = 10 \Rightarrow y^2 + 5y - 14 = 0$

which gives $(y-2)(y+7) = 0 \Rightarrow y = 2$ or $y = -7$

notice that there are two points to consider (1, 2) and (1, -7)

Differentiating the function $x^2 + y^2 + 2xy - 5x + 3y = 10$

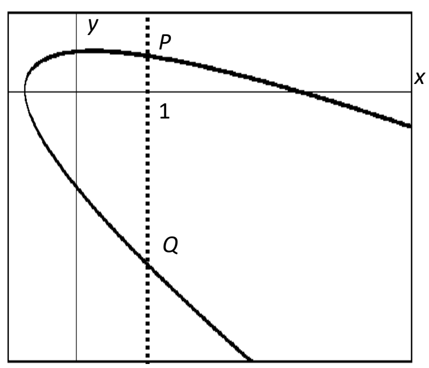
gives $\dfrac{d(x^2)}{dx} + \dfrac{d(y^2)}{dx} + \dfrac{d(2xy)}{dx} - \dfrac{d(5x)}{dx} + \dfrac{d(3y)}{dx} = \dfrac{d(10)}{dx}$

$$2x + 2y\frac{dy}{dx} + 2y + 2x\frac{dy}{dx} - 5 + 3\frac{dy}{dx} = 0$$

giving $\dfrac{dy}{dx} = \dfrac{5 - 2x - 2y}{2y + 2x + 3}$

at $P = (1, 2)$ $\quad \dfrac{dy}{dx} = \dfrac{5-2-4}{4+2+3} = -\dfrac{1}{9}$

at $Q = (1, -7)$ $\quad \dfrac{dy}{dx} = \dfrac{5-2+14}{-14+2+3} = -\dfrac{17}{9}$



Note you could substitute in and find the value of $\dfrac{dy}{dx}$ without making it the subject.

The sketch of the graph shows the two points *P* and *Q*. From the sketch you can see that the gradient is negative in both cases.

Exercise 3

1. In the following find $\dfrac{dy}{dx}$ in terms of x and y

(i) $x^2 + y^2 = 10$ (ii) $2x^2 + 2y^2 + 3x = 10 + 7y$

(iii) $x^2 - y^2 + 3xy = 6$ (iv) $2x^3 + 3xy^2 - y^3 = 0$

2. Find the gradient of the curve $x^2 + 6y^2 = 10$ at the points where $x = 2$.

3. Find the gradient of the curve $x^3 + 4xy = y^2 + 15$ at the points where $x = 2$.

4. Logarithmic Differentiation

The function $y = a^x$ cannot be differentiated by any of the methods developed so far. But taking the natural logarithm of both sides overcomes the problem!

To solve $\quad y = a^x$

take logs $\quad \ln y = \ln\left(a^x\right) = x \ln a$

differentiate $\quad \dfrac{d(\ln y)}{dx} = \dfrac{d(x \ln a)}{dx}$

By the chain rule the left hand side gives $\dfrac{d(\ln y)}{dx} = \dfrac{d(\ln y)}{dy}\dfrac{dy}{dx} = \dfrac{1}{y}\dfrac{dy}{dx}$

the right hand side gives $\quad \dfrac{d(x \ln a)}{dx} = \ln a \dfrac{d(x)}{dx} = \ln a$

putting these together gives $\quad \dfrac{1}{y}\dfrac{dy}{dx} = \ln a$

hence $\quad \dfrac{dy}{dx} = (\ln a) y = (\ln a) a^x$

This method can simplify differentiation in a number of cases, as shown in the following examples.

Examples (The first two could be differentiated as quotients.)

1. Find $\dfrac{dy}{dx}$ given the function $y = \dfrac{\sin x}{\cos x}$ (ie tan$x$)

Taking logs gives $\quad \ln y = \ln \sin x - \ln \cos x$

Differentiate $\frac{1}{y}\frac{dy}{dx} = \frac{\cos x}{\sin x} - \frac{-\sin x}{\cos x} = \frac{\cos^2 x + \sin^2 x}{\sin x \cos x} = \frac{1}{\sin x \cos x}$

$$\frac{dy}{dx} = \frac{1}{\sin x \cos x} \times y = \frac{1}{\sin x \cos x} \times \frac{\sin x}{\cos x} = \frac{1}{\cos^2 x} = \sec^2 x$$

The result should be known $\frac{d}{dx}(\tan x) = \sec^2 x$

2. Find $\frac{dy}{dx}$ given the function $y = \frac{x \sin x}{(x+1)\cos x}$

$$\ln(y) = \ln\left(\frac{x \sin x}{(x+1)\cos x}\right) = \ln(x \sin x) - \ln[(x+1)\cos x]$$

$$= \ln(x) + \ln(\sin x) - \ln(x+1) - \ln(\cos x)$$

Differentiating gives

$$\frac{1}{y}\frac{dy}{dx} = \frac{1}{x} + \frac{\cos x}{\sin x} - \frac{1}{x+1} + \frac{\sin x}{\cos x}$$

$$\frac{dy}{dx} = y\left[\frac{1}{x} + \frac{\cos x}{\sin x} - \frac{1}{x+1} + \frac{\sin x}{\cos x}\right]$$

$$= \frac{x \sin x}{(x+1)\cos x}\left[\frac{1}{x} + \frac{\cos x}{\sin x} - \frac{1}{x+1} + \frac{\sin x}{\cos x}\right]$$

which is a lot easier than using the quotient method. It could be 'simplified' but this rarely needs to be done.

3. Find $\frac{dy}{dx}$ given the function $y = x^x$

Take natural logs $\ln(y) = \ln(x^x) = x\ln(x)$

Differentiate $\frac{1}{y}\frac{dy}{dx} = x \times \frac{1}{x} + \ln(x) = 1 + \ln(x)$ (using the product rule)

$$\frac{dy}{dx} = y[1 + \ln(x)] = x^x[1 + \ln(x)]$$

Exercise 4

Use logarithmic differentiation to differentiate the following:

1. $r = 2^{\theta}$
2. $y = x^x$
3. $s = \sin^t t = (\sin t)^t$
4. $v = \sin(u^u)$
5. $y = \dfrac{xe^x + 1}{e^x(x+1)}$
6. $y = \dfrac{\sin^2 x}{1 + \cos x}$
7. $y = \dfrac{(x+1)}{(2x+3)^2(x-4)}$

5. Parametric Differentiation

When a function is given in parametric form it means that x and y are given in terms of another variable, the parameter. i.e. $x = f(t)$, $y = g(t)$.

$x = t^2, y = 2t$ are parametric equations. Frequently the parameter can be eliminated.

$y = 2t \Rightarrow t = \frac{1}{2} y$

but $x = t^2$ hence $x = \left(\frac{1}{2} y\right)^2 = \frac{1}{4} y^2$

or $y^2 = 4x$, the equation of a parabola

To find the gradient of such a function in parametric form we need to use the chain rule

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx} \text{ which can be written as } \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \text{ or } \frac{y'}{x'}$$

Given $x = t^2$, $y = 2t$

we have $\dfrac{dx}{dt} = 2t$, $\dfrac{dy}{dt} = 2$

hence $\dfrac{dy}{dx} = \dfrac{2}{2t} = \dfrac{1}{t}$

In this case we can also find the gradient using the Cartesian equations:

Given $y^2 = 4x$

we have $2y\dfrac{dy}{dx} = 4$

hence $\dfrac{dy}{dx} = \dfrac{4}{2y} = \dfrac{2}{y}$

Comparing the two answers, as $y = 2t$ then $\frac{2}{y} = \frac{1}{t}$ so the two answers are the same (as expected!)

Examples

1. Find the gradient of the curve given by $x = \sin t,\ y = \cos 2t$ when $t = \frac{\pi}{3}$.

$$x = \sin t,\ y = \cos 2t$$

Find $\frac{dy}{dt}$ and $\frac{dx}{dt}$

$$\frac{dx}{dt} = \cos t,\ \frac{dy}{dt} = -2\sin 2t$$

and use $\frac{dy}{dx} = \frac{dy}{dt} \div \frac{dx}{dt} = \frac{y'}{x'}$

$$\frac{dy}{dx} = \frac{y'}{x'} = \frac{-2\sin 2t}{\cos t}$$

$$\text{when } t = \tfrac{\pi}{3},\ \frac{dy}{dx} = \frac{-2\sin\left(\frac{2\pi}{3}\right)}{\cos\left(\frac{\pi}{3}\right)} = \frac{-2\left(\frac{\sqrt{3}}{2}\right)}{\frac{1}{2}} = -2\sqrt{3}$$

Finally substitute for t

Notes a) It would be possible to eliminate t and obtain the Cartesian equation $y = 1 - 2x^2$ which will give the same value for the gradient.

b) By putting $\sin 2x = 2\sin x\cos x$, $\frac{dy}{dx} = \frac{-2\sin 2t}{\cos t}$ can be simplified to

$\frac{dy}{dx} = \frac{-4\sin t\cos t}{\cos t} = -4\sin t$ if necessary.

2. Find the gradient of the curve given by $x = \theta + \sin\theta,\ \ y = 1 - \cos\theta$ when $\theta = \frac{\pi}{2}$ and when $\theta = \pi$.

$$x = \theta + \sin\theta \Rightarrow \frac{dx}{d\theta} = 1 + \cos\theta; \quad y = 1 - \cos\theta \Rightarrow \frac{dy}{d\theta} = \sin\theta$$

$$\frac{dy}{dx} = \frac{y'}{x'} = \frac{\sin\theta}{1 + \cos\theta}$$

when $\theta = \frac{\pi}{2}$ $\quad \dfrac{dy}{dx} = \dfrac{\sin\left(\frac{\pi}{2}\right)}{1+\cos\left(\frac{\pi}{2}\right)} = \dfrac{1}{1+0} = 1,$ tangent at 45^{o}

when, $\theta = \pi$ $\quad \dfrac{dy}{dx} = \dfrac{\sin\pi}{1+\cos\pi} = \dfrac{0}{1-1} = \infty,$ tangent vertical

(really the value is indeterminate)

Notes a) $x = \theta + \sin\theta$ and $y = 1 - \cos\theta$ cannot be made into a simple Cartesian equation!

b) $\dfrac{dy}{dx} = \dfrac{\sin\theta}{1+\cos\theta}$ can be simplified by putting $\sin\theta = 2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)$

and $\cos\theta = 2\cos^2\left(\frac{\theta}{2}\right) - 1$ giving $\dfrac{dy}{dx} = \dfrac{2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)}{1+2\cos^2\left(\frac{\theta}{2}\right)-1} = \dfrac{2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)}{2\cos^2\left(\frac{\theta}{2}\right)} = \tan\left(\frac{\theta}{2}\right)$

Exercise 5 In questions 1 to 8 find $\dfrac{dy}{dx}$ in terms of the given parameter.

1. $x = 1+3t, \quad y = 3-t^2$
2. $x = t^2 + 3t, \quad y = 2t - t^2$
3. $x = 1-4s, \quad y = 1+s^2$
4. $x = \sin\phi, \quad y = \cos 2\phi + 1$
5. $x = e^{\theta}, \quad y = 1+\theta$
6. $x = e^{u} + e^{-u}, \quad y = e^{u} - e^{-u}$
7. $x = \dfrac{1}{s}, \quad y = \dfrac{s+1}{s-1}$
8. $x = \dfrac{1}{1+t}, \quad y = \dfrac{t^2}{1+t}$

In questions 9 to 14 find the gradient of the curve at the given point.

9. $x = 1-3t, \quad y = 3+2t^2; \quad t = 2$
10. $x = t^2 - t, \quad y = 2t - t^2; \quad t = -2$
11. $x = \cos\alpha, \quad y = \sin 2\alpha; \quad \alpha = \frac{\pi}{3}$
12. $x = \sin\phi - 1, \quad y = \cos\phi + 1; \quad \phi = 0$
13. $x = \ln s, \quad y = s\ln(s+1); \quad s = 1$
14. $x = e^{2r} + e, \quad y = 2e^{r} - r; \quad r = 0$

ANSWERS

Exercise 1

1. $7e^{7x}$ 2. $-\sin x e^{\cos x}$ 3. $4e^{2x}+6xe^{x^2}$ 4. $(\cos x-\sin x)e^{(\sin x+\cos x)}$

5. $\left(x^2+2x\right)e^x$ 6. $\dfrac{xe^x-3e^x-3}{x^4}$ 7. $e^{-x}=\dfrac{1}{e^x}$ 8 . $\dfrac{e^x\left(2x^2-1\right)}{x^2}$

Exercise 2

1. $\dfrac{1}{x}$ 2. $-\dfrac{1}{x}$ 3. $\dfrac{2x}{x^2+3}$ 4. $\dfrac{2x+2}{x^2+2x-1}$ 5. $\dfrac{1}{x}-\dfrac{2}{2x-3}=\dfrac{-3}{x(2x-3)}$

6. $\dfrac{2}{x}-\dfrac{1}{x+1}=\dfrac{x+2}{x(x+1)}$ 7. $\cot x+\tan x-\dfrac{1}{x}=\dfrac{1}{\cos x\sin x}-\dfrac{1}{x}$

8. Simplify to $2\ln(x-1)-\frac{1}{2}\ln(x+1)$; answer $\dfrac{2}{x-1}-\dfrac{1}{2(x+1)}=\dfrac{3x+5}{2(x-1)(x+1)}$

9. Hint first simplify as above; answer $\dfrac{2}{x+4}-\dfrac{10}{3(5x+2)}=\dfrac{4(5x-7)}{3(x+4)(5x+2)}$

Exercise 3

1. (i) $-\dfrac{x}{y}$ (ii) $\dfrac{3+4x}{7-4y}$ (iii) $\dfrac{3y+2x}{2y-3x}$ (iv) $\dfrac{6x^2+3y^2}{3y^2-6xy}=\dfrac{2x^2+y^2}{y(y-2x)}$

2. $x=2\Rightarrow y=\pm1$ $\dfrac{dy}{dx}=\dfrac{-x}{6y}$ $\text{grad}=\pm\dfrac{1}{3}$

3. $x=2\Rightarrow y=1,\ 7$ $\dfrac{dy}{dx}=\dfrac{3x^2+4y}{2y-4x}$ $\text{grad}=-\dfrac{8}{3},\ \dfrac{20}{3}$

Exercise 4

1. $(\ln 2)2^{\theta}$ 2. $(1+\ln x)x^x$ 3. $s=\left(\ln(\sin t)+\dfrac{t\cos t}{\sin t}\right)(\sin t)^t$ 4. $\cos\left(u^u\right)(1+\ln u)\,u^u$

5. $\dfrac{e^x-x-2}{e^x(x+1)^2}$ 6. $\sin x$ 7. $\dfrac{-4x^2+2x+1}{(x-4)^2(2x+3)^2}$

Exercise 5

1. $\frac{-2t}{3}$ 2. $\frac{2-2t}{2t+3}$ 3. $\frac{-s}{2}$ 4. $\frac{-2\sin 2\phi}{\cos\phi} = -4\sin\phi$

5. $e^{-\theta}$ 6. $\frac{e^u + e^{-u}}{e^u - e^{-u}}$ 7. $\frac{2s^2}{(s-1)^2}$ 8. $-(2t+t^2)$

9. $-\frac{8}{3}$ 10. $-\frac{6}{5}$ 11. $\frac{2}{\sqrt{3}}$ 12. 0 13. $\frac{1}{2} + \ln 2$ 14. $\frac{1}{2}$

Taylor & Maclaurin Series

Notation – If $f(x)$ is continuously differentiable n times then we say that $f(x) = f^{(0)}(x)$, $f'(x) = f^{(1)}(x)$, $f''(x) = f^{(2)}(x)$, and that the i^{th} derivative of f is $f^{(i)}(x)$.

Suppose that a function f has a power series representation $f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + c_4(x-a)^4 + \ldots$ for $|x-a| < R$. Differentiating term by term find the following:

- $f^{(0)}(x)$
- $f^{(1)}(x)$
- $f^{(2)}(x)$
- $f^{(3)}(x)$
- $f^{(i)}(x)$

Find a formula for c_n given the above information (Recall the convention $0! = 1$)

Theorem (Taylor Series Representation) – If f has a power series expansion about a (i.e. if $f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$) then the power series must come in the form

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Definition – The power series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$ is called the Taylor Series associated with the function f about a. In the special case when $a = 0$ we call $\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$ the Maclaurin series of f.

Show that the Maclaurin series associated with $f(x) = e^x$ is given by $\sum_{n=0}^{\infty} \frac{x^n}{n!}$. Show that this power series converges for *all* real values for x.

Notice: We do NOT know that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ yet.

Question: *When is it true that a function f(x) with derivatives of all orders is equal to its Taylor series? Equivalently, when is it true that a function f(x) has a power series?*

Notation – Let $T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$ be called the nth degree Taylor Polynomial of f about a. Let $R_n(x) = f(x) - T_n(x)$ be called the remainder of the Taylor Series of f about a.

Find and graph the 1st 2nd and 3rd degree Taylor Polynomials of $f(x) = e^x$ about 0 along with $f(x) = e^x$.

Notice: If we can show that $\lim_{n\to\infty} R_n(x) = 0$ then

$$0 = \lim_{n\to\infty} R_n(x) = \lim_{n\to\infty} [f(x) - T_n(x)] = f(x) - \lim_{n\to\infty} T_n(x)$$

Hence $f(x)$ must be equal to its Taylor Series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$

Theorem (Taylor's Inequality)(A bound on the remainder of a Taylor Series) –

Suppose that $f(x)$ has derivatives of all orders. If $|f^{(n+1)}(x)| \le M$ for all $|x - a| \le d$, then the remainder $R_n(x)$ of the Taylor Series satisifes the following inequality

$$|R_n(x)| \le \frac{M}{(n+1)!} |x-a|^{n+1} \quad \text{for all } |x-a| \le d$$

Proof (case $n = 1$):

Assume that $|f''(x)| \le M$. Then $f''(x) \le M$ for $a - d \le x \le a + d$ so

$$\int_a^x f''(t)dt \le \int_a^x M dt$$

[Type text]

$f'(x) - f'(a) \le M(x - a)$ or $f'(x) \le f'(a) + M(x - a)$

Integrating both sides again yields

$$\int_a^x f'(t)dt \le \int_a^x [f'(a) + M(t-a)]dt$$

$$f(x) - f(a) \le f'(a)(x - a) + M\frac{(x-a)^2}{2}$$

But then we have

$$R_1(x) = f(x) - T_1(x) = f(x) - f(a) - f'(a)(x - a) \le M\frac{(x-a)^2}{2}$$

as desired.

Case (n = 2) (write out yourself)

Prove that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. Then approximate e correct to 5 decimal places.

Prove that $\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$ and approximate $\sin 2$ correct to five decimal places

Prove that $\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$ and approximate $\cos 2$ correct to five decimal places

Differentiate the Taylor series for $\sin x$ term by term. Does the result agree with the Taylor series for $\cos x$?

Write out the Taylor series for e^{-x^2}

Find the Taylor Series for $\int e^{-x^2} dx$

Evaluate $\int_0^1 e^{-x^2} dx$ correct to three decimal places.

Show that $\lim_{x \to 0} \frac{e^x - 1 - x}{x^2} = 0.5$ using the Maclaurin series for e^x. Check your result using l'Hôpital.

Evaluate $\int \frac{e^x - 1}{x} dx$ as a power series term by term.

What should your calculator output if you type in taylor(e^x, *x, 4, 0)*. Explain how you knew what to predict.

Limits – Indeterminate Forms and L'Hospital's Rule

I. Indeterminate Form of the Type $\frac{0}{0}$

We have previously studied limits with the indeterminate form $\frac{0}{0}$ as shown in the following examples:

Example 1: $\lim_{x\to 2} \frac{x^2-4}{x-2} = \lim_{x\to 2} \frac{(x+2)\cancel{(x-2)}}{\cancel{x-2}} = \lim_{x\to 2}(x+2) = 2+2 = 4$

Example 2: $\lim_{x\to 0} \frac{\tan 3x}{\sin 2x} = \lim_{x\to 0} \frac{\frac{\sin 3x}{\cos 3x}}{\sin 2x} = \lim_{x\to 0} \frac{\sin 3x}{1} \cdot \frac{1}{\cos 3x} \cdot \frac{1}{\sin 2x} =$

$$\frac{3}{2}\left(\lim_{3x\to 0} \frac{\sin 3x}{3x}\right)\left(\lim_{x\to 0} \frac{1}{\cos 3x}\right)\left(\lim_{2x\to 0} \frac{2x}{\sin 2x}\right) = \frac{3}{2}(1)(1)(1) = \frac{3}{2}$$

[*Note*: We use the given limit $\lim_{\Delta\to 0} \frac{\sin \Delta}{\Delta} = 1$.]

Example 3: $\lim_{h\to 0} \frac{\sqrt[3]{8+h}-2}{h} = f'(8) = \frac{1}{3\sqrt[3]{8^2}} = \frac{1}{12}$. [*Note*: We use the definition of the derivative $f'(a) = \lim_{h \to 0} \dfrac{f(a+h) - f(a)}{h}$ where $f(x) = \sqrt[3]{x}$ and a = 8.]

Example 4: $\lim_{x \to \pi/3} \dfrac{\cos x - 1/2}{x - \pi/3} = f'(\pi/3) = -\sin(\pi/3) = -\sqrt{3}/2$. [*Note*: We use the definition of the derivative $f'(a) = \lim_{x \to a} \dfrac{f(x) - f(a)}{x - a}$ where $f(x) = \cos x$ and $a = \pi/3$.]

1

However, there is a general, systematic method for determining limits with the indeterminate form $\dfrac{0}{0}$. Suppose that *f* and *g* are differentiable functions at *x* = *a* and that $\lim_{x \to a} \dfrac{f(x)}{g(x)}$ is an indeterminate form of the type $\dfrac{0}{0}$; that is, $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$. Since *f* and *g* are differentiable functions at *x* = *a*, then *f* and *g* are continuous at *x* = *a*; that is, $f(a) = \lim_{x \to a} f(x) = 0$ and $g(a) = \lim_{x \to a} g(x) = 0$. Furthermore, since *f* and *g* are differentiable functions at *x* = *a*, then $f'(a) = \lim_{x \to a} \dfrac{f(x) - f(a)}{x - a}$ and $g'(a) = \lim_{x \to a} \dfrac{g(x) - g(a)}{x - a}$. Thus, if $g'(a) \neq 0$, then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f(x) - f(a)}{g(x) - g(a)} = \lim_{x \to a} \frac{\dfrac{f(x) - f(a)}{x - a}}{\dfrac{g(x) - g(a)}{x - a}} = \frac{f'(a)}{g'(a)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$ if f' and g' are continuous at *x* = *a*. This illustrates a special case of the technique known as L'Hospital's Rule.

L'Hospital's Rule for Form $\dfrac{0}{0}$

Suppose that f and g are differentiable functions on an open interval containing $x = a$, except possibly at $x = a$, and that $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$. If $\lim_{x \to a} \frac{f'(x)}{g'(x)}$ has a finite limit, or if this limit is $+\infty$ or $-\infty$, then $\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$. Moreover, this statement is also true in the case of a limit as $x \to a^-, x \to a^+, x \to -\infty,$ or as $x \to +\infty$.

In the following examples, we will use the following three-step process:

Step 1. Check that the limit of $\frac{f(x)}{g(x)}$ is an indeterminate form of type $\frac{0}{0}$. If it is not, then L'Hospital's Rule cannot be used.

Step 2. Differentiate f and g separately. [*Note*: Do not differentiate $\frac{f(x)}{g(x)}$ using the quotient rule!]

Step 3. Find the limit of $\frac{f'(x)}{g'(x)}$. If this limit is finite, $+\infty$, or $-\infty$, then it is equal to the limit of $\frac{f(x)}{g(x)}$. If the limit is an indeterminate form of type $\frac{0}{0}$, then simplify $\frac{f'(x)}{g'(x)}$ algebraically and apply L'Hospital's Rule again.

Example 1: $\lim_{x \to 2} \frac{x^2 - 4}{x - 2} = \lim_{x \to 2} \frac{2x}{1} = 2(2) = 4$

Example 2: $\lim_{x \to 0} \frac{\tan 3x}{\sin 2x} = \lim_{x \to 0} \frac{3\sec^2 3x}{2\cos 2x} = \frac{3(1)}{2(1)} = \frac{3}{2}$

Example 3: $\lim\limits_{h\to 0}\dfrac{\sqrt[3]{8+h}-2}{h}=\lim\limits_{h\to 0}\dfrac{\frac{1}{3}(8+h)^{-2/3}(1)}{1}=\lim\limits_{h\to 0}\dfrac{1}{3(8+h)^{2/3}}=\dfrac{1}{3(8)^{2/3}}=\dfrac{1}{12}$

Example 4: $\lim\limits_{x\to \pi/3}\dfrac{\cos x-1/2}{x-\pi/3}=\lim\limits_{x\to \pi/3}\dfrac{-\sin x}{1}=-\sin\left(\pi/3\right)=-\dfrac{\sqrt{3}}{2}$

Example 5: $\lim\limits_{x\to 0}\dfrac{e^x-x-1}{x^2}=\lim\limits_{x\to 0}\dfrac{e^x-1}{2x}=\lim\limits_{x\to 0}\dfrac{e^x}{2}=\dfrac{1}{2}$ [Use L'Hospital's Rule twice.]

Example 6: $\lim\limits_{x\to +\infty}\dfrac{1/x^2}{\sin\left(1/x\right)}=\lim\limits_{x\to +\infty}\dfrac{-2/x^3}{\cos\left(1/x\right)\left(-1/x^2\right)}=\lim\limits_{x\to +\infty}\dfrac{2/x}{\cos\left(1/x\right)}=\dfrac{0}{1}=0$, or

$\lim\limits_{x\to +\infty}\dfrac{1/x^2}{\sin\left(1/x\right)}=\lim\limits_{y\to 0^+}\dfrac{y^2}{\sin y}=\lim\limits_{y\to 0^+}\dfrac{2y}{\cos y}=\dfrac{2(0)}{1}=0$ where $y=1/x$.

Example 7: $\lim\limits_{x\to 0}\dfrac{x}{\ln x}=\lim\limits_{x\to 0}x\left(1/\ln x\right)=0(0)=0$ [This limit is not an indeterminate form of the type $\dfrac{0}{0}$, so L'Hospital's Rule cannot be used.]

3

II. Indeterminate Form of the Type $\dfrac{\infty}{\infty}$

We have previously studied limits with the indeterminate form $\dfrac{\infty}{\infty}$ as shown in the following examples:

Example 1: $\lim\limits_{x\to+\infty}\dfrac{3x^2+5x-7}{2x^2-3x+1}=\lim\limits_{x\to+\infty}\dfrac{\frac{3x^2}{x^2}+\frac{5x}{x^2}-\frac{7}{x^2}}{\frac{2x^2}{x^2}-\frac{3x}{x^2}+\frac{1}{x^2}}=$

$$\lim_{x\to+\infty}\frac{3+\frac{5}{x}-\frac{7}{x^2}}{2-\frac{3}{x}+\frac{1}{x^2}}=\lim_{x\to+\infty}\frac{3+0-0}{2-0+0}=\frac{3}{2}$$

Example 2: $\lim\limits_{x\to-\infty}\dfrac{3x-1}{x^2+1}=\lim\limits_{x\to-\infty}\dfrac{\frac{3x}{x^2}-\frac{1}{x^2}}{\frac{x^2}{x^2}+\frac{1}{x^2}}=\lim\limits_{x\to-\infty}\dfrac{\frac{3}{x}-\frac{1}{x^2}}{1+\frac{1}{x^2}}=\dfrac{0-0}{1+0}=\dfrac{0}{1}=0$

Example 3: $\lim\limits_{x\to\infty}\dfrac{3x^3-4}{2x^2+1}=\lim\limits_{x\to\infty}\dfrac{\frac{3x^3}{x^3}-\frac{4}{x^3}}{\frac{2x^2}{x^3}+\frac{1}{x^3}}=\lim\limits_{x\to\infty}\dfrac{3-\frac{4}{x^3}}{\frac{2}{x}+\frac{1}{x^3}}=\dfrac{3-0}{0+0}=\dfrac{3}{0}\Rightarrow$

limit does not exist.

Example 4: $\lim\limits_{x\to-\infty}\dfrac{\sqrt{4x^2+1}}{x+1}=\lim\limits_{x\to-\infty}\dfrac{\frac{\sqrt{4x^2+1}}{x}}{\frac{x+1}{x}}=\lim\limits_{x\to-\infty}\dfrac{\frac{\sqrt{4x^2+1}}{-\sqrt{x^2}}}{\frac{x+1}{x}}$ ($\sqrt{x^2}=-x$

because $x<0$ and thus $x=-\sqrt{x^2}$) = $\lim\limits_{x\to-\infty}\dfrac{-\sqrt{\frac{4x^2+1}{x^2}}}{\frac{x+1}{x}}=$

4

$$\lim_{x\to-\infty}\frac{-\sqrt{4+\frac{1}{x^2}}}{1+\frac{1}{x^2}}=\frac{-\sqrt{4}}{1}=-2\,.$$

However, we could use another version of L'Hospital's Rule.

L'Hospital's Rule for Form $\frac{\infty}{\infty}$

Suppose that *f* and *g* are differentiable functions on an open interval containing *x* = *a*, except possibly at *x* = *a*, and that $\lim_{x\to a} f(x)=\infty$ and $\lim_{x\to a} g(x)=\infty$. If $\lim_{x\to a}\frac{f'(x)}{g'(x)}$ has a finite limit, or if this limit is $+\infty$ or $-\infty$, then $\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}$. Moreover, this statement is also true in the case of a limit as $x\to a^-, x\to a^+, x\to -\infty,$ or as $x\to +\infty$.

Example 1: $\lim_{x\to+\infty}\frac{3x^2+5x-7}{2x^2-3x+1}=\lim_{x\to+\infty}\frac{6x+5}{4x-3}=\lim_{x\to+\infty}\frac{6}{4}=\frac{3}{2}$

Example 2: $\lim_{x\to-\infty}\frac{3x-1}{x^2+1}=\lim_{x\to-\infty}\frac{3}{2x}=\frac{3}{2}\lim_{x\to-\infty}\frac{1}{x}=\frac{3}{2}(0)=0$

Example 3: $\lim_{x\to\infty}\frac{3x^3-4}{2x^2+1}=\lim_{x\to\infty}\frac{9x^2}{4x}=\lim_{x\to\infty}\frac{18x}{4}=\infty$

Example 4: $\lim_{x\to-\infty}\frac{\sqrt{4x^2+1}}{x+1}=\lim_{x\to-\infty}\frac{\frac{8x}{2\sqrt{4x^2+1}}}{1}=\lim_{x\to-\infty}\frac{4x}{\sqrt{4x^2+1}}\Rightarrow$ L'Hospital's Rule does not help in this situation. We would find the limit as we did previously.

5

Example 5: $\lim\limits_{x\to+\infty}\dfrac{\ln(x^2+1)}{\ln(x^3+1)}=\lim\limits_{x\to+\infty}\dfrac{\dfrac{2x}{x^2+1}}{\dfrac{3x^2}{x^3+1}}=\lim\limits_{x\to+\infty}\dfrac{2x(x^3+1)}{3x^2(x^2+1)}=\lim\limits_{x\to+\infty}\dfrac{2x^4+2x}{3x^4+3x^2}=$

$\lim\limits_{x\to+\infty}\dfrac{8x^3+2}{12x^3+6x}=\lim\limits_{x\to+\infty}\dfrac{24x^2}{36x^2+6}=\lim\limits_{x\to+\infty}\dfrac{48x}{72x}=\dfrac{48}{72}=\dfrac{2}{3}$

Example 6: $\lim\limits_{x\to0^+}\dfrac{\ln x}{1/x^2}=\lim\limits_{x\to0^+}\dfrac{1/x}{-2/x^3}=\lim\limits_{x\to0^+}\dfrac{x^3}{-2x}=\lim\limits_{x\to0^+}\dfrac{x^2}{-2}=\dfrac{0^2}{-2}=0$

Example 7: $\lim\limits_{x\to+\infty}\dfrac{\arctan x}{x}=\left(\lim\limits_{x\to+\infty}\dfrac{1}{x}\right)\left(\lim\limits_{x\to+\infty}\arctan x\right)=(0)\left(\dfrac{\pi}{2}\right)=0$ [This limit is not an indeterminate form of the type $\dfrac{\infty}{\infty}$, so L'Hospital's Rule cannot be used.]

III. Indeterminate Form of the Type $0\cdot\infty$

Indeterminate forms of the type $0\cdot\infty$ can sometimes be evaluated by rewriting the product as a quotient, and then applying L'Hospital's Rule for the indeterminate forms of type $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$.

Example 1: $\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{1/x} = \lim_{x\to 0^+} \frac{1/x}{-1/x^2} = \lim_{x\to 0^+} \frac{-x^2}{x} = \lim_{x\to 0^+} (-x) = 0$

Example 2: $\lim_{x\to 0^+} (\sin x)\ln x = \lim_{x\to 0^+} \frac{\ln x}{\csc x} = \lim_{x\to 0^+} \frac{1/x}{-\csc x \cot x} = \lim_{x\to 0^+} \frac{-\sin x \tan x}{x} =$

$$\left(\lim_{x\to 0^+} \frac{-\sin x}{x}\right)\left(\lim_{x\to 0^+} \tan x\right) = (-1)(0) = 0$$

Example 3: $\lim_{x\to +\infty} x\sin\left(1/x\right) = \lim_{x\to +\infty} \frac{\sin\left(1/x\right)}{1/x} = \lim_{y\to 0^+} \frac{\sin y}{y} = 1$ [Let $y = 1/x$.]

IV. Indeterminate Form of the Type $\infty - \infty$

A limit problem that leads to one of the expressions

$$(+\infty)-(+\infty),\quad (-\infty)-(-\infty),\quad (+\infty)+(-\infty),\quad (-\infty)+(+\infty)$$

is called an indeterminate form of type $\infty - \infty$. Such limits are indeterminate because the two terms exert conflicting influences on the expression; one pushes it in the positive direction and the other pushes it in the negative direction. However, limits problems that lead to one the expressions

$$(+\infty)+(+\infty),\quad (+\infty)-(-\infty),\quad (-\infty)+(-\infty),\quad (-\infty)-(+\infty)$$

are not indeterminate, since the two terms work together (the first two produce a limit of $+\infty$ and the last two produce a limit of $-\infty$). Indeterminate forms of the type $\infty - \infty$ can sometimes be evaluated by combining the terms and manipulating the result to produce an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$.

Example 1: $\lim\limits_{x\to 0^+}\left(\frac{1}{x}-\frac{1}{\sin x}\right)=\lim\limits_{x\to 0^+}\left(\frac{\sin x - x}{x\sin x}\right)=\lim\limits_{x\to 0^+}\frac{\cos x - 1}{x\cos x+\sin x}=$

$$\lim_{x\to 0^+}\frac{-\sin x}{-x\sin x+\cos x+\cos x}=\frac{0}{2}=0$$

Example 2: $\lim\limits_{x\to 0}\left[\ln(1-\cos x)-\ln\left(x^2\right)\right]=\lim\limits_{x\to 0}\left[\ln\left(\frac{1-\cos x}{x^2}\right)\right]=$

$$\ln\left[\lim_{x\to 0}\left(\frac{1-\cos x}{x^2}\right)\right]=\ln\left[\lim_{x\to 0}\left(\frac{\sin x}{2x}\right)\right]=\ln\left(\frac{1}{2}\right)$$

V. Indeterminate Forms of the Types 0^0, ∞^0, 1^∞

Limits of the form $\lim\limits_{x\to a}\left[f(x)\right]^{g(x)}$ $\left\{or \lim\limits_{x\to\infty}\left[f(x)\right]^{g(x)}\right\}$ frequently give rise to indeterminate forms of the types 0^0, ∞^0, 1^∞. These indeterminate forms can sometimes be evaluated as follows:

(1) $y=\left[f(x)\right]^{g(x)}$

(2) $\ln y=\ln\left[f(x)\right]^{g(x)}=g(x)\ln\left[f(x)\right]$

(3) $\lim\limits_{x\to a}\left[\ln y\right]=\lim\limits_{x\to a}\left\{g(x)\ln\left[f(x)\right]\right\}$

7

The limit on the righthand side of the equation will usually be an indeterminate limit of the type $0\cdot\infty$. Evaluate this limit using the technique previously described. Assume that $\lim\limits_{x\to a}\left\{g(x)\ln\left[f(x)\right]\right\}=L.$

(4) Finally, $\lim\limits_{x\to a}\left[\ln y\right]=L\Rightarrow\ln\left[\lim\limits_{x\to a} y\right]=L\Rightarrow\lim\limits_{x\to a} y=e^L$.

Example 1: Find $\lim_{x\to 0^+} x^x$.

This is an indeterminate form of the type 0^0. Let $y = x^x \Rightarrow \ln y = \ln x^x =$ $x\ln x$. $\lim_{x\to 0^+} \ln y = \lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \dfrac{\ln x}{1/x} = \lim_{x\to 0^+} \dfrac{1/x}{-1/x^2} = \lim_{x\to 0^+} (-x) = 0.$

Thus, $\lim_{x\to 0^+} x^x = e^0 = 1$.

Example 2: Find $\lim_{x\to +\infty} (e^x+1)^{-2/x}$.

This is an indeterminate form of the type ∞^0. Let $y = (e^x+1)^{-2/x} \Rightarrow$ $\ln y = \ln\left[(e^x+1)^{-2/x}\right] = \dfrac{-2\ln(e^x+1)}{x}$. $\lim_{x\to +\infty} \ln y = \lim_{x\to +\infty} \dfrac{-2\ln(e^x+1)}{x} =$

$\lim_{x\to +\infty} \dfrac{-2\left(\dfrac{e^x}{e^x+1}\right)}{1} = \lim_{x\to +\infty} \dfrac{-2e^x}{e^x+1} = \lim_{x\to +\infty} \dfrac{-2e^x}{e^x} = -2$. Thus, $\lim_{x\to +\infty} (e^x+1)^{-2/x} =$

e^{-2}.

Example 3: Find $\lim_{x\to 0^+} (\cos x)^{1/x}$.

This is an indeterminate form of the type 1^∞. Let $y = (\cos x)^{1/x} \Rightarrow$ $\ln y = \ln\left[(\cos x)^{1/x}\right] = \dfrac{\ln(\cos x)}{x}$. $\lim_{x\to 0^+} \ln y = \lim_{x\to 0^+} \dfrac{\ln(\cos x)}{x} =$

$\lim_{x\to 0^+} (-\tan x) = 0$. Thus, $\lim_{x\to 0^+} (\cos x)^{1/x} = e^0 = 1$.

Practice Sheet for L'Hospital's Rule

(1) $\lim\limits_{x\to 0}\dfrac{xe^{3x}-x}{1-\cos(2x)}=$

(2) $\lim\limits_{x\to +\infty}\dfrac{x}{(\ln x)^3}=$

(3) $\lim\limits_{x\to 0}[\ln(1-\cos x)-\ln(x^2)]=$

(4) $\lim\limits_{x\to +\infty}\left(1-\dfrac{2}{x}\right)^{3x}=$

(5) $\lim\limits_{x\to +\infty}\dfrac{\cos\left(\frac{1}{x}\right)-1}{\cos\left(\frac{2}{x}\right)-1}=$

(6) $\lim\limits_{x\to 0}\dfrac{\sqrt{1-x}-\sqrt{1-x^2}}{x}=$

(7) $\lim\limits_{x\to 0}(\cos x)^{1/x^2}=$

(8) $\lim\limits_{x\to 1}\dfrac{5x^4-7x^3+x^2-x+2}{3x^4-8x^3+6x^2-1}=$

(9) $\lim\limits_{x\to 0}\dfrac{9-\sqrt{81-5x}}{x}=$

(10) $\lim\limits_{x\to +\infty}\dfrac{\ln\left(x^3+2\right)}{\ln\left(5x^3-1\right)}=$

(11) $\lim\limits_{x\to +\infty}\left(e^x+1\right)^{-2/x}=$

(12) $\lim\limits_{x\to 0}\dfrac{\sin(4x)-2\sin(2x)}{x^3}=$

(13) $\lim\limits_{x\to 0}\left[\dfrac{1}{\sin x}-\dfrac{1}{x}\right]=$

(14) $\lim\limits_{x\to +\infty}x\left(e^{1/x}-1\right)=$

(15) $\lim\limits_{x\to 0^+}\sqrt[3]{x}\ln x=$

(16) $\lim\limits_{x\to 0}\dfrac{\ln\left(\dfrac{2x+1}{5x+1}\right)}{x}=$

(17) $\lim\limits_{x\to +\infty}\left(1+\dfrac{e}{x}\right)^{x/2}=$

(18) $\lim\limits_{x\to 0}\dfrac{\arctan(\sin 3x)}{x}=$

[Type text]

(19) $\lim\limits_{x\to 0^+} \dfrac{\int_0^x \sin\left(t^2\right)dt}{x^3} =$

(20) $\lim\limits_{x\to 0} \left(e^{2x} + x\right)^{1/x} =$

(21) $\lim\limits_{x\to +\infty} \dfrac{\arctan x}{x} =$

(22) $\lim\limits_{x\to 0} \dfrac{\arctan(\sin(3x))}{\arcsin(2\tan x)} =$

(23) $\lim\limits_{x\to 0} \dfrac{\ln(\cos x)}{x^2} =$

(24) $\lim\limits_{x\to +\infty} \left(1 + \dfrac{1}{2x}\right)^x =$

(25) $\lim\limits_{x\to +\infty} (\ln x)^{1/x} =$

(26) $\lim\limits_{x\to +\infty} \left(\ln(e^x + 1) - x\right) =$

Solution Key for L'Hospital's Rule

(1) $\lim_{x\to 0}\frac{xe^{3x}-x}{1-\cos(2x)}=\lim_{x\to 0}\frac{3xe^{3x}+e^{3x}-1}{2\sin 2x}=\lim_{x\to 0}\frac{9xe^{3x}+3e^{3x}+3e^{3x}}{4\cos 2x}=\frac{6}{4}=\frac{3}{2}$

(2) $\lim_{x\to+\infty}\frac{x}{(\ln x)^3}=\lim_{x\to+\infty}\frac{1}{3(\ln x)^2\left(1/x\right)}=\lim_{x\to+\infty}\frac{x}{3(\ln x)^2}=\lim_{x\to+\infty}\frac{1}{6\ln x\left(1/x\right)}=\lim_{x\to+\infty}\frac{x}{6\ln x}=$

$\lim_{x\to+\infty}\frac{1}{6\left(1/x\right)}=\lim_{x\to+\infty}\frac{x}{6}=+\infty$

(3) $\lim_{x\to 0}[\ln(1-\cos x)-\ln(x^2)]=\lim_{x\to 0}\ln\left(\frac{1-\cos x}{x^2}\right)=\ln\left\{\lim_{x\to 0}\left(\frac{1-\cos x}{x^2}\right)\right\}=$

$\ln\left\{\lim_{x\to 0}\left(\frac{\sin x}{2x}\right)\right\}=\ln\left(\frac{1}{2}\right)=-\ln 2$

(4) Let $y=\frac{1}{x}\Rightarrow\lim_{x\to+\infty}\left(1-\frac{2}{x}\right)^{3x}=\lim_{y\to 0^+}(1-2y)^{3/y}$. Now, let $z=(1-2y)^{3/y}\Rightarrow\ln z=$

$\ln(1-2y)^{3/y}=\frac{3\ln(1-2y)}{y}\Rightarrow\lim_{y\to 0^+}\ln z=\lim_{y\to 0^+}\frac{3\ln(1-2y)}{y}=\lim_{y\to 0^+}\frac{\frac{-6}{1-2y}}{1}=-6$. Thus,

$\lim_{y\to 0^+}\ln z=-6\Rightarrow\ln\left(\lim_{y\to 0^+}z\right)=-6\Rightarrow\lim_{y\to 0^+}z=e^{-6}\Rightarrow\lim_{x\to+\infty}\left(1-\frac{2}{x}\right)^{3x}=\lim_{y\to 0^+}(1-2y)^{3/y}=$

$\lim_{y\to 0^+}z=e^{-6}$.

(5) Let $y=\frac{1}{x}\Rightarrow\lim_{x\to+\infty}\frac{\cos\left(\frac{1}{x}\right)-1}{\cos\left(\frac{2}{x}\right)-1}=\lim_{y\to 0^+}\frac{\cos(y)-1}{\cos(2y)-1}=\lim_{y\to 0^+}\frac{-\sin(y)}{-2\sin(2y)}=$

$\lim_{y\to 0^+}\frac{\sin y}{4\sin y\cos y}=\lim_{y\to 0^+}\frac{1}{4\cos y}=\frac{1}{4}$.

(6) $\lim\limits_{x\to 0}\dfrac{\sqrt{1-x}-\sqrt{1-x^2}}{x}=\lim\limits_{x\to 0}\dfrac{\sqrt{1-x}-\sqrt{1-x^2}}{x}\cdot\dfrac{\sqrt{1-x}+\sqrt{1-x^2}}{\sqrt{1-x}+\sqrt{1-x^2}}=$

$\lim\limits_{x\to 0}\dfrac{(1-x)-(1-x^2)}{x\left(\sqrt{1-x}+\sqrt{1-x^2}\right)}=\lim\limits_{x\to 0}\dfrac{x^2-x}{x\left(\sqrt{1-x}+\sqrt{1-x^2}\right)}=\lim\limits_{x\to 0}\dfrac{x-1}{\sqrt{1-x}+\sqrt{1-x^2}}=-\dfrac{1}{2}.$

(7) Let $y=(\cos x)^{1/x^2}\Rightarrow \ln y=\ln(\cos x)^{1/x^2}=\dfrac{\ln(\cos x)}{x^2}\Rightarrow \lim\limits_{x\to 0}(\ln y)=\lim\limits_{x\to 0}\dfrac{\ln(\cos x)}{x^2}=$

$\lim\limits_{x\to 0}\dfrac{\frac{-\sin x}{\cos x}}{2x}=\lim\limits_{x\to 0}\dfrac{-\sin x}{2x\cos x}=\lim\limits_{x\to 0}\left(\dfrac{\sin x}{x}\right)\left(\dfrac{-1}{2\cos x}\right)=-\dfrac{1}{2}$. Thus, $\lim\limits_{x\to 0}(\ln y)=-\dfrac{1}{2}\Rightarrow$

$\ln\left(\lim\limits_{x\to 0}y\right)=-\dfrac{1}{2}\Rightarrow \lim\limits_{x\to 0}y=e^{-1/2}\Rightarrow \lim\limits_{x\to 0}(\cos x)^{1/x^2}=\lim\limits_{x\to 0}y=e^{-1/2}.$

(8) $\lim\limits_{x\to 1}\dfrac{5x^4-7x^3+x^2-x+2}{3x^4-8x^3+6x^2-1}=\lim\limits_{x\to 1}\dfrac{20x^3-21x^2+2x-1}{12x^3-24x^2+12x}=\lim\limits_{x\to 1}\dfrac{60x^2-42x+2}{36x^2-48x+12}=$

$\dfrac{20}{0}\Rightarrow$ limit does not exist.

(9) $\lim\limits_{x\to 0}\dfrac{9-\sqrt{81-5x}}{x}=\lim\limits_{x\to 0}\dfrac{9-\sqrt{81-5x}}{x}\cdot\dfrac{9+\sqrt{81-5x}}{9+\sqrt{81-5x}}=\lim\limits_{x\to 0}\dfrac{81-(81-5x)}{x\left(9+\sqrt{81-5x}\right)}=$

$\lim\limits_{x\to 0}\dfrac{5x}{x\left(9+\sqrt{81-5x}\right)}=\lim\limits_{x\to 0}\dfrac{5}{9+\sqrt{81-5x}}=\dfrac{5}{18}.$

(10) $\lim\limits_{x\to+\infty}\dfrac{\ln\left(x^3+2\right)}{\ln\left(5x^3-1\right)}=\lim\limits_{x\to+\infty}\dfrac{\frac{3x^2}{x^3+2}}{\frac{15x^2}{5x^3-1}}=\lim\limits_{x\to+\infty}\dfrac{3(5x^3-1)}{15(x^3+2)}=\lim\limits_{x\to+\infty}\dfrac{15x^3-3}{15x^3+30}=1$

(11) Let $y=\left(e^x+1\right)^{-2/x} \Rightarrow \ln y=\ln\left(e^x+1\right)^{-2/x}=\dfrac{-2\ln(e^x+1)}{x} \Rightarrow \lim\limits_{x\to+\infty}\ln y=$

$$\lim_{x\to+\infty}\frac{-2\ln(e^x+1)}{x}=\lim_{x\to+\infty}\frac{\dfrac{-2e^x}{e^x+1}}{1}=\lim_{x\to+\infty}\frac{-2e^x}{e^x+1}=\lim_{x\to+\infty}\frac{-2e^x}{e^x}=-2.$$ Thus, $\lim\limits_{x\to+\infty}\ln y=$

$$-2 \Rightarrow \ln\left(\lim_{x\to+\infty}y\right)=-2 \Rightarrow \lim_{x\to+\infty}y=e^{-2} \Rightarrow \lim_{x\to+\infty}\left(e^x+1\right)^{-2/x}=\lim_{x\to+\infty}y=e^{-2}.$$

(12) $$\lim_{x\to0}\frac{\sin(4x)-2\sin(2x)}{x^3}=\lim_{x\to0}\frac{4\cos(4x)-4\cos(2x)}{3x^2}=\lim_{x\to0}\frac{-16\sin(4x)+8\sin(2x)}{6x}=$$

$$\lim_{x\to0}\frac{-64\cos(4x)+16\cos(2x)}{6}=\frac{-48}{6}=-8.$$

(13) $$\lim_{x\to0}\left[\frac{1}{\sin x}-\frac{1}{x}\right]=\lim_{x\to0}\left(\frac{x-\sin x}{x\sin x}\right)=\lim_{x\to0}\frac{1-\cos x}{x\cos x+\sin x}=$$

$$\lim_{x\to0}\frac{\sin x}{-x\sin x+\cos x+\cos x}=\frac{0}{2}=0.$$

(14) $$\lim_{x\to+\infty}x\left(e^{1/x}-1\right)=\lim_{x\to+\infty}\frac{e^{1/x}-1}{1/x}=\lim_{x\to+\infty}\frac{e^{1/x}\left(-1/x^2\right)}{-1/x^2}=\lim_{x\to+\infty}e^{1/x}=e^0=1.$$

(15) $$\lim_{x\to0^+}\sqrt[3]{x}\ln x=\lim_{x\to0^+}\frac{\ln x}{x^{-1/3}}=\lim_{x\to0^+}\frac{1/x}{-\frac{1}{3}x^{-4/3}}=\lim_{x\to0^+}\frac{-3x^{4/3}}{x}=\lim_{x\to0^+}(-3\sqrt[3]{x})=0.$$

(16) $\lim_{x\to 0}\frac{\ln\left(\frac{2x+1}{5x+1}\right)}{x}=\lim_{x\to 0}\frac{\left(\frac{5x+1}{2x+1}\right)\left(\frac{2(5x+1)-5(2x+1)}{(5x+1)^2}\right)}{1}=\lim_{x\to 0}\frac{-3}{(2x+1)(5x+1)}=-3$.

(17) Let $y=\frac{1}{x}\Rightarrow\lim_{x\to+\infty}\left(1+\frac{e}{x}\right)^{x/2}=\lim_{y\to 0^+}(1+ey)^{1/2y}$. Next, let $z=(1+ey)^{1/2y}\Rightarrow\ln z=$

$\ln(1+ey)^{1/2y}=\frac{\ln(1+ey)}{2y}\Rightarrow\lim_{y\to 0^+}\ln z=\lim_{y\to 0^+}\frac{\ln(1+ey)}{2y}=\lim_{y\to 0^+}\frac{\frac{e}{1+ey}}{2}=\frac{e}{2}$. Thus,

13

$\lim_{y\to 0^+}\ln z=\frac{e}{2}\Rightarrow\ln\left(\lim_{y\to 0^+}z\right)=\frac{e}{2}\Rightarrow\lim_{y\to 0^+}z=e^{e/2}\Rightarrow\lim_{x\to+\infty}\left(1+\frac{e}{x}\right)^{x/2}=\lim_{y\to 0^+}(1+ey)^{1/2y}=$

$\lim_{y\to 0^+}z=e^{e/2}$. $y=\left(e^{2x}+x\right)^{1/x}\Rightarrow\ln y=\ln\left(e^{2x}+x\right)^{1/x}=\frac{\ln\left(e^{2x}+x\right)}{x}\Rightarrow\lim_{x\to 0}\ln y=$

(18) $\lim_{x\to 0}\frac{\arctan(\sin 3x)}{x}=\lim_{x\to 0}\frac{\frac{3\cos 3x}{1+\sin^2 3x}}{1}=3$.

(19) $\lim_{x\to 0^+}\frac{\int_0^x\sin\left(t^2\right)dt}{x^3}=\lim_{x\to 0^+}\frac{\sin(x^2)}{3x^2}=\lim_{x\to 0^+}\frac{2x\cos(x^2)}{6x}=\lim_{x\to 0^+}\frac{\cos(x^2)}{3}=\frac{1}{3}$.

(20) Let $y=\left(e^{2x}+x\right)^{1/x}\Rightarrow\ln y=\ln\left(e^{2x}+x\right)^{1/x}=\frac{\ln\left(e^{2x}+x\right)}{x}\Rightarrow\lim_{x\to 0}\ln y=$

$\lim_{x\to 0}\frac{\ln\left(e^{2x}+x\right)}{x}=\lim_{x\to 0}\frac{\frac{2e^{2x}+1}{e^{2x}+x}}{1}=3$. Thus $\lim_{x\to 0}\ln y=3\Rightarrow\ln\left(\lim_{x\to 0}y\right)=3\Rightarrow$

$$\lim_{x\to 0} y = e^3 \Rightarrow \lim_{x\to 0}\left(e^{2x}+x\right)^{1/x} = \lim_{x\to 0} y = e^3.$$

(21) $\lim_{x\to+\infty}\frac{\arctan x}{x} = \frac{\pi/2}{+\infty} = 0$.

(22) $\lim_{x\to 0}\frac{\arctan(\sin(3x))}{\arcsin(2\tan x)} = \lim_{x\to 0}\frac{\frac{3\cos 3x}{1+\sin^2 3x}}{\frac{2\sec^2 x}{\sqrt{1-4\tan^2 x}}} = \frac{3/1}{2/1} = \frac{3}{2}$.

(23) $\lim_{x\to 0}\frac{\ln(\cos x)}{x^2} = \lim_{x\to 0}\frac{\frac{-\sin x}{\cos x}}{2x} = \lim_{x\to 0}\left(\frac{\sin x}{x}\right)\left(\frac{-1}{2\cos x}\right) = -\frac{1}{2}$.

(24) Let $y = \frac{1}{x} \Rightarrow \lim_{x\to+\infty}\left(1+\frac{1}{2x}\right)^x = \lim_{y\to 0^+}\left(1+\frac{1}{2}y\right)^{1/y}$. Let $z = \left(1+\frac{1}{2}y\right)^{1/y} \Rightarrow \ln z =$

$$\ln\left(1+\frac{1}{2}y\right)^{1/y} = \frac{\ln\left(1+\frac{1}{2}y\right)}{y} \Rightarrow \lim_{y\to 0^+}\ln z = \lim_{y\to 0^+}\frac{\ln\left(1+\frac{1}{2}y\right)}{y} = \lim_{y\to 0^+}\frac{\frac{1/2}{1+1/2\,y}}{1} = \frac{1}{2}.$$

Thus, $\lim_{y\to 0^+}\ln z = \frac{1}{2} \Rightarrow \ln\left(\lim_{y\to 0^+} z\right) = \frac{1}{2} \Rightarrow \lim_{y\to 0^+} z = e^{1/2} \Rightarrow \lim_{x\to+\infty}\left(1+\frac{1}{2x}\right)^x =$

$$\lim_{y\to 0^+}\left(1+\frac{1}{2}y\right)^{1/y} = \lim_{y\to 0^+} z = e^{1/2}.$$

(25) Let $y = (\ln x)^{1/x} \Rightarrow \ln y = \ln(\ln x)^{1/x} = \frac{\ln(\ln x)}{x} \Rightarrow \lim_{x\to+\infty}\ln y = \lim_{x\to+\infty}\frac{\ln(\ln x)}{x} =$

$$\lim_{x\to+\infty} \frac{\frac{1}{x\ln x}}{1} = \lim_{x\to+\infty} \frac{1}{x\ln x} = 0.$$ Thus, $\lim_{x\to+\infty} \ln y = 0 \Rightarrow \ln\left(\lim_{x\to+\infty} y\right) = 0 \Rightarrow \lim_{x\to+\infty} y =$

$$e^0 = 1 \Rightarrow \lim_{x\to+\infty} (\ln x)^{1/x} = \lim_{x\to+\infty} y = 1.$$

(26) $\lim_{x\to+\infty} (\ln(e^x+1) - x) = \lim_{x\to+\infty} \left(\ln\left(e^x+1\right) - \ln e^x\right) = \lim_{x\to+\infty} \left(\ln\left(\frac{e^x+1}{e^x}\right)\right) =$

$$\ln\left(\lim_{x\to+\infty}\left(\frac{e^x+1}{e^x}\right)\right) = \ln\left(\lim_{x\to+\infty}\left(\frac{e^x}{e^x}\right)\right) = \ln 1 = 0.$$

The Mean-value Theorem

Consider the quantity Q defined by the equation

$$\frac{f(b)-f(a)}{b-a} = Q, \tag{13.1}$$

or

$$f(b) - f(a) - (b-a)Q = 0. \tag{13.2}$$

Let $F(x)$ be a function formed by replacing b by x in the left-hand member of (13.2); that is,

$$F(x) = f(x) - f(a) - (x-a)Q. \tag{13.3}$$

From (13.2), $F(b) = 0$, and from (13.3), $F(a) = 0$; therefore, by Rolle's Theorem (see §13.1), $F'(x)$ must be zero for at least one value of x between a and b, say for x_1. But by differentiating (13.3) we get

$$F'(x) = f'(x) - Q.$$

Therefore, since $F'(x_1) = 0$, then also $f'(x_1) - Q = 0$, and $Q = f'(x_1)$. Substituting this value of Q in (13.1), we get the Theorem of Mean Value[13.1],

$$\frac{f(b) - f(a)}{b - a} = f'(x_1),\ a < x_1 < b \qquad (13.4)$$

where in general all we know about x_1 is that it lies between a and b.

The Theorem of Mean Value interpreted Geometrically.

Let the curve in the figure be the locus of $y = f(x)$.

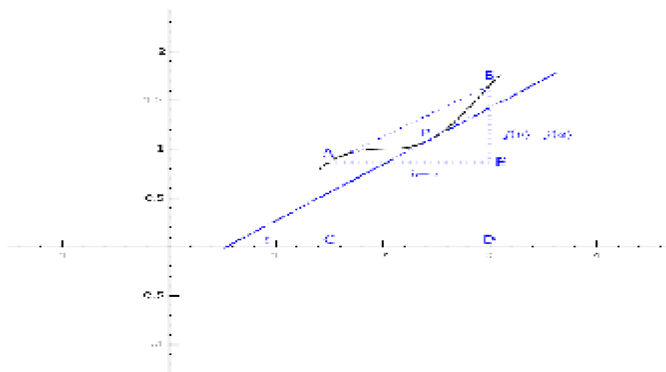


Figure : Geometric illustration of the Mean value theorem.

Take $OC = a$ and $OD = b$; then $f(a) = CA$ and $f(b) = DB$, giving $AE = b - a$ and $EB = f(b) - f(a)$. Therefore the slope of the chord AB is

$$\tan EAB = \frac{EB}{AE} = \frac{f(b) - f(a)}{b - a}.$$

There is at least one point on the curve between A and B (as P) where the tangent (or curve) is parallel to the chord AB. If the abscissa of P is x_1 the slope at P is

$$\tan t = f'(x_1) = \tan EAB.$$

Equating these last two equations, we get

$$\frac{f(b)-f(a)}{b-a}=f'(x_1),$$

which is the Theorem of Mean Value.

The student should draw curves to show that there may be more than one such point in the interval; and curves to illustrate, on the other hand, that the theorem may not be true if $f(x)$ becomes discontinuous for any value of x between a and b or if $f'(x)$ becomes discontinuous

Clearing of fractions, we may also write the theorem in the form

$$f(b)=f(a)+(b-a)f'(x_1). \tag{13.5}$$

Let $b=a+\Delta a$; then $b-a=\Delta a$, and since x_1 is a number lying between a and b, we may write

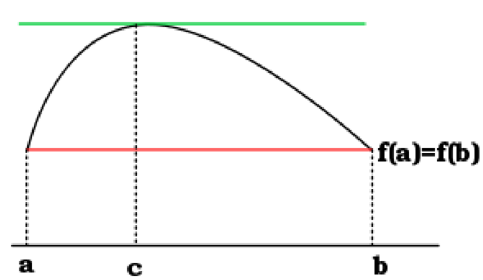
$$x_1=a+\theta\cdot\Delta a,$$

where θ is a positive proper fraction. Substituting in we get another form of the Theorem of Mean Value.

$$f(a+\Delta a)-f(a)=\Delta a f'(a+\theta\cdot\Delta a),\quad 0<\theta<1.$$

Rolle's Theorem

A differentiable and continuous function, which attains equal values at two points, must have a point somewhere between them where the slope of the tangent line to the graph of the function is zero.



The application of the Mean Value Theorem applies to Rolle's Theorem but only where the slope of the tangent is equal to zero. $\Rightarrow \left(f'(c) = 0 = \dfrac{f(b) - f(a)}{b - a} \right)$

MAX. & MIN

Introduction

A great many practical problems occur where we have to deal with functions of such a nature that they have a greatest (maximum) value or a least (minimum) valueand it is very important to know what particular value of the variable gives such a value of the function.

Example For instance, suppose that it is required to find the dimensions of the rectangle of greatest area that can be inscribed in a circle of radius 5inches. Consider the circle in Figure:

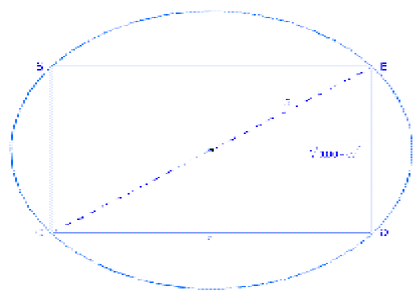


Figure: A rectangle with circumscribed circle.

Inscribe any rectangle, as BCDE. Let $CD = x$; then $DE = \sqrt{100 - x^2}$, and the area of the rectangle is evidently

$$A = A(x) = x\sqrt{100 - x^2}.$$

That a rectangle of maximum area must exist may be seen as follows: Let the base $CD (= x)$ increase to 10inches (the diameter); then the altitude $DE (= \sqrt{100 - x^2})$ will decrease to zero and the area will become zero. Now let the base decrease to zero; then the altitude will increase to 10inches and the area will again become zero. It is therefore intuitionally evident that there exists a greatest rectangle. By a careful study of the figure we might suspect that when the rectangle becomes a square its area would be the greatest, but this would at best be mere guesswork. A better way would evidently be to plot the graph of the function $A = A(x)$ and note its behavior. To aid us in drawing the graph of $A(x)$, we observe that

(a)

from the nature of the problem it is evident that x and A must both be positive; and

(b)

the values of x range from zero to 10 inclusive.

Now construct a table of values and draw the graph. What do we learn from the graph?

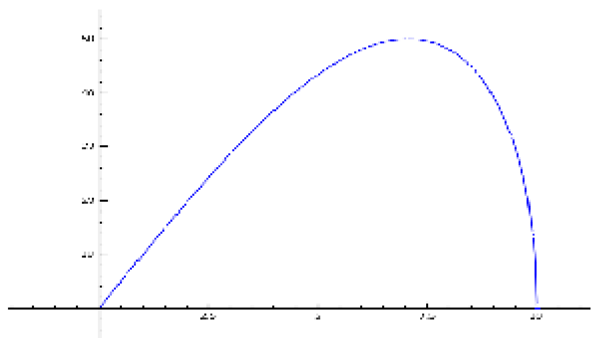


Figure: The area of a rectangle with fixed circumscribed circle.

(a) If the rectangle is carefully drawn, we may find quite accurately the area of the rectangle corresponding to any value x by measuring the length of the corresponding ordinate. Thus, when $x = OM = 3$ inches, then $A = MP = 28.6$ square inches; and when $x = ON = \frac{9}{2}$ inches, then $A = NQ \approx 33.8$ sq. in. (found by measurement).

(b) There is one horizontal tangent (RS). The ordinate TH from its point of contact T is greater than any other ordinate. Hence this discovery: One of the inscribed rectangles has evidently a greater area than any of the others. In other words, we may infer from this that the function defined by $A = A(x)$ has a maximum value. We cannot find this value (= HT) exactly by measurement, but it is very easy to find, using Calculus methods. We observed that at T the tangent was horizontal; hence the slope will be zero at that point . To find the abscissa of T we then find the first derivative of $A(x)$, place it equal to zero, and solve for x. Thus

$$
\begin{aligned}
A &= x\sqrt{100 - x^2}, \\
\frac{dA}{dx} &= \frac{100-2x^2}{\sqrt{100-x^2}}, \\
\frac{100-2x^2}{\sqrt{100-x^2}} &= 0.
\end{aligned}
$$

Solving, $x = 5\sqrt{2}$. Substituting back, we get $DE = \sqrt{100 - x^2} = 5\sqrt{2}$. Hence the rectangle of maximum area inscribed in the circle is a square of area $A = CD \times DE = 5\sqrt{2} \times 5\sqrt{2} = 50$ square inches. The length of HT is therefore 50.

Example A wooden box is to be built to contain 108 cu. ft. It is to have an open top and a square base. What must be its dimensions in order that the amount of material required shall be a minimum; that is, what dimensions will make the cost the least?

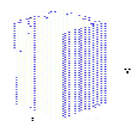


Figure : A box with square $x \times x$ base, height $y = 108/x^2$, and fixed volume.

Let x denote the length of side of square base in feet, and y denote the height of box. Since the volume of the box is given, y may be found in terms of x. Thus $\text{volume} = x^2y = 108$, so $y = \frac{108}{x^2}$. We may now express the number (= M) of square feet of lumber required as a function of x as follows:

area of base = x^2 sq. ft.,

area of four sides = $4xy = \frac{432}{x}$ sq. ft.

Hence

$$M = M(x) = x^2 + \frac{432}{x}$$

is a formula giving the number of square feet required in any such box having a capacity of 108 cu. ft. Draw a graph of $M(x)$.

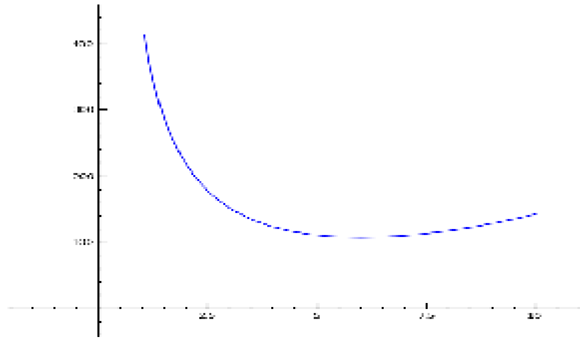


Figure : *SAGE* plot of $y = x^2 + \frac{432}{x}$, $1 < x < 10$.

What do we learn from the graph?

(a) If the box is carefully drawn, we may measure the ordinate corresponding to any length ($= x$) of the side of the square base and so determine the number of square feet of lumber required.

(b) There is one horizontal tangent (RS). The ordinate from its point of contact T is less than any other ordinate. Hence this discovery: One of the boxes evidently takes less lumber than any of the others. In other words, we may infer that the function defined by $M = M(x)$ has a minimum value. Let us find this point on the graph exactly, using our Calculus. Differentiating $M(x)$ to get the slope at any point, we have

$$\frac{dM}{dx} = 2x - \frac{432}{x^2}.$$

At the lowest point T the slope will be zero. Hence

$$2x - \frac{432}{x^2} = 0;$$

that is, when $x = 6$ the least amount of lumber will be needed.

Substituting in $M(x)$, we see that this is $M = 108$ sq. ft.

The fact that a least value of M exists is also shown by the following reasoning. Let the base increase from a very small square to a very large one. In the former case the height must be very great and therefore the amount of lumber required will be large. In the latter case, while the height is small, the base will take a great deal of lumber. Hence M varies from a large value, grows less, then increases again to another large value. It follows, then, that the graph must have a ``lowest" point corresponding to the dimensions which require the least amount of lumber, and therefore would involve the least cost.

Here is how to compute the critical points in *SAGE*:

[fontsize=\small,fontfamily=courier,fontshape=tt,frame=single,label=\sage]

[Type text]

```
sage: x = var("x")
sage: f = x^2 + 432/x
sage: solve(f.diff(x)==0,x)
[x == 3*sqrt(3)*I - 3, x == -3*sqrt(3)*I - 3, x == 6]
```

This says that $(x^2 + 432/x)' = 0$ has three roots, but only one real root - the one reported above at $x = 6$.

We will now proceed to the treatment in detail of the subject of maxima and minima.

Increasing and decreasing functions

A function is said to be *increasing* when it increases as the variable increases and decreases as the variable decreases. A function is said to be *decreasing* when it decreases as the variable increases and increases as the variable decreases.

The graph of a function indicates plainly whether it is increasing or decreasing.

Example (1) For instance, consider the function a^x whose graph is the locus of the equation $y = a^x$, $a > 1$:

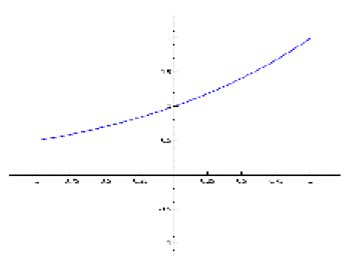


Figure: *SAGE* plot of $y = 2^x$, $-1 < x < 1$.

As we move along the curve from left to right the curve is rising; that is, as x increases the function ($= y$) always increases. Therefore a^x is an increasing function for all values of x.

(2) On the other hand, consider the function $(a-x)^3$ whose graph is the locus of the equation $y = (a-x)^3$.

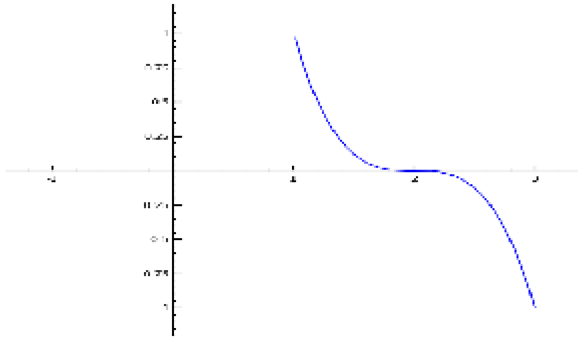


Figure : *SAGE* plot of $y=(2-x)^3$, $1<x<3$.

Now as we move along the curve from left to right the curve is falling; that is, as x increases, the function ($=y$) always decreases. Hence $(a-x)^3$ is a decreasing function for all values of x.

(3) That a function may be sometimes increasing and sometimes decreasing is shown by the graph of $y=2x^3-9x^2+12x-3$.

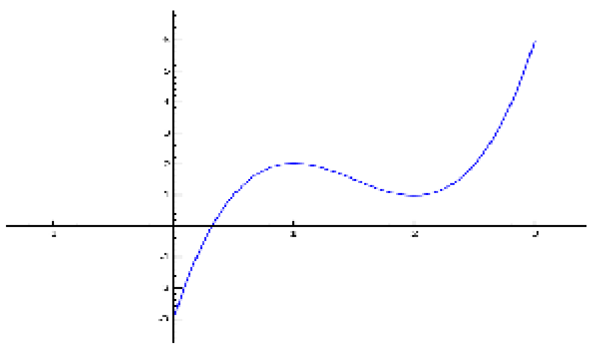


Figure : *SAGE* plot of $y=2x^3-9x^2+12x-3$, $0<x<3$.

As we move along the curve from left to right the curve rises until we reach the point A when $x=1$, then it falls from A to the point B when $x=2$, and to the right of B it is always rising. Hence

(a)

from $x=-\infty$ to $x=1$ the function is increasing;

(b)

from $x=1$ to $x=2$ the function is decreasing;

(c)

from $x = 2$ to $x = +\infty$ the function is increasing.

The student should study the curve carefully in order to note the behavior of the function when $x = 1$ and $x = 2$. Evidently A and B are turning points. At A the function ceases to increase and commences to decrease; at B, the reverse is true. At A and B the tangent (or curve) is evidently parallel to the x-axis, and therefore the slope is zero.

Tests for determining when a function is increasing or decreasing

It is evident from that at a point where a function

$y = f(x)$

is increasing, the tangent in general makes an acute angle with the x-axis; hence

slope $= \tan\tau = \frac{dy}{dx} = f'(x) =$ a positive number.

Similarly, at a point where a function is decreasing, the tangent in general makes an obtuse angle with the x-axis; therefore

slope $= \tan\tau = \frac{dy}{dx} = f'(x) =$ a negative number.

In order, then, that the function shall change from an increasing to a decreasing function, or vice versa, it is a necessary and sufficient condition that the first derivative shall change sign. But this can only happen for a continuous derivative by passing through the value zero. Thus in Figure as we pass along the curve the derivative (= slope) changes sign at the points where $x = 1$ and $x = 2$. In general, then, we have at ``turning points,"

$$\frac{dy}{dx} = f'(x) = 0.$$

A value of $y = f(x)$ satisfying this condition is called a *critical point* of the function $f(x)$. The derivative is continuous in nearly all our important applications, but it is interesting to note the case when the derivative (= slope) changes sign by passing through ∞. This would evidently happen at the points one a curve where the tangents (and curve) are perpendicular to the x-axis. At such exceptional critical points

$$\frac{dy}{dx} = f'(x) = \inf;$$

or, what amounts to the same thing,

$$\frac{1}{f'(x)} = 0.$$

Maximum and minimum values of a function

A *maximum value* of a function is one that is greater than any values immediately preceding or following. A *minimum value* of a function is one that is less than any values immediately preceding or following.

For example, it is clear that the function has a maximum value ($y = 2$) when $x = 1$, and a minimum value ($y = l$) when $x = 2$.

The student should observe that a maximum value is not necessarily the greatest possible value of a function nor a minimum value the least. It is seen that the function ($= y$) has values to the right of $x = 1$ that are greater than the maximum 2, and values to the left of $x = 1$ that are less than the minimum 1.

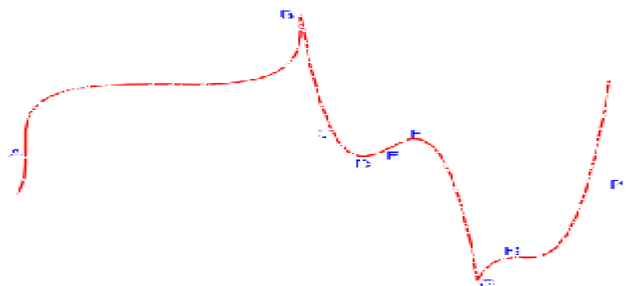


Figure : A continuous function.

A function may have several maximum and minimum values. Suppose that represents the graph of a function $f(x)$.

At B, F the function is at a local maximum, and at D, G a minimum. That some particular minimum value of a function may be greater than some particular maximum value is shown in the figure, the minimum value at D being greater than the maximum value at G.

At the ordinary critical points D, F, H the tangent (or curve) is parallel to the x-axis; therefore

$$\text{slope} = \frac{dy}{dx} = f'(x) = 0.$$

At the exceptional critical points A, B, G the tangent (or curve) is perpendicular to the x-axis, giving

$$\text{slope} = \frac{dy}{dx} = f'(x) = \infty.$$

One of these two conditions is then necessary in order that the function shall have a maximum or a minimum value. But such a condition is not sufficient; for at H the slope is zero and at A it is infinite, and yet the function has neither a maximum nor a minimum value at either point. It is necessary for us to know, in addition, how the function behaves in the neighborhood of each point. Thus at the points of maximum value, B, F, the function changes from an increasing to a decreasing function, and at the points of minimum value, D, G, the function changes from a decreasing to an increasing function. It therefore follows from that at maximum points

slope $= \frac{dy}{dx} = f'(x)$ must change from + to -,

and at minimum points

slope $= \frac{dy}{dx} = f'(x)$ must change from - to +

when we move along the curve from left to right.

At such points as A and H where the slope is zero or infinite, but which are neither maximum nor minimum points,

slope $= \frac{dy}{dx} = f'(x)$ does not change sign.

We may then state the conditions in general for maximum and minimum values of $f(x)$ for certain values of the variable as follows:

$$f(x) \text{ is a maximum if } f'(x) = 0, \text{ and } f'(x) \text{ changes from } + \text{ to } -. \tag{8.1}$$

$$f(x) \text{ is a minimum if } f'(x) = 0, \text{ and } f'(x) \text{ changes from } - \text{ to } +. \tag{8.2}$$

The values of the variable at the turning points of a function are called *critical values*; thus $x = 1$ and $x = 2$ are the critical values of the variable for the function whose graph is shown in Figure. The critical values at turning points where the tangent is parallel to the x-axis are evidently found by placing the first derivative equal to zero and solving for real values of x, just as under. (Similarly, if we wish to examine a function at exceptional turning points where the tangent is perpendicular to the x-axis, we set the reciprocal of the first derivative equal to zero and solve to find critical values.)

To determine the sign of the first derivative at points near a particular turning point, substitute in it, first, a value of the variable just a little less than the corresponding critical value, and then one a little greaterIf the first gives $+$ (as at L, Figure) and the second - (as at M), then the function ($= y$) has a maximum value in that interval (as at I). If the first gives $-$ (as at P) and the second $+$ (as at N), then the function ($=y$) has a minimum value in that interval (as at C).

If the sign is the same in both cases (as at Q and R), then the function ($=y$) has neither a maximum nor a minimum value in that interval (as at F).

We shall now summarize our results into a compact working rule.

Examining a function for extremal values: first method

Working rule:

- FIRST STEP. Find the first derivative of the function.
- SECOND STEP. Set the first derivative equal to zero[8.7] and solve the resulting equation for real roots in order to find the critical values of the variable.
- THIRD STEP. Write the derivative in factor form; if it is algebraic, write it in linear form.
- FOURTH STEP. Considering one critical value at a time, test the first derivative, first for a value a trifle less and then for a value a trifle greater than the critical value. If the sign of the derivative is first $+$ and then $-$, the function has a maximum value for that particular critical value of the variable; but if the reverse is true, then it has a minimum value. If the sign does not change, the function has neither.

Example In the problem worked out, we showed by means of the graph of the function

$$A = x\sqrt{100 - x^2}$$

that the rectangle of maximum area inscribed in a circle of radius 5 inches contained 50 square inches. This may now be proved analytically as follows by applying the above rule.

Solution. $f(x) = x\sqrt{100 - x^2}$.

First step. $f'(x) = \frac{100-2x^2}{\sqrt{100-x^2}}$.

Second step. $\frac{100-2x^2}{\sqrt{100-x^2}} = 0$ implies $x = 5\sqrt{2}$, which is the critical value. Only the positive sign of the radical is taken, since, from the nature of the problem, the negative sign has no meaning.

Third step. $f'(x) = \frac{2(5\sqrt{2}-x)(5\sqrt{2}+x)}{\sqrt{(10-x)(10+x)}}$.

Fourth step. When $x < 5\sqrt{2}$, $f'(x) = \frac{2(+)(+)}{\sqrt{(+)(+)}} = +$. When $x > 5\sqrt{2}$, $f'(x) = \frac{2(+)(+)}{\sqrt{(-)(+)}} = -$.

Since the sign of the first derivative changes from $+$ to $-$ at $x = 5\sqrt{2}$, the function has a maximum value

$$f(5\sqrt{2}) = 5\sqrt{2} \cdot 5\sqrt{2} = 50.$$

This tells us that $x_0 = 5\sqrt{2}$ is a critical point, at which the area is 50 square inches and at which the area changes from increasing to decreasing. This implies that the area is a maximum at this point

Problems

1. It is desired to make an open-top box of greatest possible volume from a square piece of tin whose side is a, by cutting equal squares out of the corners and then folding up the tin to form the sides. What should be the length of a side of the squares cut out?

   Solution. Let x = side of small square = depth of box; then $a - 2x$ = side of square forming bottom of box, and volume is $V = (a - 2x)^2 x$, which is the function to be made a maximum by varying x. Applying rule:

   First step. $\frac{dV}{dx} = (a - 2x)^2 - 4x(a - 2x) = a^2 - 8ax + 12x^2$.

   Second step. Solving $a^2 - 8ax + 12x^2 = 0$ gives critical values $x = \frac{a}{2}$ and $\frac{a}{6}$.

   It is evident that $x = \frac{a}{2}$ must give a minimum, for then all the tin would be cut away, leaving no material out of which to make a box. By the usual test, $x = \frac{a}{6}$ is found to give a maximum volume $\frac{2a^3}{27}$. Hence the side of the square to be cut out is one sixth of the side of the given square.

   The drawing of the graph of the function in this and the following problems is left to the student.

2. Assuming that the strength of a beam with rectangular cross section varies directly as the breadth and as the square of the depth, what are the dimensions of the strongest beam that can be sawed out of a round log whose diameter is d?

   Solution. If x = breadth and y = depth, then the beam will have maximum strength when the function xy^2 is a maximum. From the construction and the Pythagorean theorem, $y^2 = d^2 - x^2$; hence we should test the function

$f(x) = x(d^2 - x^2).$

First step. $f'(x) = -2x^2 + d^2 - x^2 = d^2 - 3x^2$.

Second step. $d^2 - 3x^2 = 0$. Therefore, $x = \frac{d}{\sqrt{3}}$ = critical value which gives a maximum.

Therefore, if the beam is cut so that depth = $\sqrt{\frac{2}{3}}$ of diameter of log, and breadth = $\sqrt{\frac{1}{3}}$ of diameter of log, the beam will have maximum strength.

3. What is the width of the rectangle of maximum area that can be inscribed in a given segment OAA' of a parabola?

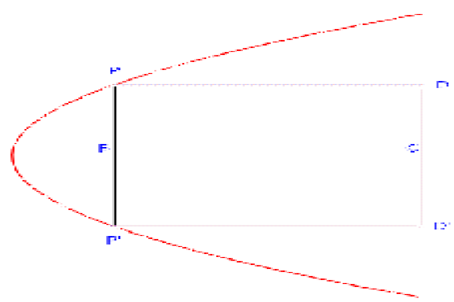


Figure 8.9: An inscribed rectangle in a parabola, $P = (x, y)$.

4. HINT. If $OC = h$, $BC = h - x$ and $PP' = 2y$; therefore the area of rectangle $PDD'P'$ is $2(h - x)y$.
5. But since P lies on the parabola $y^2 = 2px$, the function to be tested is $2(h - x)\sqrt{2px}$
6. Ans. Width = $\frac{2}{3}h$.
7. Find the altitude of the cone of maximum volume that can be inscribed in a sphere of radius r

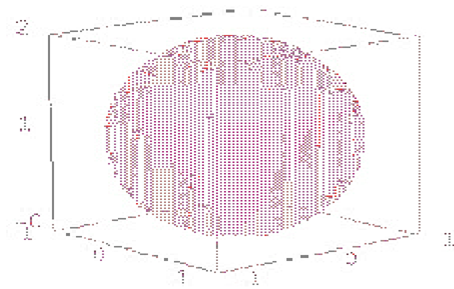


Figure An inscribed cone, height y and base radius x, in a sphere.

8. HINT. Volume of cone = $\frac{1}{3}\pi x^2 y$. But $x^2 = BC \times CD = y(2r-y)$; therefore the function to be tested is $f(y) = \frac{\pi}{3} y^2 (2r - y)$.
9. Ans. Altitude of cone = $\frac{4}{3}r$.
10. Find the altitude of the cylinder of maximum volume that can be inscribed in a given right cone

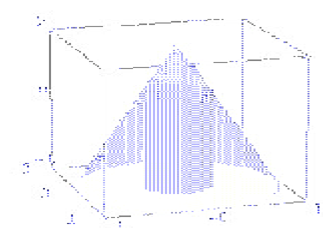


Figure: An inscribed cylinder in a cone.

11. HINT. Let $AU = r$ and $BC = h$. Volume of cylinder = $\pi x^2 y$. But from similar triangles ABC and DBG, $r/x = h/(h-y)$, so $x = \frac{r(h-y)}{h}$. Hence the function to be tested is $f(y) = \frac{r^2}{h^2} y (h-y)^2$.
12. Ans. Altitude = $\frac{1}{3}h$.
13. Divide a into two parts such that their product is a maximum.

    Ans. Each part $= \frac{a}{2}$.

14. Divide 10 into two such parts that the sum of the double of one and square of the other may be a minimum.

Ans. 9 and 1.

15. Find the number that exceeds its square by the greatest possible quantity.
Ans. $\frac{1}{2}$.

16. What number added to its reciprocal gives the least possible sum?
Ans. 1.

17. Assuming that the stiffness of a beam of rectangular cross section varies directly as the breadth and the cube of the depth, what must be the breadth of the stiffest beam that can be cut from a log 16 inches in diameter?
Ans. Breadth $= 8$ inches.

18. A water tank is to be constructed with a square base and open top, and is to hold 64 cubic yards. If the cost of the sides is \$ 1 a square yard, and of the bottom \$ 2 a square yard, what are the dimensions when the cost is a minimum? What is the minimum cost?
Ans. Side of base = 4 yd., height = 4 yd., cost \$ 96.

19. A rectangular tract of land is to be bought for the purpose of laying out a quarter-mile track with straightaway sides and semicircular ends. In addition a strip 35 yards wide along each straightaway is to be bought for grand stands, training quarters, etc. If the land costs \$ 200 an acre, what will be the maximum cost of the land required?
Ans. \$ 856.

20. A torpedo boat is anchored 9 miles from the nearest point of a beach, and it is desired to send a messenger in the shortest possible time to a military camp situated 15 miles from that point along the shore. If he can walk 5 miles an hour but row only 4 miles an hour, required the place he must land.
Ans. 3 miles from the camp.

21. A gas holder is a cylindrical vessel closed at the top and open at the bottom, where it sinks into the water. What should be its proportions for a given volume to require the least material (this would also give least weight)?
Ans. Diameter = double the height.

22. What should be the dimensions and weight of a gas holder of $\underline{8,000,000}$ cubic feet capacity, built in the most economical manner out of sheet iron $\frac{1}{16}$ of an inch thick and weighing $\frac{5}{2}$ lb. per sq. ft.?
Ans. Height = 137 ft., diameter = 273 ft., weight = 220 tons.

23. A sheet of paper is to contain 18 sq. in. of printed matter. The margins at the top and bottom are to be 2 inches each and at the sides 1 inch each. Determine the dimensions of the sheet which will require the least amount of paper.
Ans. 5 in. by 10 in.

24. A paper-box manufacturer has in stock a quantity of strawboard 30 inches by 14 inches. Out of this material he wishes to make open-top boxes by cutting equal squares out of each corner and then folding up to form the sides. Find the side of the square that should be cut out in order to give the boxes maximum volume.

Ans. 3 inches.

25. A roofer wishes to make an open gutter of maximum capacity whose bottom and sides are each 4 inches wide and whose sides have the same slope. What should be the width across the top?
Ans. 8 inches. 4

26. Assuming that the energy expended in driving a steamboat through the water varies as the cube of her velocity, find her most economical rate per hour when steaming against a current running c miles per hour.
HINT. Let v = most economical speed; then av^3 = energy expended each hour, a being a constant depending upon the particular conditions, and $v - c$ = actual distance advanced per hour. Hence $\frac{av^3}{v-c}$ is the energy expended per mile of distance advanced, and it is therefore the function whose minimum is wanted.

27. Prove that a conical tent of a given capacity will require the least amount of canvas when the height is $\sqrt{2}$ times the radius of the base. Show that when the canvas is laid out flat it will be a circle with a sector of $152^0 9' = 2.6555...$ cut out. A bell tent 10 ft. high should then have a base of diameter 14 ft. and would require 272 sq. ft. of canvas.
28. A cylindrical steam boiler is to be constructed having a capacity of 1000 cu. ft. The material for the side costs $ 2 a square foot, and for the ends $ 3 a square foot. Find radius when the cost is the least.
Ans. $\frac{1}{\sqrt[3]{3\pi}}$ ft.

29. In the corner of a field bounded by two perpendicular roads a spring is situated 6 rods from one road and 8 rods from the other.
(a) How should a straight road be run by this spring and across the corner so as to cut off as little of the field as possible?

(b) What would be the length of the shortest road that could be run across?

Ans. (a) 12 and 16 rods from corner. (b) $(6^{\frac{2}{3}} + 8^{\frac{2}{3}})^{\frac{3}{2}}$ rods.

30. Show that a square is the rectangle of maximum perimeter that can be inscribed in a given circle.
31. Two poles of height a and b feet are standing upright and are c feet apart. Find the point on the line joining their bases such that the sum of the squares of the distances from this point to the tops of the poles is a minimum. (Ans. Midway between the poles.) When will the sum of these distances be a minimum?
32. A conical tank with open top is to be built to contain V cubic feet. Determine the shape if the material used is a minimum.
33. An isosceles triangle has a base 12 in. long and altitude 10 in. Find the rectangle of maximum area that can be inscribed in it, one side of the rectangle coinciding with the base of the triangle.
34. Divide the number 4 into two such parts that the sum of the cube of one part and three times the square of the other shall have a maximum value.

35. Divide the number a into two parts such that the product of one part by the fourth power of the other part shall be a maximum.
36. A can buoy in the form of a double cone is to be made from two equal circular iron plates of radius r. Find the radius of the base of the cone when the buoy has the greatest displacement (maximum volume).

    Ans. $r\sqrt{\frac{2}{3}}$.

37. Into a full conical wineglass of depth a and generating angle α there is carefully dropped a sphere of such size as to cause the greatest overflow. Show that the radius of the sphere is $\frac{a \sin \alpha}{\sin \alpha \cos 2\alpha}$.
38. A wall 27 ft. high is 8 ft. from a house. Find the length of the shortest ladder that will reach the house if one end rests on the ground outside of the wall.

    Ans. $13\sqrt{13}$.

    Here's how to solve this using *SAGE*: Let h be the height above ground at which the ladder hits the house and let d be the distance from the wall that the ladder hits the ground on the other side of the wall. By similar triangles, $h/27 = (8+d)/d = 1 + \frac{8}{d}$, so $d + 8 = 8\frac{h}{h-27}$. The length of the ladder is, by the Pythagorean theorem, $f(h) = \sqrt{h^2 + (8+d)^2} = \sqrt{h^2 + (8\frac{h}{h-27})^2}$.

    [fontsize=\small,fontfamily=courier,fontshape=tt,frame=single,label=\sage]

```
sage: h = var("h")

sage: f(h) = sqrt(h^2+(8*h/(h-27))^2)

sage: f1(h) = diff(f(h),h)

sage: f2(h) = diff(f(h),h,2)

sage: crit_pts = solve(f1(h) == 0,h); crit_pts

[h == 21 - 6*sqrt(3)*I, h == 6*sqrt(3)*I + 21, h == 39, h == 0]

sage: h0 = crit_pts[2].rhs(); h0

39

sage: f(h0)

13*sqrt(13)

sage: f2(h0)
```

3/(4*sqrt(13))

This says $f(h)$ has four critical points, but only one of which is meaningful, $h_0 = 39$. At this point, $f(h)$ is a minimum.

39. A vessel is anchored 3miles offshore, and opposite a point 5miles further along the shore another vessel is anchored 9miles from the shore. A boat from the first vessel is to land a passenger on the shore and then proceed to the other vessel. What is the shortest course of the boat?
Ans. 13miles.

40. A steel girder 25ft. long is moved on rollers along a passageway 12.8ft. wide and into a corridor at right angles to the passageway. Neglecting the width of the girder, how wide must the corridor be?
Ans. 5.4ft.

41. A miner wishes to dig a tunnel from a point A to a point B 300feet below and 500feet to the east of A. Below the level of A it is bed rock and above A is soft earth. If the cost of tunneling through earth is $ 1 and through rock $ 3 per linear foot, find the minimum cost of a tunnel.
Ans. $ 1348.53.

42. A carpenter has 108sq. ft. of lumber with which to build a box with a square base and open top. Find the dimensions of the largest possible box he can make.
Ans. $\underline{6 \times 6 \times 3}$.

43. Find the right triangle of maximum area that can be constructed on a line of length $h$as hypotenuse.
Ans. $\frac{h}{\sqrt{2}}$ = length of both legs.

44. What is the isosceles triangle of maximum area that can be inscribed in a given circle?
Ans. An equilateral triangle.

45. Find the altitude of the maximum rectangle that can be inscribed in a right triangle with base $b$and altitude h.
Ans. Altitude = $\frac{h}{2}$.

46. Find the dimensions of the rectangle of maximum area that can be inscribed in the ellipse $b^2x^2 + a^2y^2 = a^2b^2$.
Ans. $a\sqrt{2} \times b\sqrt{2}$; area = $2ab$.

47. Find the altitude of the right cylinder of maximum volume that can be inscribed in a sphere of radius r.
Ans. Altitude of cylinder = $\frac{2r}{\sqrt{3}}$.

48. Find the altitude of the right cylinder of maximum convex (curved) surface that can be inscribed in a given sphere.
Ans. Altitude of cylinder = $r\sqrt{2}$.

49. What are the dimensions of the right hexagonal prism of minimum surface whose volume is 36 cubic feet?
Ans. Altitude = $2\sqrt{3}$; side of hexagon = 2.

50. Find the altitude of the right cone of minimum volume circumscribed about a given sphere.
Ans. Altitude = $4r$, and volume = $2\times$ vol. of sphere.

51. A right cone of maximum volume is inscribed in a given right cone, the vertex of the inside cone being at the center of the base of the given cone. Show that the altitude of the inside cone is one third the altitude of the given cone.

52. Given a point on the axis of the parabola $y^2 = 2px$ at a distance a from the vertex; find the abscissa of the point of the curve nearest to it.
Ans. $x = a - p$.

53. What is the length of the shortest line that can be drawn tangent to the ellipse $b^2x^2 + a^2y^2 = a^2b^2$ and meeting the coordinate axes?
Ans. $a+b$.

54. A Norman window consists of a rectangle surmounted by a semicircle. Given the perimeter, required the height and breadth of the window when the quantity of light admitted is a maximum.
Ans. Radius of circle = height of rectangle.

55. A tapestry 7 feet in height is hung on a wall so that its lower edge is 9 feet above an observer's eye. At what distance from the wall should he stand in order to obtain the most favorable view? (HINT. The vertical angle subtended by the tapestry in the eye of the observer must be at a maximum.)
Ans. 12 feet.

56. What are the most economical proportions of a tin can which shall have a given capacity, making allowance for waste? (HINT. There is no waste in cutting out tin for the side of the can, but for top and bottom a hexagon of tin circumscribing the circular pieces required is used up. NOTE 1. If no allowance is made for waste, then height = diameter. NOTE 2. We know that the shape of a bee cell is hexagonal, giving a certain capacity for honey with the greatest possible economy of wax.)
Ans. Height = $\frac{2\sqrt{3}}{\pi}\times$ diameter of base.

57. An open cylindrical trough is constructed by bending a given sheet of tin at breadth $2a$. Find the radius of the cylinder of which the trough forms a part when the capacity of the trough is a maximum.

Ans. Rad. = $\frac{2a}{\pi}$; i.e. it must be bent in the form of a semicircle.

58. A weight W is to be raised by means of a lever with the force F at one end and the point of support at the other. If the weight is suspended from a point at a distance a from the point of support, and the weight of the beam is w pounds per linear foot, what should be the length of the lever in order that the force required to lift it shall be a minimum?

Ans. $x = \sqrt{\frac{2aW}{w}}$ feet.

59. An electric arc light is to be placed directly over the center of a circular plot of grass 100 feet in diameter. Assuming that the intensity of light varies directly as the sine of the angle under which it strikes an illuminated surface, and inversely as the square of its distance from the surface, how high should the light he hung in order that the best possible light shall fall on a walk along the circumference of the plot?

Ans. $\frac{50}{\sqrt{2}}$ feet

60. The lower corner of a leaf, whose width is a, is folded over so as just to reach the inner edge of the page.

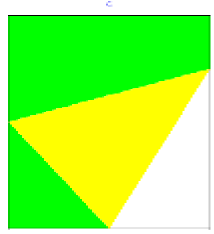


Figure: A leafed page of width a.

61. (a) Find the width of the part folded over when the length of the crease is a minimum.
62. (b) Find the width when the area folded over is a minimum.
63. Ans. (a) $\frac{3}{4}a$; (b) $\frac{2}{3}a$.
64. A rectangular stockade is to be built which must have a certain area. If a stone wall already constructed is available for one of the sides, find the dimensions which would make the cost of construction the least.
    Ans. Side parallel to wall = twice the length of each end.

65. When the resistance of air is taken into account, the inclination of a pendulum to the vertical may be given by the formula $\theta = ae^{-kt}\cos(nt+\eta)$. Show that the greatest elongations occur at equal intervals $\frac{\pi}{n}$ of time.

66. It is required to measure a certain unknown magnitude x with precision. Suppose that n equally careful observations of the magnitude are made, giving the results $a_1, a_2, a_3, \dots, a_n$. The errors of these observations are evidently $x - a_1, x - a_2, x - a_3, \cdots, x - a_n$, some of which are positive and some negative. It has been agreed that the most probable value of x is such that it renders the sum of the squares of the errors, namely $(x-a_1)^2 + (x-a_2)^2 + (x-a_3)^2 + \cdots + (x-a_n)^2$, a minimum. Show that this gives the arithmetical mean of the observations as the most probable value of x.
(This is related to the method of least squares, discovered by Gauss, a commonly used technique in statistical applications.)

67. The bending moment at x of a beam of length ℓ, uniformly loaded, is given by the formula $M = \frac{1}{2}w\ell x - \frac{1}{2}wx^2$, where w= load per unit length. Show that the maximum bending moment is at the center of the beam.

68. If the total waste per mile in an electric conductor is $W = c^2 r + \frac{t^2}{r}$, where c= current in amperes (a constant), r= resistance in ohms per mile, and t= a constant depending on the interest on the investment and the depreciation of the plant, what is the relation between c, r, and t when the waste is a minimum?
Ans. $cr = t$.

69. A submarine telegraph cable consists of a core of copper wires with a covering made of nonconducting material. If x denote the ratio of the radius of the core to the thickness of the covering, it is known that the speed of signaling varies as

$$x^2 \log \frac{1}{x}.$$

Show that the greatest speed is attained when $x = \frac{1}{\sqrt{e}}$.

70. Assuming that the power given out by a voltaic cell is given by the formula

$$P = \frac{E^2 R}{(r+R)^2},$$

when E= constant electromotive force, r= constant internal resistance, R= external resistance, prove that P is a maximum when $r = R$.

71. The force exerted by a circular electric current of radius a on a small magnet whose axis coincides with the axis of the circle varies as

$$\frac{x}{(a^2+x^2)^{\frac{5}{2}}}.$$

where x= distance of magnet from plane of circle. Prove that the force is a maximum when $x = \frac{a}{2}$.

72. We have two sources of heat at A and B, which we visualize on the real line (with B to the right or A), with intensities a and b respectively. The total intensity of heat at a point P between A and B at a distance of x from A is given by the formula $I = \frac{a}{x^2} + \frac{b}{(d-x)^2}$. Show that the temperature at P will be the lowest when $\frac{d-x}{x} = \frac{\sqrt[3]{b}}{\sqrt[3]{a}}$. that is, the distances BP and AP have the same ratio as the cube roots of the corresponding heat intensities. The distance of P from A is $x = \frac{a^{\frac{1}{3}} d}{a^{\frac{1}{3}} + b^{\frac{1}{3}}}$.

73. The range of a projectile in a vacuum is given by the formula $R = \frac{v_0^2 \sin 2\phi}{g}$, where v_0 = initial velocity, g = acceleration due to gravity, ϕ = angle of projection with the horizontal. Find the angle of projection which gives the greatest range for a given initial velocity.
Ans. $\phi = 45^o = \pi/4$.

74. The total time of flight of the projectile in the last problem is given by the formula $T = \frac{2v_0 \sin \phi}{g}$. At what angle should it be projected in order to make the time of flight a maximum?
Ans. $\phi = 90^o = \pi/2$.

75. The time it takes a ball to roll down an inclined plane with angle ϕ (with respect to the x-axis) is given by the formula $T = 2\sqrt{\frac{2}{g \sin 2\phi}}$. Neglecting friction, etc., what must be the value of ϕ to make the quickest descent?
Ans. $\phi = 45^o = \pi/4$.

76. Examine the function $(x-1)^2(x+1)^3$ for maximum and minimum values. Use the first method.
Solution. $f(x) = (x-1)^2(x+1)^3$.

First step. $f'(x) = 2(x-1)(x+1)^3 + 3(x-1)^2(x+1)^2 = (x-1)(x+1)^2(5x-1)$.

Second step. $(x-1)(x+1)^2(5x-1) = 0$, $x = 1, -1, \frac{1}{5}$, which are critical values.

Third step. $f'(x) = 5(x-1)(x+1)^2(x - \frac{1}{5})$.

Fourth step. Examine first for critical value $x = 1$.

When $x < 1$, $f'(x) = 5(-)(+)2(+) = -$. When $x > 1$, $f'(x) = 5(+)(+)2(+) = +$.

Therefore, when $x = 1$ the function has a minimum value $f(l) = 0$. Examine now for the critical value $x = \frac{1}{5}$. When $x < \frac{1}{5}$, $f'(x) = 5(-)(+)^2(-) = +$. When $x > \frac{1}{5}$, $f'(x) = 5(-)(+)^2(+) = -$. Therefore, when $x = \frac{1}{5}$ the function has a maximum value $f(\frac{1}{5}) = 1.11$. Examine lastly for the critical value $x = -1$. When $x < -1$, $f'(x) = 5(-)(-)2(-) = +$. When $x > -1$, $f'(x) = 5(-)(+)2(-) = +$. Therefore, when $x = -1$ the function has neither a maximum nor a minimum value.

77.

Examine the following functions for maximum and minimum values:

1. $(x - 3)^2(x - 2)$.

   Ans. $x = \frac{7}{3}$, gives max. = $\frac{4}{27}$; $x = 3$, gives min. = 0.

2. $(x - 1)^3(x - 2)^2$.

   Ans. $x = \frac{8}{5}$, gives max. = 0.03456; $x = 2$, gives min. = 0; $x = 1$, gives neither.

3. $(x - 4)^5(x + 2)^4$.

   Ans. $x = -2$, gives max.; $x = \frac{2}{3}$ gives min; $x = 4$, gives neither.

4. $(x - 2)^5(2x + 1)^4$.

   Ans. $x = -\frac{1}{2}$, gives max.; $x = \frac{11}{18}$, gives min.; $x = 2$, gives neither.

5. $(x + 1)^{\frac{2}{3}}(x - 5)^2$.

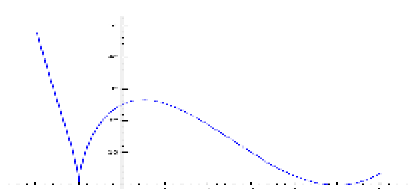


Figure: *SAGE* plot of $y = (x+1)^{\frac{2}{3}}(x-5)^{2}$.

6. Ans. $x = \frac{1}{2}$, gives max.; $x = -1$ and 5, give min.

7. $(2x-a)^{\frac{1}{3}}(x-a)^{\frac{2}{3}}$.

   Ans. $x = \frac{2a}{3}$, gives max.; $x = 1$ and $-\frac{1}{3}$, gives min.; $x = \frac{a}{2}$, gives neither.

8. $x(x-1)^2(x+1)^3$.

   Ans. $x = \frac{1}{2}$, gives max.; $x = 1$ and $-\frac{1}{3}$, gives min.; $x = -1$, gives neither.

9. $x(a+x)^2(a-x)^3$

   Ans. $x = -a$ and $\frac{a}{3}$, give max.; $x = -\frac{a}{2}$; $x = a$, gives neither.

10. $b + c(x-c)^{\frac{2}{3}}$.

    Ans. $x = a$, gives min. = b.

11. $a - b(x-c)^{\frac{1}{3}}$.

    Ans. No max. or min.

12. $\frac{x^2-7x+6}{x-10}$.

    Ans. $x = 4$, gives max. $x = 16$, gives min.

13. $\frac{(a-x)^3}{a-2x}$.

    Ans. $x = \frac{a}{4}$, gives min.

14. $\frac{1-x+x^2}{1+x-x^2}$.

    Ans. $x = \frac{1}{2}$, gives min.

15. $\frac{x^2-3x+2}{x^2+3x+2}$.

    Ans. $x = \sqrt{2}$, gives min. = $12\sqrt{2} - 17$; $x = -\sqrt{2}$, gives max. = $-12\sqrt{2} - 17$; $x = -1, -2$, give neither.

16. $\frac{(x-a)(b-x)}{x^2}$.

$x = \frac{2ab}{a+b}$, gives max. = $\frac{(a-b)^2}{4ab}$.

17. $\frac{a^2}{x} + \frac{b^2}{a-x}$.

Ans. $x = \frac{a^2}{a-b}$, gives min.; $x = \frac{a^2}{a+b}$, gives max.

18. Examine $x^3 - 3x^2 - 9x + 5$ for maxima and minima, Use the second method,

Solution. $f(x) = x^3 - 3x^2 - 9x + 5$.

First step. $f'(x) = 3x^2 - 6x - 9$.

Second step, $3x^2 - 6x - 9 = 0$; hence the critical values are $x = -1$ and 3.

Third step. $f''(x) = 6x - 6$.

Fourth step. $f''(-1) = -12$.

Therefore, $f(-1) = 10 =$ maximum value. $f''(3) = +12$. Therefore, $f(3) = -22 =$ minimum value.

19. Examine $\sin^2 x \cos x$ for maximum and minimum values.

Solution. $f(x) = \sin^2 x \cos x$.

First step. $f'(x) = 2\sin x \cos^2 x - \sin^3 x$.

Second step. $2\sin x \cos^2 x - \sin^3 x = 0$; hence the critical values are $x = n\pi$ and $x = n\pi \pm \arctan(-\sqrt{2}) = n\pi \pm \alpha$.

Third step. $f''(x) = \cos x(2\cos^2 x - 7\sin^2 x)$.

Fourth step. $f''(0) = +$. Therefore, $f(0) = 0$ = minimum value. $f''(\pi) = -$. Therefore, $f(\pi) = 0$ = maximum value. $f''(\alpha) = -$. Therefore, $f(\alpha)$ maximum value. $f''(\pi - \alpha) = +$. Therefore, $f(\pi - \alpha)$ minimum value.

Examine the following functions for maximum and minimum values:

1. $3x^3 - 9x^2 - 27x + 30$.

   Ans. $x = -1$, gives max. = 45; $x = 3$, gives min. = -51.

2. $2x^3 - 21x^2 + 36x - 20$.

   Ans. $x = 1$, gives max. = -3; $x = 6$, gives min. = -128.

3. $\frac{x^3}{3} - 21x^2 + 3x + 1$.

   Ans. $x = 1$, gives max. = [illegible]; $x = 3$, gives min. = 1.

4. $2x^3 - 15x^2 + 36x + 10$.

   Ans. $x = 2$, gives max. = 38; $x = 3$, gives min. = 37.

5. $x^3 - 9x^2 + 15x - 3$.

   Ans. $x = 1$, gives max. = 4; $x = 5$, gives min. = -28.

6. $x^3 - 3x^2 + 6x + 10$.

   Ans. No max. or min.

7. $x^5 - 5x^4 + 5x^3 + 1$. $x = 1$, gives max. = 2; $x = 3$, gives min. = -26; $x = 0$, gives neither.

8. $3x^5 - 125x^2 + 2160x$.

   $x = -4$ and 3, give max.; $x = -3$ and 4, give min.

9. $2x^3 - 3x^2 - 12x + 4$.

10. $2x^3 - 21x^2 + 36x - 20$.

11. $x^4 - 2x^2 + 10$.

12. $x^4 - 4$.

13. $x^3 - 8$.

14. $4 - x^6$.

15. $\sin x(1 + \cos x)$.

Ans. $x = 2n\pi + \frac{\pi}{3}$, give max. $= \frac{3}{4}\sqrt{3}$; $x = 2n\pi - \frac{\pi}{3}$, give min. $= \frac{3}{4}\sqrt{3}$; $x = n\pi$, give neither.

16. $\frac{x}{\log x}$.

Ans. $x = e$, gives min. = e; $x = 1$, gives neither.

17. $\log \cos x$.

Ans. $x = n\pi$, gives max.

18. $ae^{kx} + be^{-kx}$.

Ans. $x = \frac{1}{k} \log \sqrt{\frac{b}{a}}$, gives min. $= 2\sqrt{ab}$.

19. x^x.

$x = \frac{1}{e}$, gives min.

20. $x^{\frac{1}{x}}$.

Ans. $x = e$, gives max.

21. $\cos x + \sin x$.

Ans. $x = \frac{\pi}{4}$, gives max. = $\sqrt{2}$. $x = \frac{5\pi}{4}$, gives min. = $-\sqrt{2}$.

22. $\sin 2x - x$.

Ans. $x = \frac{\pi}{6}$, gives max.; $x = -\frac{\pi}{6}$, gives min.

23. $x + \tan x$.

Ans. No max. or min.

24. $\sin^3 x \cos x$.

Ans. $x = n\pi + \frac{\pi}{3}$, gives max. = $\frac{3}{16}\sqrt{3}$; $x = n\pi - \frac{\pi}{3}$, gives min. = $-\frac{3}{16}\sqrt{3}$; $x = n\pi$, gives neither.

25. $x \cos x$.

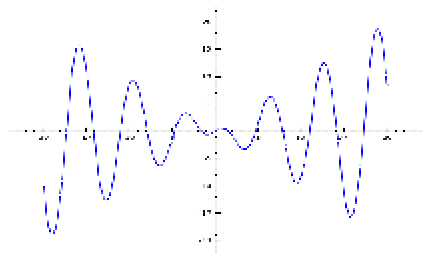


Figure: *SAGE* plot of $y = x\cos(x)$.

26. Ans. x such that $x \sin x = \cos x$, gives max/min.

27. $\sin x + \cos 2x$.

Ans. $\arcsin \frac{1}{4}$, gives max.; $x = \frac{\pi}{2}$, gives min.

28. $2\tan x - \tan^2 x$.

Ans. $x = \frac{\pi}{4}$, gives max.

29. $\frac{\sin x}{1+\tan x}$.

Ans. $x = \frac{\pi}{4}$, gives max.

30. $\frac{x}{1+x\tan x}$.

$x = \cos x$, gives max.; $x = -\cos x$, gives min.

Definition of successive derivatives

We have seen that the derivative of a function of x is in general also a function of x. This new function may also be differentiable, in which case the derivative of the first derivative is called the second derivative of the original function. Similarly, the derivative of the second derivative is called the third derivative; and so on to the n-th derivative. Thus, if

$$\begin{aligned} y &= 3x^4, \\ \frac{dy}{dx} &= 12x^3, \\ \frac{d}{dx}\left(\frac{dy}{dx}\right) &= 36x^2, \\ \frac{d}{dx}\left[\frac{d}{dx}\left(\frac{dy}{dx}\right)\right] &= 72x, \end{aligned}$$

Notation

The symbols for the successive derivatives are usually abbreviated as follows:

$$\begin{aligned}
\frac{d}{dx}\left(\frac{dy}{dx}\right) &= \frac{d^2y}{dx^2}, \\
\frac{d}{dx}\left[\frac{d}{dx}\left(\frac{dy}{dx}\right)\right] &= \frac{d}{dx}\left(\frac{d^2y}{dx^2}\right) = \frac{d^3y}{dx^3}, \\
\dots &\quad \dots \\
\frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) &= \frac{d^ny}{dx^n}.
\end{aligned}$$

If $y = f(x)$, the successive derivatives are also denoted by

$$f'(x),\ f''(x),\ f'''(x),\ f^{(4)}(x),\ \dots,\ f^{(n)}(x);$$

or

$$y',\ y'',\ y''',\ y^{(4)},\ \dots,\ y^{(n)};$$

or,

$$\frac{d}{dx}f(x),\ \frac{d^2}{dx^2}f(x),\ \frac{d^3}{dx^3}f(x),\ \frac{d^4}{dx^4}f(x),\ \dots,\ \frac{d^n}{dx^n}f(x).$$

The n-th derivative

For certain functions a general expression involving n may be found for the n-th derivative. The usual plan is to find a number of the first successive derivatives, as many as may be necessary to discover their law of formation, and then by induction write down the n-th derivative.

Example Given $y = e^{ax}$, find $\frac{d^ny}{dx^n}$.

Solution. $\frac{dy}{dx} = ae^{ax}$, $\frac{d^2y}{dx^2} = a^2e^{ax}$, ..., $\frac{d^ny}{dx^n} = a^ne^{ax}$.

Example Given $y = \log x$, find $\frac{d^ny}{dx^n}$.

Solution. $\frac{dy}{dx} = \frac{1}{x}$, $\frac{d^2y}{dx^2} = -\frac{1}{x^2}$, $\frac{d^3y}{dx^3} = \frac{1\cdot 2}{x^3}$, $\frac{d^4y}{dx^4} = \frac{1\cdot 2\cdot 3}{x^4}$, ... $\frac{d^ny}{dx^n} = (-1)^{n-1}\frac{(n-1)!}{x^n}$.

Example Given $y = \sin x$, find $\frac{d^ny}{dx^n}$.

Solution. $\frac{dy}{dx} = \cos x = \sin\left(x + \frac{\pi}{2}\right)$,

[Type text]

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\sin\left(x+\frac{\pi}{2}\right) = \cos\left(x+\frac{\pi}{2}\right) = \sin\left(x+\frac{2\pi}{2}\right),$$

$$\frac{d^3y}{dx^3} = \frac{d}{dx}\sin\left(x+\frac{2\pi}{2}\right) = \cos\left(x+\frac{2\pi}{2}\right) = \sin\left(x+\frac{3\pi}{2}\right)$$

. . .

$$\frac{d^ny}{dx^n} = \sin\left(x+\frac{n\pi}{2}\right).$$

Leibnitz's Formula for the n-th derivative of a product

This formula expresses the n-th derivative of the product of two variables in terms of the variables themselves and their successive derivatives.

If u and v are functions of x, we have, from equation (V) in above,

$$\frac{d}{dx}(uv) = \frac{du}{dx}v + u\frac{dv}{dx}.$$

Differentiating again with respect to x,

$$\frac{d^2}{dx^2}(uv) = \frac{d^2u}{dx^2}v + \frac{du}{dx}\frac{dv}{dx} + \frac{du}{dx}\frac{dv}{dx} + u\frac{d^2v}{dx^2} = \frac{d^2u}{dx^2}v + 2\frac{du}{dx}\frac{dv}{dx} + u\frac{d^2v}{dx^2}.$$

Similarly,

$$\begin{aligned}\frac{d^3}{dx^3}(uv) &= \frac{d^3u}{dx^3} + \frac{d^2u}{dx^2}\frac{dv}{dx} + 2\frac{d^2u}{dx^2}\frac{dv}{dx} + 2\frac{du}{dx}\frac{d^2v}{dx^2} + \frac{du}{dx}\frac{d^2v}{dx^2} + u\frac{d^3v}{dx^3}\\ &= \frac{d^3u}{dx^3}v + 3\frac{d^2u}{dx^2}\frac{dv}{dx} + 3\frac{du}{dx}\frac{d^2v}{dx^2} + u\frac{d^3v}{dx^3}.\end{aligned}$$

However far this process may be continued, it will be seen that the numerical coefficients follow the same law as those of the Binomial Theorem, and the indices of the derivatives correspond[7.1] to the exponents of the Binomial Theorem. Reasoning then by mathematical induction from the m-th to the $(m+1)$-st derivative of the product, we can prove *Leibnitz's Formula*

$$\frac{d^n}{dx^n}(uv) = \frac{d^nu}{dx^n}v + n\frac{d^{n-1}u}{dx^{n-1}}\frac{dv}{dx} + \frac{n(n-1)}{2!}\frac{d^{n-2}u}{dx^{n-2}}\frac{d^2v}{dx^2} + \cdots + n\frac{du}{dx}\frac{d^{n-1}v}{dx^{n-1}} + u\frac{d^nv}{dx^n}, \qquad (7.1)$$

Example Given $y = e^x \log x$, find $\frac{d^3y}{dx^3}$ by Leibnitz's Formula.

Solution. Let $u = e^x$, and $v = \log x$; then $\frac{du}{dx} = e^x$, $\frac{dv}{dx} = \frac{1}{x}$, $\frac{d^2u}{dx^2} = e^x$, $\frac{d^2v}{dx^2} = -\frac{1}{x^2}$, $\frac{d^3u}{dx^3} = e^x$, $\frac{d^3v}{dx^3} = \frac{2}{x^3}$.

Substituting in we get

$$\frac{d^3y}{dx^3} = e^x \log x + \frac{3e^x}{x} - \frac{3e^x}{x^2} = e^x \left(\log x + \frac{3}{x} - \frac{3}{x^2} + \frac{2}{x^3} \right).$$

This can be verified using the *SAGE* commands:

[fontsize=\small,fontfamily=courier,fontshape=tt,frame=single,label=\sage]

```
sage: x = var("x")
sage: f = exp(x)*log(x)
sage: diff(f,x,3)
e^x*log(x) + 3*e^x/x - 3*e^x/x^2 + 2*e^x/x^3
```

Example Given $y = x^2 e^{ax}$, find $\frac{d^ny}{dx^n}$ by Leibnitz's Formula.

Solution. Let $u = x^2$, and $v = e^{ax}$; then $\frac{du}{dx} = 2x$, $\frac{dv}{dx} = ae^{ax}$, $\frac{d^2u}{dx^2} = 2x$, $\frac{d^2v}{dx^2} = a^2e^{ax}$, $\frac{d^3u}{dx^3} = 0$, $\frac{d^3v}{dx^3} = a^3e^{ax}$, ..., $\frac{d^nu}{dx^n} = 0$, $\frac{d^nv}{dx^n} = a^ne^{ax}$. Substituting in (7.1), we get

$$\frac{d^ny}{dx^n} = x^2a^ne^{ax} + 2na^{n-1}xe^{ax} + n(n-1)a^{n-2}e^{ax} = a^{n-2}e^{ax}[x^2a^2 + 2nax + n(n-1)].$$

Successive differentiation of implicit functions

To illustrate the process we shall find $\frac{d^2y}{dx^2}$ from the equation of the hyperbola

$$b^2x^2 - a^2y^2 = a^2b^2.$$

Differentiating with respect to x, as ,

$$2b^2x - 2a^2y\frac{dy}{dx} = 0,$$

or,

[Type text]

$$\frac{dy}{dx} = \frac{b^2x}{a^2y}. \tag{7.2}$$

Differentiating again, remembering that y is a function of x,

$$\frac{d^2y}{dx^2} = \frac{a^2yb^2 - b^2xa^2\frac{dy}{dx}}{a^4y^2}.$$

Substituting for $\frac{dy}{dx}$ its value from

$$\frac{d^2y}{dx^2} = \frac{a^2b^2y - a^2b^2x\left(\frac{b^2y}{a^2y}\right)}{a^4y^2} = -\frac{b^2(b^2x^2 - a^2y^2)}{a^4y^3}.$$

The given equation, $b^2x^2 - a^2y^2 = a^2b^2$, therefore gives,

$$\frac{d^2y}{dx^2} = -\frac{b^4}{a^2y^3}.$$

This basically says

$$y' = \frac{dy}{dx} = \frac{b^2x}{a^2y},$$

and

$$y'' = \frac{d^2y}{dx^2} = -\frac{b^2 - a^2(y')^2}{a^2y}.$$

Exercises

Verify the following derivatives:

1. $y = 4x^3 - 6x^2 + 4x + 7$.

   Ans. $\frac{d^2y}{dx^2} = 12(2x - 1)$.

2. $f(x) = \frac{x^3}{1-x}$.

Ans. $f^{(4)}(x) = \frac{4!}{(1-x)^5}$.

3. $f(y) = y^6$.

Ans. $f^{(6)}(y) = 6!$.

4. $y = x^3 \log x$.

Ans. $\frac{d^4y}{dx^4} = \frac{6}{x}$.

5. $y = \frac{c}{x^n}$. $y'' = \frac{n(n+1)c}{x^{n+2}}$.

6. $y = (x-3)e^{2x} + 4xe^x + x$.

Ans. $y'' = 4e^x[(x-2)e^x + x + 2]$.

7. $y = \frac{a}{2}(e^{\frac{x}{a}} + e^{-\frac{x}{a}})$.

Ans. $y'' = \frac{1}{2a}(e^{\frac{x}{a}} + e^{-\frac{x}{a}}) = \frac{y}{a^2}$.

8. $f(x) = ax^2 + bx + c$.

Ans. $f'''(x) = 0$.

9. $f(x) = \log(x+1)$.

Ans. $f^{(4)}(x) = -\frac{6}{(x+1)^4}$.

10. $f(x) = log(e^x + e^{-x})$.

Ans. $f'''(x) = -\frac{8(e^x - e^{-x})}{(e^x - e^{-x})^3}$.

11. $r = \sin a\theta$.

Ans. $\frac{d^4r}{d\theta^4} = a^4 \sin a\theta = a^4 r$.

12. $r = \tan\phi$.

Ans. $\frac{d^3r}{d\phi^3} = 6\sec^6\phi - 4\sec^2\phi$.

13. $r = \log\sin\phi$.

Ans. $r''' = 2\cot\phi\csc^2\phi$.

14. $f(t) = e^{-t}\cos t$.

Ans. $f^{(4)}(t) = -4e^{-t}\cos t = -4f(t)$.

15. $f(\theta) = \sqrt{\sec 2\theta}$.

Ans. $f''(\theta) = 3[f(\theta)]5 - f(\theta)$.

16. $p = (q^2 + a^2)\arctan\frac{q}{a}$.

Ans. $\frac{d^3p}{dq^3} = \frac{4a^3}{(a^2+q^2)^2}$.

17. $y = a^x$.

Ans. $\frac{d^ny}{dx^n} = (\log a)^n a^x$.

18. $y = \log(1+x)$.

Ans. $\frac{d^ny}{dx^n} = (-1)^{n-1}\frac{(n-1)!}{(1+x)^n}$.

19. $y = \cos ax$.

Ans. $\frac{d^ny}{dx^n} = a^n\cos\left(ax + \frac{n\pi}{2}\right)$.

20. $y = x^{n-1}\log x$.

Ans. $\frac{d^ny}{dx^n} = \frac{(n-1)!}{x}$.

21. $y = \frac{1-x}{1+x}$.

Ans. $\frac{d^ny}{dx^n} = 2(-1)^n\frac{n!}{(1+x)^{n+1}}$.

Hint: Reduce fraction to form $-1+\frac{2}{1+x}$ before differentiating.

22. If $y = e^x \sin\ x$, prove that $\frac{d^2y}{dx^2} - 2\frac{dy}{dx} + 2y = 0$.

23. If $y = a\cos(\log\ x) + b\sin(\log\ x)$, prove that $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + y = 0$.

Use Leibnitz's Formula in the next four examples:

1. $y = x^2a^x$.

   Ans. $\frac{d^ny}{dx^n} = a^x(\log a)^{n-2}[(x\log a + n)^2 - n]$.

2. $y = xe^x$.

   Ans. $\frac{d^ny}{dx^n} = (x+n)e^x$.

3. $f(x) = e^x \sin\ x$.

   Ans. $f^{(n)}(x) = (\sqrt{2})^n e^x \sin\left(x + \frac{n\pi}{4}\right)$.

4. $f(\theta) = \cos\ a\theta \cos\ b\theta$.

   Ans. $f^n(\theta) = \frac{(a+b)^n}{2}\cos\left[(a+b)\theta + \frac{n\pi}{2}\right] + \frac{(a-b)^n}{2}\cos\left[(a-b)\theta + \frac{n\pi}{2}\right]$.

5. Show that the formulas for acceleration, , may be written $a = \frac{d^2s}{dt^2}$, $a_x = \frac{d^2x}{dt^2}$, $a_y = \frac{d^2y}{dt^2}$.

6. $y^2 = 4ax$.

   Ans. $\frac{d^2y}{dx^2} = -\frac{4a^2}{y^3}$.

7. $b^2x^2 + a^2y^2 = a^2b^2$.

   Ans. $\frac{d^2y}{dx^2} = -\frac{b^4}{a^2y^3}$; $\frac{d^3y}{dx^2} = -$ [illegible].

8. $x^2 + y^2 = r^2$. $\frac{d^2y}{dx^2} = -\frac{r^2}{y^3}$.

9. $y^2 + y = x^2$.

   Ans. $\frac{d^3y}{dx^3} = -\frac{24x}{(1+2y)^5}$.

[Type text]

10. $ax^2 + 2hxy + by^2 = 1$.

Ans. $\frac{d^2y}{dx^2} = \frac{h^2-ab}{(hx+by)^3}$.

11. $y^2 - 2xy = a^2$.

Ans. $\frac{d^2y}{dx^2} = \frac{a^2}{(y-x)^3}; \frac{d^3y}{dx^3} = -\frac{3a^2x}{(y-x)^5}$.

12. $\sec\phi \cos\theta = c$.

Ans. $\frac{d^2\theta}{d\phi^2} = \frac{\tan^2\theta - \tan^2\phi}{\tan^3\theta}$.

13. $\theta = \tan(\phi + \theta)$.

Ans. $\frac{d^3\theta}{d\phi^3} = -\frac{2(5+8\theta^2+3\theta^4)}{\theta^8}$.

14. Find the second derivative in the following:

(a) $\log(u+v) = u - v$.
(b) $e^u + u = e^v + v$.
(c) $s = 1 + te^s$.
(d) $e^s + st - e = 0$.
(e) $y^3 + x^3 - 3axy = 0$.
(f) $y^2 - 2mxy + x^2 - a = 0$.
(g) $y = \sin(x+y)$.
(h) $e^{x+y} = xy$.

UNIT-IV
Integration

What is integration?

The dictionary definition of *integration* is combining parts so that they work together or form a whole. Mathematically, integration stands for finding the area under a curve from one point to another. It is represented by

$$\int_a^b f(x)dx$$

where the symbol $\int$ is an integral sign, and a and b are the lower and upper limits of integration, respectively, the function f is the integrand of the integral, and x is the variable of integration. Figure 1 represents a graphical demonstration of the concept.

Riemann Sum

Let f be defined on the closed interval $[a,b]$, and let Δ be an arbitrary partition of $[a,b]$ such as: $a = x_0 < x_1 < x_2 < ..... < x_{n-1} < x_n = b$, where Δx_i is the length of the i^{th} subinterval (Figure 2).

If c_i is any point in the i^{th} subinterval, then the sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i,\; x_{i-1} \le c_i \le x_i$$

is called a Riemann sum of the function f for the partition Δ on the interval $[a,b]$. For a given partition Δ, the length of the longest subinterval is called the norm of the partition. It is denoted by $\|\Delta\|$ (the norm of Δ). The following limit is used to define the definite integral.

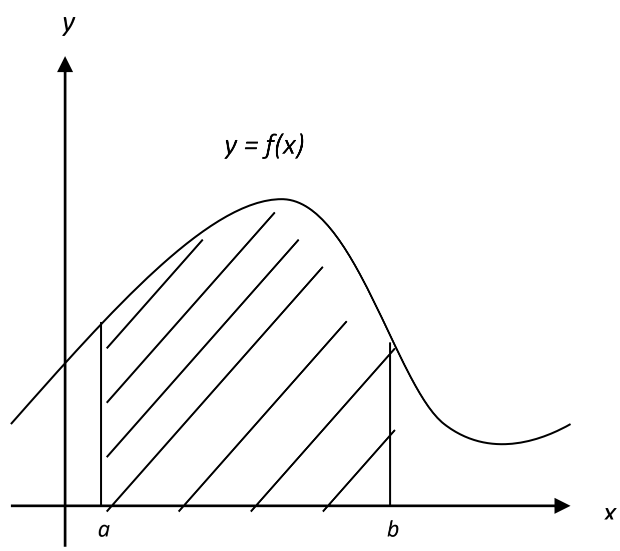


Figure The definite integral as the area of a region under the curve, $Area = \int_a^b f(x)dx$.

If c_i is any point in the i^{th} subinterval, then the sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i,\; x_{i-1} \le c_i \le x_i$$

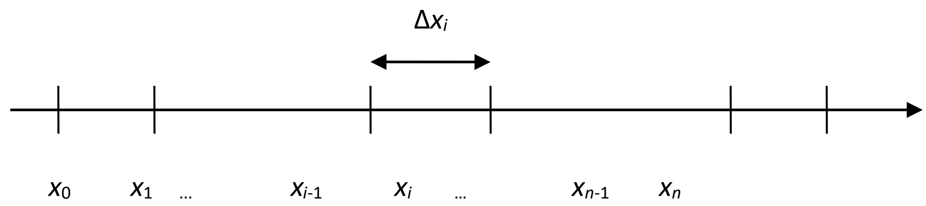


Figure Division of interval into n segments.

is called a Riemann sum of the function f for the partition Δ on the interval $[a,b]$. For a given partition Δ, the length of the longest subinterval is called the norm of the partition. It is denoted by $\|\Delta\|$ (the norm of Δ). The following limit is used to define the definite integral.

$$\lim_{\|\Delta\|\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i = I$$

This limit exists if and only if for any positive number ε, there exists a positive number δ such that for every partition Δ of $[a,b]$ with $\|\Delta\| < \delta$, it follows that

$$\left| I - \sum_{i=1}^{n} f(c_i)\Delta x_i \right| < \varepsilon$$

for any choice of c_i in the i^{th} subinterval of Δ.

If the limit of a Riemann sum of f exists, then the function f is said to be integrable over $[a,b]$ and the Riemann sum of f on $[a,b]$ approaches the number I.

$$\lim_{\|\Delta\|\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i = I$$

where

$$I = \int_a^b f(x)dx$$

Example

Find the area of the region between the parabola $y = x^2$ and the x-axis on the interval $[0,4.5]$. Use Riemann's sum with four partitions.

Solution

We evaluate the integral for the area as a limit of Riemann sums. We sketch the region (Figure 3), and partition $[0,4.5]$ into four subintervals of length

$$\Delta x = \frac{4.5-0}{4} = 1.125 .$$

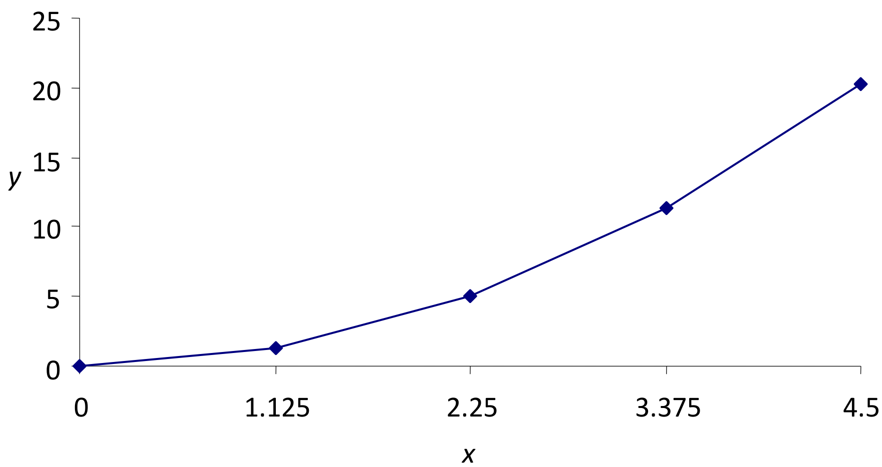


Figure Graph of the function $y = x^2$.

The points of partition are

$$x_0 = 0,\ x_1 = 1.125,\ x_2 = 2.25,\ x_3 = 3.375,\ x_4 = 4.5$$

Let's choose c_i's to be right hand endpoint of its subinterval. Thus,

$$c_1 = x_1 = 1.125,\ c_2 = x_2 = 2.25,\ c_3 = x_3 = 3.375,\ c_4 = x_4 = 4.5$$

The rectangles defined by these choices have the following areas:

$$f(c_1)\Delta x = f(1.125)\times(1.125) = (1.125)^2(1.125) = 1.4238$$

$$f(c_2)\Delta x = f(2.25)\times(1.125) = (2.25)^2(1.125) = 5.6953$$

$$f(c_3)\Delta x = f(3.375)\times(1.125) = (3.375)^2(1.125) = 12.814$$

$$f(c_4)\Delta x = f(4.5)\times(1.125) = (4.5)^2(1.125) = 22.781$$

The sum of the areas then is

$$\int_0^{4.5} x^2 dx \approx \sum_{i=1}^{4} f(c_i)\Delta x,$$

$$=1.4238+5.6953+12.814+22.781$$

$$= 42.715$$

How does this compare with the exact value of the integral $\int_0^{4.5} x^2 dx$?

Example

Find the exact area of the region between the parabola $y = x^2$ and the $x-$axis on the interval $[0, b]$. Use Riemann's sum.

Solution

Note that in Example 1 for $y = x^2$ that

$$f(c_i)\Delta x = i^2(\Delta x)^3$$

Thus, the sum of these areas, if the interval is divided into n equal segments is

$$S_n = \sum_{i=1}^{n} f(c_i)\Delta x$$

$$= \sum_{i=1}^{n} i^2 (\Delta x)^3$$

$$= (\Delta x)^3 \sum_{i=1}^{n} i^2$$

Since

$$\Delta x = \frac{b}{n}, \text{ and}$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

then

$$S_n = \frac{b^3}{n^3}\frac{n(n+1)(2n+1)}{6}$$

$$= \frac{b^3}{6}\frac{2n^2 + n + 2n + 1}{n^2}$$

$$= \frac{b^3}{6}\left(2 + \frac{3}{n} + \frac{1}{n^2}\right)$$

The definition of a definite integral can now be used

$$\int_a^b f(x)dx = \lim_{\|\Delta x\|\to 0} \sum_{i=1}^{n} f(c_i)\Delta x$$

To find the area under the parabola from $x=0$ to $x=b$, we have

$$\int_0^b x^2 dx = \lim_{|\Delta|\to 0} \sum_{i=1}^{n} f(c_i)\Delta x$$

$$= \lim_{n\to\infty} S_n$$

$$= \lim_{n\to\infty} \frac{b^3}{6}\left(2+\frac{3}{n}+\frac{1}{n^2}\right)$$

$$= \frac{b^3}{6}(2+0+0)$$

$$= \frac{b^3}{3}$$

For the value of $b=4.5$ as given in Example 1,

$$\int_0^{4.5} x^2 dx = \frac{4.5^3}{3}$$

$$= 30.375$$

The Mean Value Theorem for Integrals

The area of a region under a curve is usually greater than the area of an inscribed rectangle and less than the area of a circumscribed rectangle. The mean value theorem for integrals states that somewhere between these two rectangles, there exists a rectangle whose area is exactly equal to the area of the region under the curve, as shown in Figure 4. Another variation states that if a function f is continuous between a and b, then there is at least one point in $[a,b]$ where the function equals the average value of the function f over $[a,b]$.

Theorem: If the function f is continuous on the closed interval $[a,b]$, then there exists a number c in $[a,b]$ such that:

$$f(c) = \frac{1}{b-a}\int_a^b f(x)dx$$

Example

Graph the function $f(x) = (x-1)^2$, and find its average value over the interval $[0,3]$. At what point in the given interval does the function assume its average value?

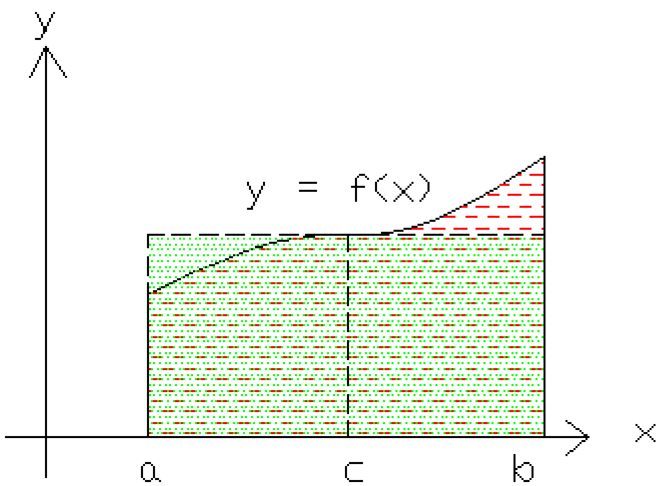


Figure Mean value rectangle.

Solution

$$Average(f) = \frac{1}{b-a}\int_a^b f(x)dx$$

$$= \frac{1}{3-0}\int_0^3 (x-1)^2 dx$$

$$= \frac{1}{3}\int_0^3 (x^2 - 2x + 1)dx$$

$$= \frac{1}{3}\left[\left(\frac{1}{3}\times 27 - 9 + 3\right) - 0\right]$$

$$= 1$$

The average value of the function f over the interval $[0,3]$ is 1. Thus, the function assumes its average value at

$$f(c) = 1$$

$$(c-1)^2 = 1$$

$$c = 0,\ 2$$

The connection between integrals and area can be exploited in two ways. When a formula for the area of the region between the x-axis and the graph of a continuous function is known, it can be used to evaluate the integral of the function. However, if the area of region is not known, the integral of the function can be used to define and calculate the area. Table 1 lists a number of standard indefinite integral forms.

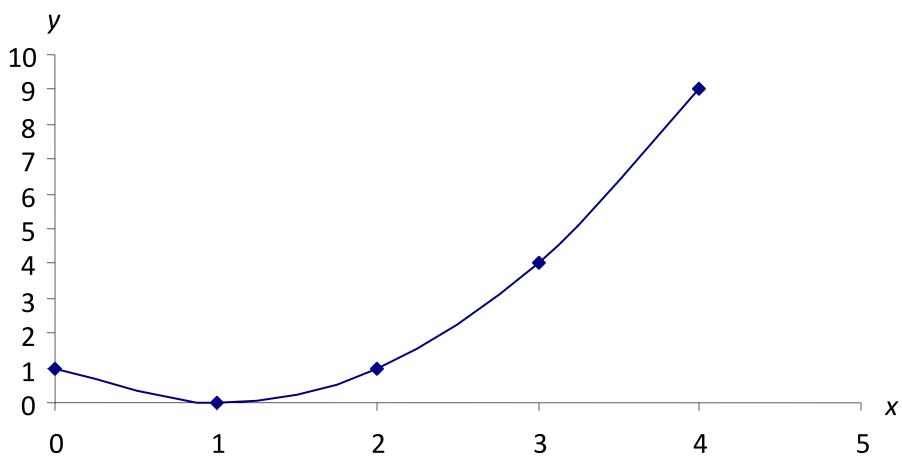


Figure The function $f(x) = (x-1)^2$.

Example 4

Find the area of the region between the circle $x^2 + y^2 = 1$ and the x-axis on the interval $[0,1]$ (the shaded region) in two different ways.

Solution

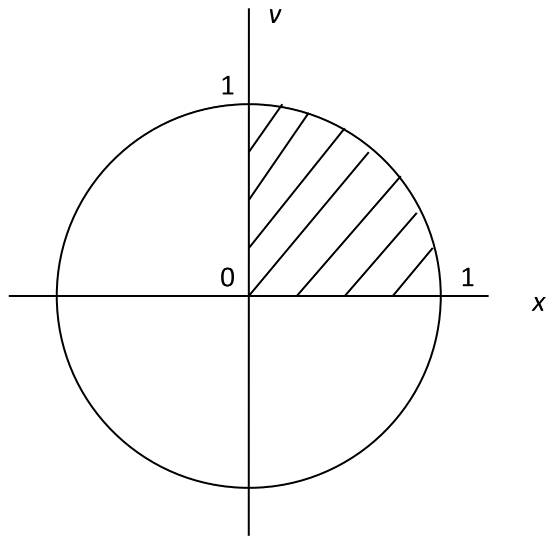


Figure Graph of the function $x^2 + y^2 = 1$.

The first and easy way to solve this problem is by recognizing that it is a quarter circle. Hence the area of the shaded area is

$$A = \frac{1}{4}\pi r^2$$

$$= \frac{1}{4}\pi(1)^2$$

$$= \frac{\pi}{4}$$

The second way is to use the integrals and the trigonometric functions. First, let's simplify the function $x^2 + y^2 = 1$.

$$x^2 + y^2 = 1$$
$$y^2 = 1 - x^2$$
$$y = \sqrt{1 - x^2}$$

The area of the shaded region is the equal to

$$A = \int_0^1 \sqrt{1 - x^2}\,dx$$

We set $x = \sin\theta,\ dx = \cos\theta\, d\theta$

[Type text]

$$A = \int_0^1 \sqrt{1-x^2}\,dx$$

$$= \int_0^{\pi/2} \sqrt{\left(1-\sin^2\theta\right)}\cos\theta\;d\theta$$

$$= \int_0^{\pi/2} \sqrt{\left(\cos^2\theta\right)}\cos\theta\;d\theta$$

$$= \int_0^{\pi/2} \cos^2\theta\;d\theta$$

By using the following formula

$$\cos^2\theta = \frac{1+\cos 2\theta}{2},$$

we have

$$A = \int_0^{\pi/2} \frac{1+\cos 2\theta}{2}\,d\theta$$

$$= \int_0^{\pi/2} \left(\frac{1}{2}+\frac{\cos 2\theta}{2}\right)d\theta$$

$$= \left[\frac{1}{2}\theta + \frac{\sin 2\theta}{4}\right]_0^{\pi/2}$$

$$= \left(\frac{\pi}{4}+0\right)-\left(0+0\right)$$

$$= \frac{\pi}{4}$$

The following are some more examples of exact integration. You can use the brief table of integrals given in Table 1.

Table A brief table of integrals

$\int dx = x + C$	$\int \sin xdx = -\cos x + C$
$\int a\, f(x)\, dx = a\int f(x)\, dx + C$	$\int \cos xdx = \sin x + C$
$\int [u(x) \pm v(x)]dx = \int u(x)dx \pm \int v(x)dx + C$	$\int \tan xdx = -\ln\lvert\cos x\rvert + C = \ln\lvert\sec x\rvert + C$
$\int x^n dx = \frac{x^{n+1}}{n+1} + C$	$\int \sec(ax)dx = \frac{1}{a}\ln\lvert\sec(ax) + \tan(ax)\rvert + C$
$\int u\, dv = u\, v - \int v\, du + C$	$\int \cot xdx = -\ln\lvert\csc x\rvert + C = \ln\lvert\sin x\rvert + C$
$\int \frac{dx}{ax+b} = \frac{1}{a}\ln\lvert ax+b\rvert + C$	$\int \sec^2 axdx = \frac{1}{a}\tan(ax) + C$
$\int a^x dx = \frac{a^x}{\ln a} + C$	$\int \sec(x)\tan(x)dx = \sec(x) + C$
$\int e^{ax} dx = \frac{e^{ax}}{a} + C$	$\int \csc(x)\cot(x)dx = -\csc(x) + C$

Example 5

Evaluate the following integral

$$\int_0^1 2xe^{-x^2}\,dx$$

Solution

Let $u = -x^2,\ du = -2xdx$

At $x = 0,\ u = -(0)^2 = 0$

At $x = 1,\ u = -(1)^2 = -1$

$$\int_0^1 2xe^{-x^2}\,dx = \int_0^1 (-e^{-x^2})(-2xdx)$$

$$= \int_0^{-1} (-e^u)(du)$$

$$= \left[-e^u\right]_0^{-1}$$

$$= -e^{-1} - (-e^0)$$

$$=0.6321$$

Example 6

Evaluate

$$\int_0^{\pi/4} \frac{1+\sin x}{\cos^2 x}\,dx$$

Solution

$$\int_0^{\pi/4} \frac{1+\sin x}{\cos^2 x}\,dx = \int_0^{\pi/4} \left(\frac{1}{\cos^2 x} + \frac{\sin x}{\cos^2 x}\right)dx$$

$$= \int_0^{\pi/4} \left(\sec^2 x + \sec x \times \tan x\right)dx$$

$$= \int_0^{\pi/4} \left(\sec^2 x\right)dx + \int_0^{\pi/4} (\sec x)(\tan x)\,dx$$

$$= [\tan x]_0^{\pi/4} + [\sec x]_0^{\pi/4}$$

$$= (1-0) + \left(\sqrt{2} - 1\right)$$

$$= \sqrt{2}$$

Example 7

Evaluate $\int x \sec^2 x\, dx$

Solution

We use the formula

$$\int u dv = uv - \int v du$$

Let $u = x,\ du = dx$, and $dv = \sec^2 x\, dx, v = \tan x$

So the new integral is

$$\int x \sec^2 x dx = x \tan x - \int \tan x dx$$

$$= x \tan x + \ln|\cos x| + C$$

Example 8

Evaluate

$$\int_1^2 x \ln x dx$$

Solution

Let $u = \ln x,\ du = \frac{1}{x} dx$ and $dv = x dx, v = \frac{x^2}{2}$

Using the formula $\int u dv = uv - \int v du$, the new integral is

$$\int_1^2 (x)(\ln x) dx = \left[\ln x \times \frac{x^2}{2}\right]_1^2 - \int_1^2 \left(\frac{x^2}{2}\right)\left(\frac{1}{x} dx\right)$$

$$= \left[\ln x \times \frac{x^2}{2}\right]_1^2 - \int_1^2 \frac{x}{2} dx$$

$$= \left[\ln x \times \frac{x^2}{2}\right]_1^2 - \left[\frac{x^2}{4}\right]_1^2$$

$$= \left[\left(\ln 2 \times \frac{2^2}{2}\right) - \left(\ln 1 \times \frac{1^2}{2}\right)\right] - \left[\left(\frac{2^2}{4}\right) - \left(\frac{1^2}{4}\right)\right]$$

$$= \left[(2\ln 2) - \left(\frac{1}{2}\ln 1\right)\right] - \left[\left(\frac{4}{4}\right) - \left(\frac{1}{4}\right)\right]$$

$$= \left[(2\ln 2) - \left(\frac{1}{2} \times 0\right)\right] - \left[1 - \frac{1}{4}\right]$$

$$= 0.6362$$

Example 9

Evaluate

$$\int_0^1 \frac{5x}{(4+x^2)^2}\,dx$$

Solution

We use the formula $\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du$, by substituting $u = g(x)$, $du = g'(x)dx$ then integrating from $g(a)$ to $g(b)$.

Let

$$u = g(x) = 4 + x^2,$$

so

$$g(0) = 4,\ g(1) = 5, \text{ and}$$

$$du = (2x)dx$$

The new integral is

$$\int_0^1 \frac{5x}{(4+x^2)^2}dx = \int_0^1 \frac{1}{(4+x^2)^2} \times \frac{5}{2} \times (2x)dx$$

$$= \frac{5}{2}\int_4^5 \frac{1}{u^2}\,du$$

$$= \frac{5}{2}\left[-\frac{1}{u}\right]_4^5$$

$$= \frac{5}{2}\left[(-\frac{1}{5})-(-\frac{1}{4})\right]$$

$$= \frac{5}{2}\times\frac{1}{20}$$

$$= 0.125$$

Example 10

Evaluate

$$\int_0^4 |2x-1|\,dx$$

Solution

First, let's analyze the expression $|2x-1|$.

$$|2x-1| = -(2x-1)\,,\ \ x < \frac{1}{2}$$

$$= (2x-1)\,,\ \ x \geq \frac{1}{2}$$

$$\int_0^4 |2x-1|\,dx = \int_0^{1/2} -(2x-1)dx + \int_{1/2}^4 (2x-1)dx$$

$$= \left[-x^2 + x\right]_0^{1/2} + \left[x^2 - x\right]_{1/2}^4$$

$$= \left[\left(-\frac{1}{4}+\frac{1}{2}\right)-0\right] + \left[(16-4)-\left(\frac{1}{4}-\frac{1}{2}\right)\right]$$

$$= 12.5$$

Example 11

Evaluate

$$\int_{-\infty}^{-2} \frac{2}{x^2-1}\,dx$$

Solution

$$\int_{-\infty}^{-2} \frac{2}{x^2-1}\,dx = \int_{-\infty}^{-2} \frac{2}{(x-1)\times(x+1)}\,dx$$

$$= \int_{-\infty}^{-2} \frac{(x+1)-(x-1)}{(x-1)\times(x+1)}\,dx$$

$$= \int_{-\infty}^{-2} \frac{x+1}{(x-1)\times(x+1)} - \frac{x-1}{(x-1)\times(x+1)}\,dx$$

$$= \int_{-\infty}^{-2} \frac{1}{x-1}\,dx - \int_{-\infty}^{-2} \frac{1}{x+1}\,dx$$

$$= \lim_{b\to-\infty} \left[\ln|x-1|\right]_b^{-2} - \lim_{b\to-\infty} \left[\ln|x+1|\right]_b^{-2}$$

$$= \lim_{b\to-\infty} \left[\ln\left|\frac{x-1}{x+1}\right|\right]_{-2}^{b}$$

$$= \lim_{b\to-\infty} \left[\ln\left|\frac{-3}{-1}\right| - \ln\left|\frac{b-1}{b+1}\right|\right]$$

$$= \ln(3) - \ln\left(\lim_{b\to-\infty}\left|\frac{b-1}{b+1}\right|\right)$$

$$= \ln(3) - \ln(1)$$

$$= \ln(3)$$

$$= 1.0986$$

Example 12

Graph the function $y = \frac{1}{3}(x^2+2)^{3/2}$, and find the length of the curve from $x = 0$ to $x = 3$.

Solution

We use the equation

$$L = \int_a^b \sqrt{1 + (\frac{dy}{dx})^2}\, dx$$

We have:

$$y = \frac{1}{3}(x^2 + 2)^{3/2}$$

So,

$$\frac{dy}{dx} = \left(\frac{1}{3}\right) \times \left(\frac{3}{2}\right) \times \left(x^2 + 2\right)^{3/2-1} \times (2x)$$

$$= x\sqrt{x^2 + 2}$$

$$L = \int_0^3 \sqrt{1 + \left(x\sqrt{x^2 + 2}\right)^2}\, dx$$

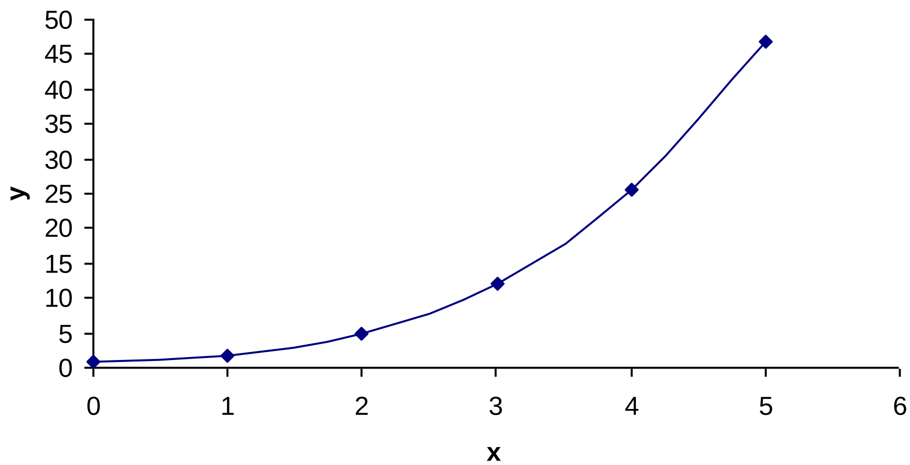


Figure 7 Graph of the function $y = \frac{1}{3}(x^2 + 2)^{3/2}$

$$= \int_0^3 \sqrt{1+x^2(x^2+2)}dx$$

$$= \int_0^3 \sqrt{1+x^4+2x^2}dx$$

$$= \int_0^3 \sqrt{(x^2+1)^2}dx$$

$$= \int_0^3 (x^2+1)dx$$

$$= \left[\frac{x^3}{3}+x\right]_0^3$$

$$= 12$$

Example 13

Find the area of the shaded region given in Figure 8.

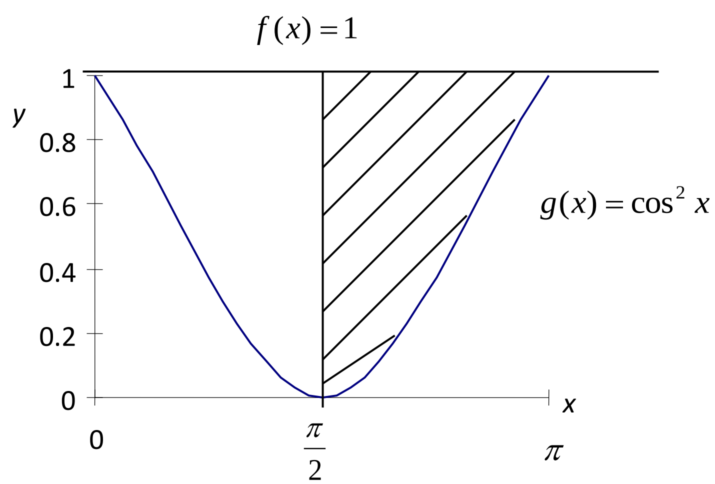


Figure 8 Graph of the function $\cos^2 x$.

Solution

For the sketch given,

$$a = \frac{\pi}{2},\ b = \pi \text{, and}$$

$$f(x) - g(x) = 1 - \cos^2 x = \sin^2 x$$

$$A = \int_{\pi/2}^{\pi} \sin^2(x)\,dx$$

$$= \int_{\pi/2}^{\pi} \frac{1 - \cos 2x}{2}\,dx$$

$$= \int_{\pi/2}^{\pi} \left[\frac{1}{2} - \frac{\cos 2x}{2}\right] dx$$

$$= \left[\frac{x}{2} - \frac{\sin 2x}{4}\right]_{\pi/2}^{\pi}$$

$$= \left[\left(\frac{\pi}{2} - \frac{\sin(2\pi)}{4}\right) - \left(\frac{\pi}{4} - \frac{\sin 2\left(\frac{\pi}{2}\right)}{4}\right)\right]$$

$$= \left[\left(\frac{\pi}{2} - 0\right) - \left(\frac{\pi}{4} - 0\right)\right]$$

$$= \frac{\pi}{4}$$

Example 14

Find the volume of the solid generated by revolving the shaded region in Figure 9 about the *y*-axis.

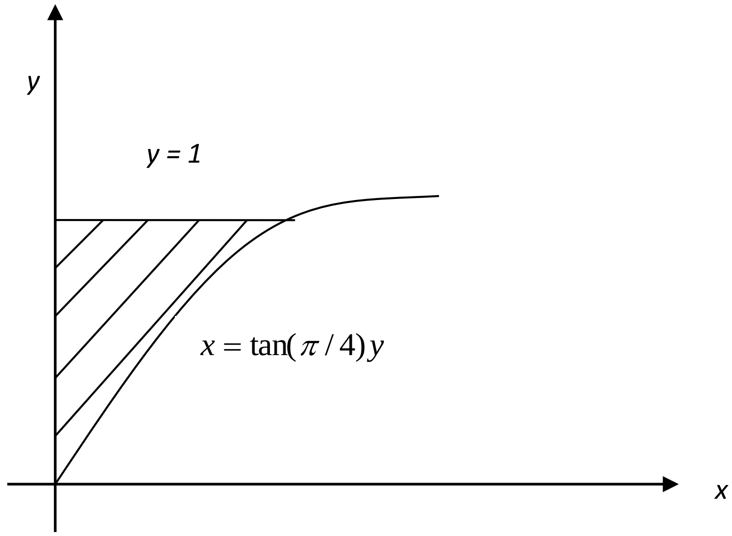


Figure 9 Volume generated by revolving shaded region.

Solution

We use the formula $V = \int_a^b \pi (radius)^2 dy$

Let

$$u = \frac{\pi}{4} y,\ du = \frac{\pi}{4} dy\ .$$

Therefore, at $y = 0,\ u = 0$

$$y = 1,\ u = \frac{\pi}{4}$$

$$V = \int_0^1 \pi [R(y)]^2 dy$$

$$= \pi \int_0^1 \left[\tan\left(\frac{\pi}{4} y \right) \right]^2 dy$$

$$= \pi \times \frac{4}{\pi} \int_0^1 \left[\tan\left(\frac{\pi}{4} y \right) \right]^2 \frac{\pi}{4} dy$$

$$= 4 \int_0^{\pi/4} (\tan u)^2 du \quad \text{(Choosing } u = \frac{\pi}{4} y \text{)}$$

$$= 4\int_0^{\pi/4}\left(-1+\sec^2 u\right)du$$

$$= 4\left[-u+\tan u\right]_0^{\pi/4}$$

$$= 4\left[\left(-\frac{\pi}{4}+\tan\frac{\pi}{4}\right)-\left(0+\tan 0\right)\right]$$

$$= 4\left[\left(-\frac{\pi}{4}+1\right)-\left(0+0\right)\right]$$

$$= 0.8584$$

Partial Fractions:-

Partial Fractions provides a way to integrate all rational functions.

Rational functions= $\dfrac{P(x)}{Q(x)}$ when P and Q are polynomials

This is the technique to find $\int \dfrac{P(x)}{Q(x)}dx$

Rule 1: The degree of the numerator must be less than the degree of the denominator. If this is not the case we first must divide the numerator into the denominator.

Step 1: If Q has a quadratic factor $ax^2 + bx + c$ which corresponds to a complex root of order k, then the partial fraction expansion of $\dfrac{P}{Q}$ contains a term of the form

$$\frac{B_1x+C_1}{(ax^2+bx+c)}+\frac{B_2x+C_2}{(ax^2+bx+c)^2}+\ldots+\frac{B_k+C_k}{(ax^2+bx+c)^k}$$

Where $B_1, B_2, \ldots B_k$ and $C_1, C_2, \ldots, C_k$ are unknown constants.

Step 2: Set the sum of the terms of equal to the partial fraction expansion

[Type text]

Example: $\frac{1}{(x-2)(x-5)} = \frac{A}{x-2} + \frac{B}{x-5}$

Step 3: When then multiply both sides by Q to get some expression that is equal to P

Example: 1= A(x-5) + B(x-2)

1= (A+B)x-5A-2B

Step 4: Use the theory that 2 polynomials are equal if and only if the corresponding coefficients are equal

Example: 5A-2B=1 and A+B=0

Step 5: Solve for A, B, and C

Example: A= -1/3 B= 1/3

Step 6: Express integral of $\frac{P}{Q}$ as the sum of the integrals of the terms of partial fraction expansion.

Example: $\int \frac{1}{(x-2)(x-5)} dx = \int \frac{\frac{-1}{3}}{(x-2)} dx + \int \frac{\frac{1}{3}}{(x-5)} dx$

$= \frac{-1}{3} \ln|x-2| + \frac{1}{3} \ln|x-5| + C$

Example 2:

Find $\int \frac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1} dx$

$$\frac{x^4-2x^2+4x+1}{x^3-x^2-x+1}=x+1+\frac{4x}{x^3-x^2-x+1}$$ Note: long division

$$\frac{4x}{(x-1)^2(x+1)}$$ Note: Factor Q(x)= $x^3 - x^2 - x + 1$

$$\frac{A}{(x-1)}+\frac{B}{(x-1)^2}+\frac{C}{x+1}$$ Note: Partial fraction decomposition since $(x-1)^2$'s factor is linear. There is a constant on top for the and power and first power

4x= A(x-1)(x+1) + B(x+1) + C(x-1) ^{2}

Note: multiply by Least common denominator

(x-1) ^{2} (x+1)

= (A+C)x ^{2} +(B-2C)x+(-A+B+C)

A+C = 0

B-2C= 4

-A+B+C= 0 Note: Equate equations

A=1 B=2 C=-1 Note: Solve for coefficients

$$\int (x+1)dx+\int \frac{1}{x-1}dx+\int \frac{2}{(x-1)^2}dx-\int \frac{1}{x+1}dx$$

$$=\frac{x^2}{2}+x+\ln|x-1|-\frac{2}{x-1}-\ln|x+1|+C$$

$$=\frac{x^2}{2}+x-\frac{2}{x-1}+\ln\left|\frac{x-1}{x+1}\right|+C$$

Example 3:

Find $\int \frac{2x^2 - x + 4}{x^3 + 4x} dx$

$\frac{2x^2 - x + 4}{x^3 + 4x} = \frac{A}{x} + \frac{Bx + C}{x^2 + 4}$ Note: x^2+4 is quadratic

$2x^2-x + 4 = A(x^2+4)+ (Bx+C)x$ Note: multiplying $x(x^2+4)$

$= (A+B)\, x^2+Cx+4A$

A+B=2 C=-1 4A=4 Note: Equating coefficients

A=1 B=1 C=-1

$$\int \frac{2x^2 - x + 4}{x^3 + 4x} dx = \int \frac{1}{x} dx + \int \frac{x-1}{x^2+4} dx = \int \frac{1}{x} dx + \int \frac{x}{x^2+4} dx - \int \frac{1}{x^2+4} dx$$

$$= \ln|x| + \frac{1}{2}\ln|x^2+4| - \frac{1}{2}\tan^{-1}\left(\frac{x}{2}\right) + C$$

INTEGRATION BY PART

This is a method to evaluate integrals that cannot be evaluated by eye or by u-substitution. It is usually applied to expressions with varied functions within each other, or multiplied by each other. A good rule is: if the expression has a chain of functions (f(g(x)) or if the expression has a product of functions x(f(x)), integration by parts will be necessary. Here are some examples of problems that would be solved with integration by parts:

$x \cdot \sec^2(x)$ $x^3 \ln(x)$ $\theta^2 \sin\theta$

Let's start with:

$x^3 \ln(x)$

In integration by parts, you separate the expression into two parts: u, and $\partial(v)$.

The u should be easy to differentiate, and the $d(v)$ should be easy to integrate.

Once you have chosen a u and a $d(v)$, set up a chart like this:

u = ln(x) $\qquad d(v) = \frac{\ln(x)\cdot x^4}{4} - \frac{x^4}{16}$

$d(u) = x^{-1} \qquad$ v $= \frac{x^4}{4}$

Now, the formula to solve this is:

$$uv - \int\left(d(u)v\right)$$

so here, the equation to solve is:

$$\frac{\ln(x)\cdot x^4}{4} - \int\left(x^{-1}\cdot\frac{x^4}{4}\right)d(x)$$

which simplifies to:

$$\frac{\ln(x)\cdot x^4}{4} - \int\left(\frac{x^3}{4}\right)d(x)$$

and solve the integral to get:

$$\frac{\ln(x)\cdot x^4}{4} - \frac{x^4}{16}$$

Unfortunately, it is not always so simple. Sometimes, you must use u-substitution, or even integration by parts again within the solution. Take, for example:

$$\theta^2 \sin\theta$$

To solve this, you would set up a chart again.

$$\theta^2 \sin\theta$$

u = θ^2 $\quad$ $d(v)$ = $\sin\theta$

$d(u)$ = 2θ $\quad$ v = $-\cos\theta$

With this chart, you can set up the solution using the $uv-\int(d(u)v)$ formula:

$$=-\theta^2\cos\theta+\int(2\theta\cos\theta)d(\theta)$$

But the second part of this, $\int(2\theta\cos\theta)d(\theta)$ cannot be solved by eye. You must set up a second chart:

$$\int(2\theta\cos\theta)d(\theta)$$

u = 2θ $\quad$ $d(v)$ = $\cos\theta$

$d(u)$ = 2 $\quad$ v = $\sin\theta$

This gives us:

$$=2\theta\sin\theta+\int(2\sin\theta)d(\theta)$$

Which can be simplified to:

$$=2\theta\sin\theta+2\cos\theta$$

Now, you can substitute it into the original solution, in the place of $\int(2\theta\cos\theta)d(\theta)$, giving you:

$$=-\theta^2\cos\theta+2\theta\sin\theta+2\cos\theta$$

Trigonometric Integrals(REDUCTION FORMULA)

I. Integrating Powers of the Sine and Cosine Functions

A. Useful trigonometric identities

1. $\sin^2 x+\cos^2 x=1$

2. $\sin 2x = 2\sin x \cos x$

3. $\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$

4. $\sin^2 x = \dfrac{1-\cos 2x}{2}$

5. $\cos^2 x = \dfrac{1+\cos 2x}{2}$

6. $\sin x \cos y = \dfrac{1}{2}[\sin(x-y) + \sin(x+y)]$

7. $\sin x \sin y = \dfrac{1}{2}[\cos(x-y) - \cos(x+y)]$

8. $\cos x \cos y = \dfrac{1}{2}[\cos(x-y) + \cos(x+y)]$

B. Reduction formulas

1. $\displaystyle\int \sin^n x\, dx = -\frac{1}{n}\sin^{n-1} x \cos x + \frac{n-1}{n}\int \sin^{n-2} x\, dx$

2. $\displaystyle\int \cos^n x\, dx = \frac{1}{n}\cos^{n-1} x \sin x + \frac{n-1}{n}\int \cos^{n-2} x\, dx$

C. Examples

1. Find $\displaystyle\int \sin^2 x\, dx$.

Method 1(Integration by parts): $\displaystyle\int \sin^2 x\, dx = \int \sin x\, (\sin x\, dx)$. Let

$u = \sin x$ and $dv = \sin x\ dx \Rightarrow du = \cos x\ dx$ and $v = \int \sin x\, dx =$

1

$-\cos x$. Thus, $\int \sin^2 x\ dx = (\sin x)(-\cos x) + \int \cos^2 x\, dx = -\sin x \cos x +$

$\int (1 - \sin^2 x)\ dx = -\sin x \cos x + \int 1 dx - \int \sin^2 x\ dx = -\sin x \cos x + x -$

$\int \sin^2 x\ dx \Rightarrow 2\int \sin^2 x\ dx = -\sin x \cos x + x \Rightarrow \int \sin^2 x\ dx =$

$-\frac{1}{2}\sin x \cos x + \frac{1}{2}x + C$.

Method 2(Trig identity): $\int \sin^2 x\ dx = \frac{1}{2}\int (1 - \cos 2x)\ dx = \frac{1}{2}x - \frac{1}{4}\sin 2x + C$.

Method 3(Reduction formula): $\int \sin^2 x\ dx = -\frac{1}{2}\sin x \cos x + \frac{1}{2}\int 1 dx =$

$-\frac{1}{2}\sin x \cos x + \frac{1}{2}x + C$.

2. Find $\int \cos^3 x\ dx$.

Use the reduction formula: $\int \cos^3 x\ dx = \frac{1}{3}\cos^2 x \sin x + \frac{2}{3}\int \cos x\, dx =$

$\frac{1}{3}\cos^2 x \sin x + \frac{2}{3}\sin x + C = \frac{1}{3}\sin x(1 - \sin^2 x) + \frac{2}{3}\sin x + C =$

$\sin x - \frac{1}{3}\sin^3 x + C$.

3. Find $\int \sin^3 x \cos^2 x \, dx$.

$$\int \sin^3 x \cos^2 x \, dx = \int \sin^2 x \sin x \cos^2 x \, dx = \int (1-\cos^2 x)\cos^2 x \sin x dx =$$

$\int (\cos^2 x - \cos^4 x)(\sin x \, dx)$. Let $u = \cos x \Rightarrow du = -\sin x \, dx$. Thus,

1

$$\int (\cos^2 x - \cos^4 x)(\sin x \, dx) = -\int (u^2 - u^4) \, du = -\frac{1}{3}u^3 + \frac{1}{5}u^5 + C =$$

$$-\frac{1}{3}\cos^3 x + \frac{1}{5}\cos^5 x + C.$$

2

4. Find $\int \sin^2 x \cos^2 x \, dx$.

$$\int \sin^2 x \cos^2 x \, dx = \int \left(\frac{1-\cos 2x}{2}\right)\left(\frac{1+\cos 2x}{2}\right) dx = \frac{1}{4}\int (1-\cos^2 2x)\, dx =$$

$$\frac{1}{4}\int \sin^2 2x \, dx = \frac{1}{4}\int \left(\frac{1-\cos 4x}{2}\right) dx = \frac{1}{8}\int 1 dx - \frac{1}{8}\int \cos 4x \, dx =$$

$$\frac{1}{8}x - \frac{1}{32}\sin 4x + C.$$

5. Find $\int \sin 4x \cos 3x \, dx$.

Method 1(Integration by parts): Let $u = \sin 4x$ and $dv = \cos 3x \, dx \Rightarrow du =$

$4\cos 4x \, dx$ and $v = \frac{1}{3}\sin 3x$. Thus, $\int \sin 4x \cos 3x \, dx =$

$(\sin 4x)\left(\frac{1}{3}\sin 3x\right)-\frac{4}{3}\int \cos 4x \sin 3x\ dx = \frac{1}{3}\sin 4x \sin 3x -$

$\frac{4}{3}\int \cos 4x \sin 3x\ dx$. Find $\int \cos 4x \sin 3x\ dx$. Let $u=\cos 4x$ and $dv =$

$\sin 3x\ dx \Rightarrow du = -4\sin 4x\ dx$ and $v=-\frac{1}{3}\cos 3x$. Thus,

$\int \cos 4x \sin 3x\ dx = -\frac{1}{3}\cos 4x \cos 3x - \frac{4}{3}\int \sin 4x \cos 3x\ dx$. Returning to

the original integral, $\int \sin 4x \cos 3x\ dx = \frac{1}{3}\sin 4x \sin 3x -$

$\frac{4}{3}\left\{-\frac{1}{3}\cos 4x \cos 3x - \frac{4}{3}\int \sin 4x \cos 3x\ dx\right\} = \frac{1}{3}\sin 4x \sin 3x +$

$\frac{4}{9}\cos 4x \cos 3x + \frac{16}{9}\int \sin 4x \cos 3x\ dx \Rightarrow -\frac{7}{9}\int \sin 4x \cos 3x\ dx =$

$\frac{1}{3}\sin 4x \sin 3x + \frac{4}{9}\cos 4x \cos 3x \Rightarrow \int \sin 4x \cos 3x\ dx =$

$-\frac{3}{7}\sin 4x \sin 3x - \frac{4}{7}\cos 4x \cos 3x + C$.

Method 2(Trig identity): $\int \sin 4x \cos 3x\ dx = \frac{1}{2}\int (\sin x + \sin 7x)\ dx =$

$-\frac{1}{2}\cos x - \frac{1}{14}\cos 7x + C$.

3

II. Integrating Powers of the Tangent and Secant Functions

A. Useful trigonometric identity: $\tan^2 x + 1 = \sec^2 x$

B. Useful integrals

1. $\int \sec x \tan x \, dx = \sec x + C$

2. $\int \sec^2 x \, dx = \tan x + C$

3. $\int \tan x \, dx = \ln|\sec x| + C = -\ln|\cos x| + C$

4. $\int \sec x \, dx = \ln|\sec x + \tan x| + C$

C. Reduction formulas

1. $\int \sec^n x \, dx = \dfrac{\sec^{n-2} x \tan x}{n-1} + \dfrac{n-2}{n-1}\int \sec^{n-2} x \, dx$

2. $\int \tan^n x \, dx = \dfrac{\tan^{n-1} x}{n-1} - \int \tan^{n-2} x \, dx$

D. Examples

1. Find $\int \tan^2 x \, dx$.

$$\int \tan^2 x \, dx = \int (\sec^2 x - 1)\, dx = \int \sec^2 x \, dx - \int 1 \, dx = \tan x - x + C.$$

2. Find $\int \tan^3 x \, dx$.

$$\int \tan^3 x\,dx = \frac{\tan^2 x}{2} - \int \tan x\,dx = \frac{1}{2}\tan^2 x - \ln|\sec x| + C.$$

4

3. Find $\int \sec^3 x\,dx$.

$$\int \sec^3 x\,dx = \frac{\sec x \tan x}{2} + \frac{1}{2}\int \sec x\,dx = \frac{1}{2}\sec x \tan x + \frac{1}{2}\ln|\sec x + \tan x| + C.$$

4. Find $\int \tan x \sec^2 x\,dx$.

Let $u = \tan x \Rightarrow du = \sec^2 x dx \Rightarrow \int \tan x \sec^2 x\,dx = \int u du = \frac{1}{2}u^2 + C =$

$\frac{1}{2}\tan^2 x + C$.

5. Find $\int \tan x \sec^4 x\,dx$.

$\int \tan x \sec^4 x\,dx = \int \tan x \sec^2 x \sec^2 x\,dx = \int \tan x(1+\tan^2 x)\sec^2 x\,dx =$

$\int \tan x \sec^2 x\,dx + \int \tan^3 x \sec^2 dx$. Let $u = \tan x \Rightarrow du = \sec^2 x\,dx$. Thus,

$\int \tan x \sec^4 x\,dx = \int u du + \int u^3 du = \frac{1}{2}u^2 + \frac{1}{4}u^4 + C = \frac{1}{2}\tan^2 x + \frac{1}{4}\tan^4 x + C$.

6. Find $\int \tan x \sec^3 x\,dx$.

$$\int \tan x \sec^3 x\, dx = \int \sec^2 x (\sec x \tan x\, dx).$$ Let $u = \sec x \Rightarrow du = \sec x \tan x dx$.

Thus, $\int \tan x \sec^3 x\, dx = \int u^2 du = \frac{1}{3}u^3 + C = \frac{1}{3}\sec^3 x + C$.

7. Find $\int \tan^2 x \sec^3 x\, dx$.

$\int \tan^2 x \sec^3 x\, dx = \int (\sec^2 x - 1)\sec^3 x\, dx = \int \sec^5 x\, dx - \int \sec^3 x\, dx$. Using the reduction formula, $\int \sec^5 x\, dx = \frac{1}{4}\sec^3 \tan x + \frac{3}{4}\int \sec^3 x\, dx$. Thus,

$\int \tan^2 x \sec^3 x\, dx = \int \sec^5 x\, dx - \int \sec^3 x\, dx = \frac{1}{4}\sec^3 x \tan x + \frac{3}{4}\int \sec^3 x dx -$
$\int \sec^3 x\, dx = \frac{1}{4}\sec^3 x \tan x - \frac{1}{4}\int \sec^3 x\, dx = \frac{1}{4}\sec^3 x \tan x - \frac{1}{8}\sec x \tan x -$
$\frac{1}{8}\ln|\sec x + \tan x| + C$.

8. Find $\int \sqrt{\tan x}\ \sec^4 x\ dx$.

$\int \sqrt{\tan x}\ \sec^4 x\ dx = \int \sqrt{\tan x} \sec^2 x \sec^2 x\, dx = \int \sqrt{\tan x}\,(1 + \tan^2 x)\sec^2 x dx$.

Let $u = \tan x \Rightarrow du = \sec^2 x\, dx \Rightarrow \int \sqrt{\tan x}\ \sec^4 x\ dx = \int \sqrt{\tan x} \sec^2 x\, dx +$
$\int \sqrt{\tan x} \tan^2 x \sec^2 x\ dx = \int u^{1/2} du + \int u^{5/2} du = \frac{2}{3}u^{3/2} + \frac{2}{7}u^{7/2} + C =$
$\frac{2}{3}(\tan x)^{3/2} + \frac{2}{7}(\tan x)^{7/2} + C$.

9. Find $\int \sqrt{\sec x} \tan x \, dx$.

Let $u = \sqrt{\sec x} \Rightarrow u^2 = \sec x \Rightarrow 2udu = \sec x \tan x dx = u^2 \tan x dx \Rightarrow$

$\tan x \, dx = \frac{2udu}{u^2} = \frac{2}{u} du$. Thus, $\int \sqrt{\sec x} \tan x \, dx = \int u \left(\frac{2}{u} du \right) = 2 \int 1 du =$

$2u + C = 2\sqrt{\sec x} + C$.

Practice Sheet forTrigonometric Integrals

(1) Prove the reduction formula: $\int \sin^n x \, dx = -\frac{1}{n} \sin^{n-1} x \cos x + \frac{n-1}{n} \int \sin^{n-2} x \, dx$

(2) Prove the reduction formula: $\int \cos^n x \, dx = \frac{1}{n} \cos^{n-1} x \sin x + \frac{n-1}{n} \int \cos^{n-2} x \, dx$

(3) Prove the reduction formula: $\int \sec^n x \, dx = \frac{\sec^{n-2} x \tan x}{n-1} + \frac{n-2}{n-1} \int \sec^{n-2} x \, dx$

(4) Prove the reduction formula: $\int \tan^n x \, dx = \frac{\tan^{n-1} x}{n-1} - \int \tan^{n-2} x \, dx$

(5) $\int_0^{\pi/4} \tan^3(3x) \; dx =$

(6) $\int_0^{\pi/4} \cos^2(2x) \; dx =$

(7) $\int_0^{\pi/8} \sin(5x)\cos(3x)\ dx =$

(8) $\int \tan^3 x \sec^3 x\ dx =$

(9) $\int \sqrt{\sin x}\cos^3 x\ dx =$

7

(10) $\int \cos^3 x \sin^2 x\ dx =$

(11) $\int_0^{\pi/2} \frac{\sin^3 x}{\sqrt{\cos x}}\ dx =$

(12) $\int \sin^2 x \cos^2 x dx =$

(13) $\int \tan^5 x \sec x dx =$

Solution Key for Trigonometric Integrals

(1) $\int \sin^n x\, dx = \int \sin^{n-1} x \sin x\, dx$. Use integration by parts with $u = \sin^{n-1} x$ and

$dv = \sin x\,dx \Rightarrow du = (n-1)\sin^{n-2} x\cos x\,dx$ and $v = \int \sin x\,dx = -\cos x \Rightarrow$

$\int \sin^n x\,dx = \int \sin^{n-1} x \sin x\,dx = -\sin^{n-1} x\cos x + (n-1)\int \sin^{n-2} x\cos^2 x\,dx =$

$-\sin^{n-1} x\cos x + (n-1)\int \sin^{n-2} x\left(1-\sin^2 x\right)dx = -\sin^{n-1} x\cos x +$

$(n-1)\int \sin^{n-2} x\,dx - (n-1)\int \sin^n x\,dx \Rightarrow n\int \sin^n x\,dx = -\sin^{n-1} x\cos x +$

$(n-1)\int \sin^{n-2} x\,dx \Rightarrow \int \sin^n x\,dx = -\frac{1}{n}\sin^{n-1} x\cos x + \frac{n-1}{n}\int \sin^{n-2} x\,dx$.

(2) $\int \cos^n x\,dx = \int \cos^{n-1} x\cos x\,dx$. Use integration by parts with $u = \cos^{n-1} x$ and

$dv = \cos x\,dx \Rightarrow du = (n-1)\cos^{n-2} x(-\sin x)\,dx$ and $v = \int \cos x\,dx = \sin x \Rightarrow$

$\int \cos^n x\,dx = \int \cos^{n-1} x\cos x\,dx = \cos^{n-1} x\sin x + (n-1)\int \cos^{n-2} x\sin^2 x\,dx =$

$\cos^{n-1} x\sin x + (n-1)\int \cos^{n-2} x\left(1-\cos^2 x\right)dx = \cos^{n-1} x\sin x +$

$$(n-1)\int \cos^{n-2} x\,dx-(n-1)\int \cos^{n} x\,dx \Rightarrow n\int \cos^{n} x\,dx=\cos^{n-1} x\sin x+$$

$$(n-1)\int \cos^{n-2} x\,dx \Rightarrow \int \cos^{n} x\,dx=\frac{1}{n}\cos^{n-1} x\sin x+\frac{n-1}{n}\int \cos^{n-2} x\,dx .$$

(3) $\int \sec^{n} x\,dx=\int \sec^{n-2} x\sec^{2} x\,dx$. Use integration by parts with $u=\sec^{n-2} x$ and

$dv=\sec^{2} x\,dx \Rightarrow du=(n-2)\sec^{n-3} x(\sec x\tan x\,dx)$ and $v=\int \sec^{2} x\,dx=\tan x \Rightarrow$

$$\int \sec^{n} x\,dx=\int \sec^{n-2} x\sec^{2} x\,dx = \sec^{n-2} x\tan x-(n-2)\int \sec^{n-2} x\tan^{2} x\,dx=$$

$$\sec^{n-2} x\tan x-(n-2)\int \sec^{n-2} x\left(\sec^{2} x-1\right)dx=\sec^{n-2} x\tan x-(n-2)\int \sec^{n} x dx+$$

$$(n-2)\int \sec^{n-2} x\,dx \Rightarrow (n-1)\int \sec^{n} x\,dx=\sec^{n-2} x\tan x+(n-2)\int \sec^{n-2} x\,dx \Rightarrow$$

$$\int \sec^{n} x\,dx=\frac{\sec^{n-2} x\tan x}{n-1}+\frac{n-2}{n-1}\int \sec^{n-2} x\,dx .$$

(4) $\int \tan^{n} x\,dx=\int \tan^{n-2} x\tan^{2} x\,dx=\int \tan^{n-2} x\left(\sec^{2} x-1\right)dx=\int \tan^{n-2} x\sec^{2} x dx-$

$$\int \tan^{n-2} x\,dx=\frac{\tan^{n-1} x}{n-1}-\int \tan^{n-2} x\,dx .$$

[Type text]

(5) Let $u = 3x \Rightarrow du = 3\,dx \Rightarrow \int \tan^3(3x)\,dx = \frac{1}{3}\int \tan^3(3x)\;3dx = \frac{1}{3}\int \tan^3 u\,du$. Use

reduction formula #4 above to get $\frac{1}{3}\int \tan^3 u\,du = \frac{1}{3}\left(\frac{\tan^2 u}{2}\right) - \frac{1}{3}\int \tan u\,du =$

$\frac{1}{6}\tan^2 u - \frac{1}{3}\ln|\sec u| \Rightarrow \int_0^{\pi/4} \tan^3(3x)\;\;dx = \left\{\frac{1}{6}\tan^2(3x) - \frac{1}{3}\ln|\sec(3x)|\right\}_0^{\pi/4} =$

$\left\{\frac{1}{6}\tan^2\left(\frac{3\pi}{4}\right) - \frac{1}{3}\ln\left|\sec\left(\frac{3\pi}{4}\right)\right|\right\} - \left\{\frac{1}{6}\tan^2(0) - \frac{1}{3}\ln|\sec(0)|\right\} = \frac{1}{6}(-1)^2 - \frac{1}{3}\ln\left|-\sqrt{2}\right| -$

$\frac{1}{6}(0)^2 + \frac{1}{3}\ln 1 = \frac{1}{6} - \frac{1}{3}\ln\left(\sqrt{2}\right).$

(6) Use the trigonometric identity $\cos^2\Delta = \frac{1+\cos 2\Delta}{2}$ to get $\int \cos^2(2x)\,dx =$

$\int \frac{1+\cos(4x)}{2}\,dx = \frac{1}{2}\int 1dx + \frac{1}{2}\int \cos(4x)\,dx = \frac{1}{2}x + \frac{1}{8}\sin(4x) \Rightarrow \int_0^{\pi/4} \cos^2(2x)\;\;dx =$

$\left\{\frac{1}{2}\left(\frac{\pi}{4}\right) + \frac{1}{8}\sin\pi\right\} - \left\{\frac{1}{2}(0) + \frac{1}{8}\sin(0) = \frac{\pi}{8}\right\}.$

(7) Use the trigonometric identity $\sin x \cos y = \frac{1}{2}[\sin(x-y) + \sin(x+y)]$ to get

$\int \sin(5x)\cos(3x)\,dx = \frac{1}{2}\int \sin(2x)\,dx + \frac{1}{2}\int \sin(8x)\,dx = -\frac{1}{4}\cos(2x) - \frac{1}{16}\cos(8x) \Rightarrow$

$$\int_0^{\pi/8} \sin(5x)\cos(3x)\,dx = \left\{-\frac{1}{4}\cos\left(\frac{\pi}{4}\right) - \frac{1}{16}\cos(\pi)\right\} - \left\{-\frac{1}{4}\cos 0 - \frac{1}{16}\cos 0\right\} =$$

$$-\frac{1}{4}\left(\frac{\sqrt{2}}{2}\right) + \frac{1}{16} + \frac{1}{4} + \frac{1}{16} = \frac{3-\sqrt{2}}{8}$$

10

(8) $\int \tan^3 x \sec^3 x \; dx = \int \tan^2 x \sec^2 x (\sec x \tan x\, dx) =$

$$\int (\sec^2 x - 1)\sec^2 x (\sec x \tan x\, dx) = \int \sec^4 x (\sec x \tan x\, dx) -$$

$$\int \sec^2 x (\sec x \tan x\, dx) = \frac{1}{5}\sec^5 x - \frac{1}{3}\sec^3 x + C.$$

(9) $\int \sqrt{\sin x}\cos^3 x \; dx = \int \sqrt{\sin x}\,(\cos^2 x)(\cos x\, dx) = \int (\sin x)^{1/2}(1 - \sin^2 x)\cos x dx =$

$$\int (\sin x)^{1/2}\cos x\, dx - \int (\sin x)^{5/2}\cos x\, dx = \frac{2}{3}(\sin x)^{3/2} - \frac{2}{7}(\sin x)^{7/2} + C.$$

(10) $\int \cos^3 x \sin^2 x \; dx = \int \cos^2 x \sin^2 x (\cos x\, dx) = \int (1 - \sin^2 x)(\sin^2 x)\cos x\, dx =$

$$\int \sin^2 x (\cos x\, dx) - \int \sin^4 x (\cos x\, dx) = \frac{1}{3}\sin^3 x - \frac{1}{5}\sin^5 x + C.$$

(11) $\int \frac{\sin^3 x}{\sqrt{\cos x}}\,dx = \int (\cos x)^{-1/2} \sin^2 x(\sin x\,dx = \int (\cos x)^{-1/2}\left(1-\cos^2 x\right)\sin x\,dx =$

$\int (\cos x)^{-1/2}(\sin x\,dx) - \int (\cos x)^{3/2}(\sin x\,dx) = -2(\cos x)^{1/2} + \frac{2}{5}(\cos x)^{5/2} \Rightarrow$

$$\int_0^{\pi/2} \frac{\sin^3 x}{\sqrt{\cos x}}\,dx = \left\{-2\cos\left(\frac{\pi}{2}\right) + \frac{2}{5}\left(\cos\left(\frac{\pi}{2}\right)\right)^{5/2}\right\} - \left\{-2\cos 0 + \frac{2}{5}(\cos 0)^{5/2}\right\} = \frac{8}{5}.$$

(12) Use the trigonometric identities $\cos^2\Delta = \frac{1+\cos 2\Delta}{2}$ and $\sin^2\Delta = \frac{1-\cos 2\Delta}{2}$.

$\int \sin^2 x\cos^2 x\,dx = \int\left(\frac{1-\cos 2x}{2}\right)\left(\frac{1+\cos 2x}{2}\right)dx = \frac{1}{4}\int\left(1-\cos^2 2x\right)dx =$

11

$\frac{1}{4}\int 1\,dx - \frac{1}{4}\int \cos^2 2x\,dx = \frac{1}{4}x - \frac{1}{4}\int\left(\frac{1+\cos 4x}{2}\right)dx = \frac{1}{4}x - \frac{1}{8}\int 1\,dx -$

$\frac{1}{8}\int \cos 4x\,dx = \frac{1}{4}x - \frac{1}{8}x - \frac{1}{32}\sin 4x + C = \frac{1}{8}x - \frac{1}{32}\sin 4x + C.$

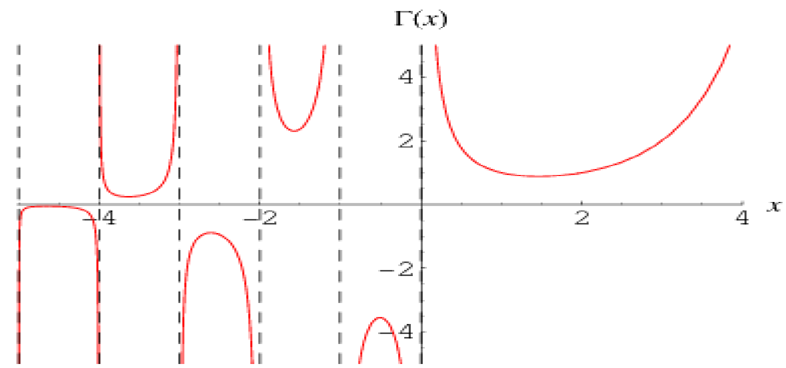
(13) $\int \tan^5 x\sec x\,dx = \int \tan^4 x\tan x\sec x\,dx = \int\left(\tan^2 x\right)^2 \tan x\sec x\,dx =$

$\int\left(\sec^2 x - 1\right)^2 \sec x\tan x\,dx = \int\left(\sec^4 x - 2\sec^2 x + 1\right)\sec x\tan x\,dx =$

$$\int \sec^4 x(\sec x \tan x\, dx) - 2\int \sec^2 x(\sec x \tan x\, dx) + \int \sec x \tan x\, dx =$$

$$\frac{1}{5}\sec^5 x - \frac{2}{3}\sec^3 x + \sec x + C\,.$$

Gamma Function



The (complete) gamma function $\Gamma(n)$ is defined to be an extension of the factorial to complex and real number arguments. It is related to the factorial by

$$\Gamma(n) = (n-1)!, \tag{1}$$

a slightly unfortunate notation due to Legendre which is now universally used instead of Gauss's simpler $\Pi(n) = n!$ (Gauss 1812; Edwards 2001, p. 8).

It is analytic everywhere except at $z = 0$, -1, -2, ..., and the residue at $z = -k$ is

$$\operatorname*{Res}_{z=-k} \Gamma(z) = \frac{(-1)^k}{k!}. \tag{2}$$

There are no points z at which $\Gamma(z) = 0$.

The gamma function is implemented in *Mathematica* as Gamma[*z*].

There are a number of notational conventions in common use for indication of a power of a gamma functions. While authors such as Watson (1939) use $\Gamma^n(z)$ (i.e., using a trigonometric function-like convention), it is also common to write $[\Gamma(z)]^n$.

The gamma function can be defined as a definite integral for $\mathbb{R}[z] > 0$ (Euler's integral form)

$$\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} \, dt \tag{3}$$

$$= 2\int_0^\infty e^{-t^2} t^{2z-1} \, dt, \tag{4}$$

or

$$\Gamma(z) \equiv \int_0^1 \left[\ln\left(\frac{1}{t}\right)\right]^{z-1} dt. \tag{5}$$

Plots of the real and imaginary parts of $\Gamma(z)$ in the complex plane are illustrated above.

Integrating equation (3) by parts for a real argument, it can be seen that

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt \tag{6}$$

$$= \left[-t^{x-1} e^{-t}\right]_0^\infty + \int_0^\infty (x-1)\, t^{x-2} e^{-t} \, dt \tag{7}$$

$$= (x-1)\int_0^\infty t^{x-2} e^{-t} \, dt \tag{8}$$

$$= (x-1)\,\Gamma(x-1). \tag{9}$$

If x is an integer $n = 1$, 2, 3, ..., then

$$\Gamma(n) = (n-1)\,\Gamma(n-1) = (n-1)(n-2)\,\Gamma(n-2) \tag{10}$$

$$= (n-1)(n-2)\cdots 1 = (n-1)!, \tag{11}$$

so the gamma function reduces to the factorial for a positive integer argument.

A beautiful relationship between $\Gamma(z)$ and the Riemann zeta function $\zeta(z)$ is given by

$$\zeta(z)\,\Gamma(z) = \int_0^\infty \frac{u^{z-1}}{e^u - 1} \, du \tag{12}$$

for $\mathbb{R}[z] > 1$ (Havil 2003, p. 60).

The gamma function can also be defined by an infinite product form (Weierstrass form)

$$\Gamma(z) \equiv \left[z e^{\gamma z} \prod_{r=1}^{\infty} \left(1 + \frac{z}{r}\right) e^{-z/r}\right]^{-1}, \tag{13}$$

where γ is the Euler-Mascheroni constant (Krantz 1999, p. 157; Havil 2003, p. 57). This can be written

$$\Gamma(z) = \frac{1}{z} \exp\left[\sum_{k=1}^{\infty} \frac{(-1)^k s_k}{k} z^k\right], \tag{14}$$

where

$$s_1 \equiv \gamma \tag{15}$$

$$s_k \equiv \zeta(k) \tag{16}$$

for $k \geq 2$, where $\zeta(z)$is the Riemann zeta function (Finch 2003). Taking the logarithm of both sides of (◇),

$$-\ln[\Gamma(z)] = \ln z + \gamma z + \sum_{n=1}^{\infty}\left[\ln\left(1+\frac{z}{n}\right) - \frac{z}{n}\right]. \tag{17}$$

Differentiating,

$$-\frac{\Gamma'(z)}{\Gamma(z)} = \frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{\frac{1}{n}}{1+\frac{z}{n}} - \frac{1}{n}\right) \tag{18}$$

$$= \frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n+z} - \frac{1}{n}\right) \tag{19}$$

$$\Gamma'(z) = -\Gamma(z)\left[\frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n+z} - \frac{1}{n}\right)\right] \tag{20}$$

$$\equiv \Gamma(z)\Psi(z) = \Gamma(z)\psi_0(z) \tag{21}$$

$$\Gamma'(1) = -\Gamma(1)\left\{1 + \gamma + \left[\left(\tfrac{1}{2} - 1\right) + \left(\tfrac{1}{3} - \tfrac{1}{2}\right) + \ldots + \left(\frac{1}{n+1} - \frac{1}{n}\right) + \ldots\right]\right\} \tag{22}$$

$$= -(1 + \gamma - 1) = -\gamma \tag{23}$$

$$\Gamma'(n) = -\Gamma(n)\left\{\frac{1}{n} + \gamma + \left[\left(\frac{1}{1+n} - 1\right) + \left(\frac{1}{2+n} - \frac{1}{2}\right) + \left(\frac{1}{3+n} - \frac{1}{3}\right) + \ldots\right]\right\} \tag{24}$$

$$= -(n-1)!\left(\frac{1}{n} + \gamma - \sum_{k=1}^{n}\frac{1}{k}\right), \tag{25}$$

where $\Psi(z)$is the digamma function and $\psi_0(z)$is the polygamma function. nth derivatives are given in terms of the polygamma functions ψ_n, ψ_{n-1}, ..., ψ_0.

The minimum value $x_0$of $\Gamma(x)$for real positive $x = x_0$is achieved when

$$\Gamma'(x_0) = \Gamma(x_0)\psi_0(x_0) = 0 \tag{26}$$

$$\psi_0(x_0) = 0. \tag{27}$$

This can be solved numerically to give $x_0 = 1.46163\ldots$(Sloane's A030169; Wrench 1968), which has continued fraction [1, 2, 6, 63, 135, 1, 1, 1, 1, 4, 1, 38, ...] (Sloane's A030170). At x_0, $\Gamma(x_0)$achieves the value 0.8856031944... (Sloane's A030171), which has continued fraction [0, 1, 7, 1, 2, 1, 6, 1, 1, ...] (Sloane's A030172).

The Euler limit form is

$$\Gamma(z) = \frac{1}{z}\prod_{n=1}^{\infty}\left[\left(1+\frac{1}{n}\right)^{z}\left(1+\frac{z}{n}\right)^{-1}\right], \tag{28}$$

so

$$\Gamma(z) = \lim_{n\to\infty}\frac{(n+1)^{z}}{z(1+z)\left(1+\frac{z}{2}\right)\left(1+\frac{z}{3}\right)\cdots\left(1+\frac{z}{n}\right)} \tag{29}$$

(30)

$$= \lim_{n\to\infty}\frac{(n+1)^{z}\,n!}{z(z+1)(z+2)(z+3)\cdots(z+n)} \tag{31}$$

$$= \lim_{n\to\infty}\frac{n!}{(z)_{n+1}}(n+1)^{z} \tag{32}$$

$$= \lim_{n\to\infty}\frac{n!}{(z)_{n+1}}n^{z} \tag{33}$$

(Krantz 1999, p. 156).

One over the gamma function $1/\Gamma(z)$ is an entire function and can be expressed as

$$\frac{1}{\Gamma(z)} = z\exp\left[\gamma z - \sum_{k=2}^{\infty}\frac{(-1)^{k}\zeta(k)z^{k}}{k}\right], \tag{34}$$

where γ is the Euler-Mascheroni constant and $\zeta(z)$ is the Riemann zeta function (Wrench 1968). An asymptotic series for $1/\Gamma(z)$ is given by

$$\frac{1}{\Gamma(z)} \sim z+\gamma z^{2}+\tfrac{1}{12}\left(6\gamma^{2}-\pi^{2}\right)z^{3}+\tfrac{1}{12}\left[2\gamma^{3}-\gamma\pi^{2}+4\zeta(3)\right]z^{4}+\ldots. \tag{35}$$

Writing

$$\frac{1}{\Gamma(z)} = \sum_{k=1}^{\infty}a_{k}z^{k}, \tag{36}$$

the a_k satisfy

$$a_{n} = n\,a_{1}\,a_{n} - a_{2}\,a_{n-1} + \sum_{k=2}^{n}(-1)^{k}\zeta(k)\,a_{n-k} \tag{37}$$

(Bourguet 1883, Davis 1933, Isaacson and Salzer 1943, Wrench 1968). Wrench (1968) numerically computed the coefficients for the series expansion about 0 of

$$\frac{1}{z\,(1+z)\,\Gamma(z)} = 1 + (\gamma - 1)\,z + \left[1 + \tfrac{1}{2}\,(\gamma - 2)\,\gamma - \tfrac{1}{12}\,\pi^2\right] z^2 + \ldots. \tag{38}$$

The Lanczos approximation gives a series expansion for $\Gamma(z+1)$ for $z > 0$ in terms of an arbitrary constant σ such that $\mathbb{R}\,[z + \sigma + 1/2] > 0$.

The gamma function satisfies the functional equations

$$\Gamma(1+z) = z\,\Gamma(z) \tag{39}$$

$$\Gamma(1-z) = -z\,\Gamma(-z). \tag{40}$$

Additional identities are

$$\Gamma(x)\,\Gamma(-x) = -\frac{\pi}{x\sin(\pi\,x)} \tag{41}$$

$$\Gamma(x)\,\Gamma(1-x) = \frac{\pi}{\sin(\pi\,x)} \tag{42}$$

$$|(i\,x)!|^2 = \frac{\pi\,x}{\sinh(\pi\,x)} \tag{43}$$

$$|(n+i\,x)!| = \sqrt{\frac{\pi\,x}{\sinh(\pi\,x)}}\prod_{s=1}^{n}\sqrt{s^2+x^2}. \tag{44}$$

Using (41), the gamma function $\Gamma(r)$ of a rational number r can be reduced to a constant times $\Gamma(\mathrm{frac}\,(r))$ or $1/\Gamma(\mathrm{frac}\,(r))$. For example,

$$\Gamma\left(\tfrac{2}{3}\right) = \frac{2\,\pi}{\sqrt{3}\,\Gamma\left(\tfrac{1}{3}\right)} \tag{45}$$

$$\Gamma\left(\tfrac{3}{4}\right) = \frac{\sqrt{2}\,\pi}{\Gamma\left(\tfrac{1}{4}\right)} \tag{46}$$

$$\Gamma\left(\tfrac{3}{5}\right) = \sqrt{2 - \frac{2}{\sqrt{5}}}\;\frac{\pi}{\Gamma\left(\tfrac{2}{5}\right)} \tag{47}$$

$$\Gamma\left(\tfrac{4}{5}\right) = \sqrt{2 + \frac{2}{\sqrt{5}}}\;\frac{\pi}{\Gamma\left(\tfrac{1}{5}\right)}. \tag{48}$$

For $\mathbb{R}\,[x] = -\frac{1}{2}$,

$$\left|\left(-\tfrac{1}{2} + i\,y\right)!\right|^2 = \frac{\pi}{\cosh(\pi\,y)}. \tag{49}$$

Gamma functions of argument $2\,z$ can be expressed using the Legendre duplication formula

$$\Gamma(2\,z) = (2\,\pi)^{-1/2}\,2^{2z-1/2}\,\Gamma(z)\,\Gamma\left(z + \tfrac{1}{2}\right). \tag{50}$$

Gamma functions of argument $3z$ can be expressed using a triplication formula

$$\Gamma(3z) = (2\pi)^{-1} 3^{3z-1/2} \Gamma(z) \Gamma\left(z+\tfrac{1}{3}\right) \Gamma\left(z+\tfrac{2}{3}\right). \tag{51}$$

The general result is the Gauss multiplication formula

$$\Gamma(z) \Gamma\left(z+\tfrac{1}{n}\right) \cdots \Gamma\left(z+\tfrac{n-1}{n}\right) = (2\pi)^{(n-1)/2} n^{1/2-nz} \Gamma(nz). \tag{52}$$

The gamma function is also related to the Riemann zeta function $\zeta(z)$ by

$$\Gamma\left(\frac{s}{2}\right) \pi^{-s/2} \zeta(s) = \Gamma\left(\frac{1-s}{2}\right) \pi^{-(1-s)/2} \zeta(1-s). \tag{53}$$

For integer $n=1$, 2, ..., the first few values of $\Gamma(n)$ are 1, 1, 2, 6, 24, 120, 720, 5040, 40320, 362880, ... (Sloane's A000142). For half-integer arguments, $\Gamma(n/2)$ has the special form

$$\Gamma\left(\tfrac{1}{2} n\right) = \frac{(n-2)!! \sqrt{\pi}}{2^{(n-1)/2}}, \tag{54}$$

where $n!!$ is a double factorial. The first few values for $n=1$, 3, 5, ... are therefore

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} \tag{55}$$

$$\Gamma\left(\tfrac{3}{2}\right) = \tfrac{1}{2} \sqrt{\pi} \tag{56}$$

$$\Gamma\left(\tfrac{5}{2}\right) = \tfrac{3}{4} \sqrt{\pi}, \tag{57}$$

$15\sqrt{\pi}/8$, $105\sqrt{\pi}/16$, ... (Sloane's A001147 and A000079; Wells 1986, p. 40). In general, for n a positive integer $n=1$, 2, ...

$$\Gamma\left(\tfrac{1}{2}+n\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi} \tag{58}$$

$$= \frac{(2n-1)!!}{2^n} \sqrt{\pi} \tag{59}$$

$$\Gamma\left(\tfrac{1}{2}-n\right) = \frac{(-1)^n 2^n}{1 \cdot 3 \cdot 5 \cdots (2n-1)} \sqrt{\pi} \tag{60}$$

$$= \frac{(-1)^n 2^n}{(2n-1)!!} \sqrt{\pi}. \tag{61}$$

Simple closed-form expressions of this type do not appear to exist for $\Gamma(1/n)$ for n a positive integer $n>2$. However, Borwein and Zucker (1992) give a variety of identities relating gamma functions to square roots and elliptic integral singular values k_n, i.e., elliptic moduli k_n such that

$$\frac{K'(k_n)}{K(k_n)} = \sqrt{n}, \tag{62}$$

where $K(k)$ is a complete elliptic integral of the first kind and $K'(k) = K(k') = K\left(\sqrt{1-k^2}\right)$ is the complementary integral. M. Trott (pers. comm.) has developed an algorithm for automatically generating hundreds of such identities.

$$\Gamma\left(\tfrac{1}{3}\right) = 2^{7/9}\, 3^{-1/12}\, \pi^{1/3}\, [K(k_3)]^{1/3} \tag{63}$$

$$\Gamma\left(\tfrac{1}{4}\right) = 2\pi^{1/4}\, [K(k_1)]^{1/2} \tag{64}$$

$$\Gamma\left(\tfrac{1}{6}\right) = 2^{-1/3}\, 3^{1/2}\, \pi^{-1/2} \left[\Gamma\left(\tfrac{1}{3}\right)\right]^2 \tag{65}$$

$$\Gamma\left(\tfrac{1}{8}\right)\Gamma\left(\tfrac{3}{8}\right) = \left(\sqrt{2}-1\right)^{1/2} 2^{13/4}\, \pi^{1/2}\, K(k_2) \tag{66}$$

$$\frac{\Gamma\left(\tfrac{1}{8}\right)}{\Gamma\left(\tfrac{3}{8}\right)} = 2\left(\sqrt{2}+1\right)^{1/2} \pi^{-1/4}\, [K(k_1)]^{1/2} \tag{67}$$

$$\Gamma\left(\tfrac{1}{12}\right) = 2^{-1/4}\, 3^{3/8} \left(\sqrt{3}+1\right)^{1/2} \pi^{-1/2}\, \Gamma\left(\tfrac{1}{4}\right)\Gamma\left(\tfrac{1}{3}\right) \tag{68}$$

$$\Gamma\left(\tfrac{5}{12}\right) = 2^{1/4}\, 3^{-1/8} \left(\sqrt{3}-1\right)^{1/2} \pi^{1/2}\, \frac{\Gamma\left(\tfrac{1}{4}\right)}{\Gamma\left(\tfrac{1}{3}\right)} \tag{69}$$

$$\frac{\Gamma\left(\tfrac{1}{24}\right)\Gamma\left(\tfrac{11}{24}\right)}{\Gamma\left(\tfrac{5}{24}\right)\Gamma\left(\tfrac{7}{24}\right)} = \sqrt{3}\,\sqrt{2+\sqrt{3}} \tag{70}$$

$$\frac{\Gamma\left(\tfrac{1}{24}\right)\Gamma\left(\tfrac{5}{24}\right)}{\Gamma\left(\tfrac{7}{24}\right)\Gamma\left(\tfrac{11}{24}\right)} = 4 \cdot 3^{1/4} \left(\sqrt{3}+\sqrt{2}\right) \pi^{-1/2}\, K(k_1) \tag{71}$$

$$\frac{\Gamma\left(\tfrac{1}{24}\right)\Gamma\left(\tfrac{7}{24}\right)}{\Gamma\left(\tfrac{5}{24}\right)\Gamma\left(\tfrac{11}{24}\right)} = 2^{25/18}\, 3^{1/3} \left(\sqrt{2}+1\right) \pi^{-1/3}\, [K(k_3)]^{2/3} \tag{72}$$

$$\Gamma\left(\tfrac{1}{24}\right)\Gamma\left(\tfrac{5}{24}\right)\Gamma\left(\tfrac{7}{24}\right)\Gamma\left(\tfrac{11}{24}\right) = 384\left(\sqrt{2}+1\right)\left(\sqrt{3}-\sqrt{2}\right)\left(2-\sqrt{3}\right)\pi\, [K(k_6)]^2 \tag{73}$$

$$\Gamma\left(\tfrac{1}{10}\right) = 2^{-7/10}\, 5^{1/4} \left(\sqrt{5}+1\right)^{1/2} \pi^{-1/2}\, \Gamma\left(\tfrac{1}{5}\right)\Gamma\left(\tfrac{2}{5}\right) \tag{74}$$

$$\Gamma\left(\tfrac{3}{10}\right) = 2^{-3/5} \left(\sqrt{5}-1\right) \pi^{1/2}\, \frac{\Gamma\left(\tfrac{1}{5}\right)}{\Gamma\left(\tfrac{2}{5}\right)} \tag{75}$$

$$\frac{\Gamma\left(\tfrac{1}{15}\right)\Gamma\left(\tfrac{4}{15}\right)\Gamma\left(\tfrac{7}{15}\right)}{\Gamma\left(\tfrac{2}{15}\right)} = 2 \cdot 3^{1/2}\, 5^{1/6} \sin\left(\tfrac{2}{15}\pi\right) \left[\Gamma\left(\tfrac{1}{3}\right)\right]^2 \tag{76}$$

$$\frac{\Gamma\left(\tfrac{1}{15}\right)\Gamma\left(\tfrac{2}{15}\right)\Gamma\left(\tfrac{7}{15}\right)}{\Gamma\left(\tfrac{4}{15}\right)} = 2^2 \cdot 3^{2/5} \sin\left(\tfrac{1}{5}\pi\right) \sin\left(\tfrac{4}{15}\pi\right) \left[\Gamma\left(\tfrac{1}{5}\right)\right]^2 \tag{77}$$

$$\frac{\Gamma\left(\frac{2}{15}\right)\Gamma\left(\frac{4}{15}\right)\Gamma\left(\frac{7}{15}\right)}{\Gamma\left(\frac{1}{15}\right)} = \frac{2^{-3/2}\,3^{-1/5}\,5^{1/4}\left(\sqrt{5}-1\right)^{1/2}\left[\Gamma\left(\frac{2}{5}\right)\right]^2}{\sin\left(\frac{4}{15}\pi\right)} \tag{78}$$

$$\frac{\Gamma\left(\frac{1}{15}\right)\Gamma\left(\frac{2}{15}\right)\Gamma\left(\frac{4}{15}\right)}{\Gamma\left(\frac{7}{15}\right)} = 60\left(\sqrt{5}-1\right)\sin\left(\frac{7}{15}\pi\right)\left[K\left(k_{15}\right)\right]^2 \tag{79}$$

$$\frac{\Gamma\left(\frac{1}{20}\right)\Gamma\left(\frac{9}{20}\right)}{\Gamma\left(\frac{3}{20}\right)\Gamma\left(\frac{7}{20}\right)} = 2^{-1}\,5^{1/4}\left(\sqrt{5}+1\right) \tag{80}$$

$$\frac{\Gamma\left(\frac{1}{20}\right)\Gamma\left(\frac{3}{20}\right)}{\Gamma\left(\frac{7}{20}\right)\Gamma\left(\frac{9}{20}\right)} = 2^{4/5}\left(10-2\sqrt{5}\right)^{1/2}\pi^{-1}\sin\left(\frac{7}{20}\pi\right)\sin\left(\frac{9}{20}\pi\right)\left[\Gamma\left(\frac{1}{5}\right)\right]^2 \tag{81}$$

$$\frac{\Gamma\left(\frac{1}{20}\right)\Gamma\left(\frac{7}{20}\right)}{\Gamma\left(\frac{3}{20}\right)\Gamma\left(\frac{9}{20}\right)} = 2^{3/5}\left(10+2\sqrt{5}\right)^{1/2}\pi^{-1}\sin\left(\frac{3}{20}\pi\right)\sin\left(\frac{9}{20}\pi\right)\left[\Gamma\left(\frac{2}{5}\right)\right]^2 \tag{82}$$

$$\Gamma\left(\frac{1}{20}\right)\Gamma\left(\frac{3}{20}\right)\Gamma\left(\frac{7}{20}\right)\Gamma\left(\frac{9}{20}\right) = 160\left(\sqrt{5}-2\right)^{1/2}\pi\left[K\left(k_5\right)\right]^2. \tag{83}$$

Several of these are also given in Campbell (1966, p. 31).

A few curious identities include

$$\prod_{n=1}^{2}\Gamma\left(\frac{1}{3}n\right) = \frac{2\pi}{\sqrt{3}} \tag{84}$$

$$\prod_{n=1}^{3}\Gamma\left(\frac{1}{3}n\right) = \frac{2\pi}{\sqrt{3}} \tag{85}$$

$$\prod_{n=1}^{4}\Gamma\left(\frac{1}{3}n\right) = \frac{2\pi\Gamma\left(\frac{1}{3}\right)}{3\sqrt{3}} \tag{86}$$

$$\prod_{n=1}^{5}\Gamma\left(\frac{1}{3}n\right) = \frac{8}{27}\pi^2 \tag{87}$$

$$\prod_{n=1}^{6}\Gamma\left(\frac{1}{3}n\right) = \frac{8}{27}\pi^2 \tag{88}$$

$$\prod_{n=1}^{7}\Gamma\left(\frac{1}{3}n\right) = \frac{32}{243}\pi^2\,\Gamma\left(\frac{1}{3}\right) \tag{89}$$

$$\prod_{n=1}^{8}\Gamma\left(\frac{1}{3}n\right) = \frac{640\,\pi^3}{2187\sqrt{3}}, \tag{90}$$

of which Magnus and Oberhettinger 1949, p. 1 give only the last case,

$$\frac{\left[\Gamma\left(\frac{1}{4}\right)\right]^4}{16\pi^2} = \frac{3^2}{3^2-1}\,\frac{5^2-1}{5^2}\,\frac{7^2}{7^2-1}\cdots, \tag{91}$$

and

$$\frac{\Gamma'(1)}{\Gamma(1)} - \frac{\Gamma'\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = 2\ln 2 \tag{92}$$

(Magnus and Oberhettinger 1949, p. 1). Ramanujan also gave a number of fascinating identities:

$$\frac{\Gamma^2(n+1)}{\Gamma(n+x\,i+1)\,\Gamma(n-x\,i+1)} = \prod_{k=1}^{\infty}\left[1+\frac{x^2}{(n+k)^2}\right] \tag{93}$$

$$\phi(m,n)\,\phi(n,m) = \frac{\Gamma^3(m+1)\,\Gamma^3(n+1)}{\Gamma(2\,m+n+1)\,\Gamma(2\,n+m+1)}\,\frac{\cosh\left[\pi(m+n)\sqrt{3}\right]-\cos\left[\pi(m-n)\right]}{2\,\pi^2\left(m^2+m\,n+n^2\right)}, \tag{94}$$

where

$$\phi(m,n) \equiv \prod_{k=1}^{\infty}\left[1+\left(\frac{m+n}{k+m}\right)^3\right], \tag{95}$$

$$\prod_{k=1}^{\infty}\left[1+\left(\frac{n}{k}\right)^3\right]\prod_{k=1}^{\infty}\left[1+3\left(\frac{n}{n+2\,k}\right)^2\right] = \frac{\Gamma\left(\frac{1}{2}\,n\right)}{\Gamma\left[\frac{1}{2}\,(n+1)\right]}\,\frac{\cosh\left(\pi\,n\sqrt{3}\right)-\cos(\pi\,n)}{2^{n+2}\,\pi^{3/2}\,n} \tag{96}$$

(Berndt 1994).

Ramanujan gave the infinite sums

$$\begin{aligned}&\sum_{k=0}^{\infty}(8\,k+1)\left[\frac{\Gamma\left(k+\frac{1}{4}\right)}{k!\,\Gamma\left(\frac{1}{4}\right)}\right]^4\\ &= 1+9\left(\frac{1}{4}\right)^4+17\left(\frac{1\cdot 5}{4\cdot 8}\right)^4+25\left(\frac{1\cdot 5\cdot 9}{4\cdot 8\cdot 12}\right)^4+\ldots\\ &= \frac{2^{3/2}}{\sqrt{\pi}\left[\Gamma\left(\frac{3}{4}\right)\right]^2}\end{aligned} \tag{97}$$

and

$$\begin{aligned}&\sum_{k=0}^{\infty}(-1)^k(4\,k+1)\left[\frac{(2\,k-1)!!}{(2\,k)!!}\right]^5\\ &= 1-5\left(\frac{1}{2}\right)^5+9\left(\frac{1\cdot 3}{2\cdot 4}\right)^5-13\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^5+\ldots\\ &= \frac{2}{\left[\Gamma\left(\frac{3}{4}\right)\right]^4}\end{aligned} \tag{98}$$

(The following asymptotic series is occasionally useful in probability theory (e.g., the one-dimensional random walk):

$$\frac{\Gamma\left(J+\frac{1}{2}\right)}{\Gamma(J)} = \sqrt{J}\left(1 - \frac{1}{8J} + \frac{1}{128J^2} + \frac{5}{1024J^3} - \frac{21}{32\,768J^4} + \ldots\right) \tag{99}$$

(Graham *et al.* 1994). This series also gives a nice asymptotic generalization of Stirling numbers of the first kind to fractional values.

It has long been known that $\Gamma\left(\frac{1}{4}\right)\pi^{-1/4}$is transcendental (Davis 1959), as is $\Gamma\left(\frac{1}{3}\right)$(Le Lionnais 1983; Borwein and Bailey 2003, p. 138), and Chudnovsky has apparently recently proved that $\Gamma\left(\frac{1}{4}\right)$is itself transcendental (Borwein and Bailey 2003, p. 138).

There exist efficient iterative algorithms for $\Gamma(k/24)$for all integers k(Borwein and Bailey 2003, p. 137). For example, a quadratically converging iteration for $\Gamma(1/4) = 3.6256099\ldots$(Sloane's A068466) is given by defining

$$x_n = \frac{1}{2}\left(x_{n-1}^{1/2} + x_{n-1}^{-1/2}\right) \tag{100}$$

$$y_n = \frac{y_{n-1}\,x_{n-1}^{1/2} + x_{n-1}^{-1/2}}{y_{n-1}+1}, \tag{101}$$

setting $x_0 = \sqrt{2}$ and $y_1 = 2^{1/4}$, and then

$$\Gamma\left(\frac{1}{4}\right) = 2\left(1+\sqrt{2}\right)^{3/4}\left[\prod_{n=1}^{\infty} x_n^{-1}\left(\frac{1+x_n}{1+y_n}\right)^3\right]^{1/4} \tag{102}$$

FINANCIAL ACCOUNTING & ANALYSIS (105)

MEANING AND SCOPE OF ACCOUNTING

DEFINITION

Financial accountancy (or financial accounting) is the field of accountancy concerned with the preparation of financial statements for decision makers, such as stockholders, suppliers, banks, employees, government agencies, owners and other stakeholders.

In 1966, American accounting Association (AAA) defined accounting as "Accounting is the process of identifying, measuring and communicating economic information to permit the informed judgements ands decisions by the users of the information."

In 1941, American Institute of Certified public Accountants (AICPA) defined accounting as "Accounting is the art of recording, classifying and summarizing in significant manner and in terms of money, transactions and events which are, in part, at least of a financial character and interpreting the results thereof."

Financial capital maintenance can be measured in either nominal monetary units or units of constant purchasing power. The central need for financial accounting is to reduce the various principal-agent problems, by measuring and monitoring the agents' performance and thereafter reporting the results to interested users.

SCOPE OF ACCOUNTING:

Accounting as compared to book-keeping has a very wide and huge role to play in the businesses whether there is a small firm engaged in few transactions to a large MNC with multiple transactions daily, Accounting is required everywhere. Accounting not only maintains the records but also analyzes and interprets the results also. Accounting is being done in any nature and size of the firm, example: schools, hospitals, banks, retail shops. Nowadays Accountants are serving the accounting requirements of various businesses and organizations for which they require expertise in the same field.

Accounting has grown in its importance today as a discipline that provides results online in a quick and accessible form that can be used by the management for decision-making.

It has been recognized as a tool for mastering the economic problems that a business organization may have to face.

There are several parties now connected with the accounting information they are as follows:

- Proprietors
- Managers
- Suppliers
- Investors
- Government
- Public

  Thus accounting is being used anywhere and everywhere as it is the need of an hour to give rise to any business event that occurs daily many a times as it accounts for the profitability and returns of the business.

NATURE OF ACCOUNTING:

Accounting as a discipline has been originated to serve the purpose of the organization in maintaining and updating the records and transactions of the various business events on a day-to-day basis.

It functions as a means from which various parties connected to the business can get the accounting information and the management is able to do the decision-making for various business plans and policies.

The accountants are hired for maintaining and preparing the records and financial statements, provide the results and also conduct analysis and conclusions drawn from them.

From the above on e can derive the basic nature as follows:

1. It keeps and maintains the record of the business.
2. It also helps to analyze and interpret the financial data.
3. It is useful in decision-making.
4. It is useful in preparation and compilation of financial results.
5. It also serves as a means to depict the financial position of the business.
6. It ignores the qualitatitive aspects of the business.
7. The information provided in accounting is based on estimates.

FUNCTIONS OF ACCOUNTING

1. Record Keeping
2. Protecting of properties
3. Communication of results

4. Meeting legal requirements

OBJECTIVES OF ACCOUNTING ARE

1. To keep systematic records: Its main objective is to keep complete record of Business transactions. It avoids the possibility of omission and fraud.
2. To calculate profit or loss: Accounting helps to ascertain the net profit earned or loss suffered on account of business transactions during a particular period.
   To ascertain profit or loss at the end of each accounting period Trading and Profit& Loss is prepared.
3. To facilitate decision making: The information collected from various financial statements is useful in decision-making and is of use to the parties connected to the business.
4. To ascertain the financial position of the business: The businessman is able to know the financial position of the firm through the profits made or losses incurred by the firm during the accounting period as well as the balance-sheet serve as a basis of serving the financial status of the firm at a particular date.

RELATIONSHIP OF ACCOUNTING WITH ECONOMICS

Prof. Robbins has defined term ‘economics’ as follows-

“Economics is the science, which studies human behavior as a relationship between ends and scarce means which have alternatives uses.”However when a person is to take any economical decision, he has to depend mainly on the accounting information. Generally an accountant is concerned with the economic problems of an only one enterprise only, but an economist is concerned with the problems of an industry as a whole. Micro levels data, arranged by the accounting system, is summed up to get macro level data base. Thus at the macro level, accounting provides the basic data, upon which the economic models are developed.

However, there exists a wide gulf between economists’ and accountants’ concept of income and capital For example, the profit according to an economist is not same thing as the profit according to an accountant. No doubt, accountants have derived the ideas of value, income and capital maintenance from economists, but suitably modified to make them usable in practical circumstances. Thus, Economics and Accounting are close related subjects.

Relationship of Accounting with Mathematics

Knowledge of arithmetic and algebra is a pre-requisite for accounting computations and measurements. Calculations of interest and annuity etc. are some examples of fundamental uses of mathematics in accounting.

Presently graphs and charts are being widely used for communicating accounting information to the users. Thus the knowledge in geometry and trigonometry has become essential to have a better understanding about the accounting communication system.

Relationship of Accounting with Statistics

Collection, Tabulation, Analysis and presentation of data are some primary functions, which are performed by both Accountants and statisticians.The accountant is mainly concerned with the monetary data, although to some extent, he is also concerned with the quantitative data. But a statistician is concerned equally with the monetary and quantitative data.The use of statistics is accounting can be appreciated better in the context of the nature of accounting records. Accounting information is very precise; it is exact to the last paisa. But for decision-making purposes, such precision is not necessary and hence the statistical approximations are sought. Accounting records generally confined to one year, while statistical analysis is more useful if a longer period is taken. For example, a longer period will be required to fit the trend line. Statistical methods are helpful in developing accounting data and in their interrelation. Therefore the study and application of statistical methods will add extra edge to the accounting data.

Relationship of Accounting with Law

Every business house has to work within legal environment. All the transactions with suppliers and customers are governed by the Contract Act, Sale of Goods Act, Negotiable Instruments Act, etc. The entity, itself, created and controlled by laws. For example a partnership business is controlled by Partnership Act, a company is created and controlled by the Companies Act. Very often the accounting system to be followed has been prescribed by the law. For example the Companies Act has prescribed the format of financial statements.However legal prescription about the accounting system is the product of development in accounting knowledge. That is to say legislation about accounting system cannot be enacted unless there is a corresponding development in the accounting discipline. In what way, accounting influences law and is also influenced by law.

Relationship of Accounting with Management

Management is a broad occupational field, which comprises many functions and application of many disciplines including statistics, mathematics, economics, etc. Accountants are well placed in the management and play a key role in the management them. A large portion of accounting information is prepared for management decision-making. In the management team, an accountant is in a better position to understand and use such data. In other words, since an accountant plays an active role in management, he understands the data requirement. So, accounting system can be molded to serve the management purpose.

<u>Advantages of Financial Accounting:</u>

• It provides legal information to stakeholders such as financial accounts in the form of trading, profit and loss account and balance sheet.

• It shows the mode of investment for shareholders.

• It provides business trade credit for suppliers.

• It notifies the risks of loan in business for banks and lenders.

<u>Limitations of Financial Accounting:</u>

One of the major limitations of financial accounting is that it does not take into
Account the non-monetary facts of the business like the competition in the market, change in the value for money etc.

The following limitations of financial accounting have led to the development of cost
Accounting:

1. **<u>No clear idea of operating efficiency</u>**: You will agree that, at times, profits may be more or less, not because of efficiency or inefficiency but because of inflation or trade depression. Financial accounting will not give you a clear picture of operating efficiency when prices are rising or decreasing because of inflation or trade depression.
2. **<u>Weakness not spotted out by collective results</u>**: Financial accounting discloses only the

net result of the collective activities of a business as a whole. It does not indicate profit or loss of each department, job, process or contract. It does not disclose the exact cause of inefficiency i.e.it does not tell where the weakness is because it discloses the net profit of all the activities of a business as a whole. Say, for instance, it can be compared with a reading on a thermometer. A reading of more than 98.4° or less than 98.4º discloses that something is wrong with the human body but the exact disease is not disclosed. Similarly, loss or less profit disclosed by the profit and loss account is a signal of bad performance of the business in whole, but the exact cause of such performance is not identified.

3. **Not helpful in price fixation:** In financial accounting, costs are not available as an aid in determining prices of the products, services, production order and lines of products.

4. **No classification of expenses and accounts**: In financial accounting, there is no such system by which accounts are classified so as to give relevant data regarding costs by departments, processes, products in the manufacturing divisions, by units of product lines and sales territories, by departments, services and functions in the administrative division. Further expenses are not attributed as to direct and indirect items. They are not assigned to the products at each stage of production to show the controllable and uncontrollable items of overhead costs.

5. **No data for comparison and decision-making**: It will not provide you with useful data for comparison with a previous period. It also does not facilitate taking various financial decisions like introduction of new products, replacement of labour by machines, price in normal or special circumstances, producing a part in the factory or sourcing it from the market, production of a product to be continued or given up, priority accorded to different products and whether investment should be made in new products etc.

6. **No control on cost**: It does not provide for a proper control of materials and supplies, wages, labour and overheads.

7. **No standards to assess the performance**: In financial accounting, there is no such well developed system of standards, which would enable you to appraise the efficiency of the

organization in using materials, labour and overhead costs. Again, it does not provide you any such information, which would help you to assess the performance of various persons and departments in order that costs do not exceed a reasonable limit for a given quantum of work of the requisite quality.

Basis of Accounting:

There are basically two systems of accounting:

1. **Cash System of Accounting.**
2. **Accrual System of Accounting.**

Cash system of Accounting:

It is a system in which accounting entries are made only when cash is received or paid. No entry is made when a payment or receipt is merely due. For example, the rent for December 2009 has not been paid till January 10the 2010. Under cash basis, rent expense for the month of December will not be recorded as payment has not been made. Government system of accounting is mostly on the cash system.

Accrual System of Accounting:

It is a system in which accounting entries are made on the basis of amount having become due for payment or receipt. This system recognizes the fact that if a transaction or an event occurred, its consequences cannot be avoided and therefore, should be brought into book in order to present a meaningful picture of profit earned or loss suffered.

Accounting Principles-Concepts and Conventions

Meaning of Accounting Principles:

Financial accounting is information that must be processed and reported objectively. Third parties, who must rely on such information, have a right to be assured that the data is free from bias and inconsistency, whether deliberate or not. For this reason, financial accounting relies on certain standards or guides that are called 'Generally Accepted Accounting Principles' (GAAP). Principles derived from tradition, such as the concept of matching. In any report of financial statements (audit, compilation, review, etc.), the preparer/auditor must indicate to the reader whether or not the information contained within the statements complies with GAAP.

ACCOUNTING PRINCIPLES

• **Principle of regularity**: Regularity can be defined as conformity to enforced rules and laws.

• **Principle of consistency**: This principle states that when a business has fixed a specific method for the accounting treatment of an item, it will enter all similar items that follow, in exactly the same way.

• **Principle of sincerity**: According to this principle, the accounting unit should reflect in good faith the reality of the company's financial status.

• **Principle of the permanence of methods**: This principle aims at maintaining the coherence and comparison of the financial information published by the company.

Principle of non-compensation: One should show the full details of the financial information and not seek to compensate a debt with an asset, revenue with an expense etc.

• **Principle of prudence**: This principle aims at showing the reality 'as is': one should not try to make things look rosier than they are. Typically, revenue should be recorded only when it is certain and a provision should be entered for an expense, which is probable.

• **Principle of continuity**: When stating financial information, one assumes that business will not be interrupted. This principle mitigates the principle of prudence: assets do not have to be accounted at their disposable value, but it is accepted that they are at their historical value.

• **Principle of periodicity**: Each accounting entry should be allocated to a given period and split accordingly if it covers several periods. If a client pre-pays a subscription (or lease, etc.), the given revenue should be split to the entire time-span and not accounted for entirely on the date of the transaction.

• **Principle of full disclosure/materiality**: All information and values pertaining to the financial position of a business must be disclosed in the records.

<u>Accounting Concepts and Conventions:</u>

An accounting convention is a modus operandi of universally accepted system of recording and presenting accounting information to the concerned parties. They are followed judiciously and rarely ignored. Accounting conventions are evolved through the regular and consistent practice over the years to aid unvarying recording in the books of accounts. Accounting conventions help in comparing accounting data of different business units or of the same unit for different periods. These have been developed over the years.

1. **<u>Convention of relevance</u>**: The convention of relevance emphasizes the fact that only such information should be made available by accounting that is pertinent and helpful for achieving its objectives. The relevance of the items to be recorded depends on its nature

and the amount involved. It includes information, which will influence the decision of its client. This is also known as convention of materiality. For example, business is interested in knowing as to what has been the total labor cost. It is neither interested in knowing the amount employees spend nor what they save.

2. **Convention of feasibility**: The convention of feasibility emphasizes that the time, labor and cost of analyzing accounting information should be comparable to the benefits arising out of it. For example, the cost of 'oiling and greasing' the machinery is so small that its break-up per unit produced will be meaningless and will amount to wastage of labor and time of the accounting staff.

3. **Convention of consistency**: The convention of consistency means that the same accounting principles should be used for preparing financial statements year on year. An evocative conclusion can be drawn from financial statements of the same enterprise when there is similarity between them over a period of time. However, these are possible only when accounting policies and practices followed by the enterprise are uniform and consistent over a period. If dissimilar accounting procedures and practices are followed for preparing financial statements of different accounting years, then the result will not be analogous.Generally, a businessman follows the above-mentioned general practices or methods year after year. Forexample, while charging depreciation on fixed assets or valuing unsold stock, if a particular method is used it should be followed year after year, so that the financial statements can be analyzed and a comparison made.

4. **Convention of objectivity**: The convention of objectivity highlights that accounting information should be measured and expressed by the standards which are universally acceptable. For example, unsold stock of goods at the end of the year should be valued at cost price or market price, whichever is less and not at a higher price even if it is likely to be sold at a higher price in the future.

5. **Convention of full disclosure**: Convention of full disclosure states that all material and relevant facts concerning financial statements should be fully disclosed. Full disclosure means that there should be complete, reasonable and sufficient disclosure of accounting information.Full refers to complete and detailed presentation of information. Thus, the

convention of full disclosure suggests that every financial statement should disclose all pertinent information. For example, the business provides financial information to all interested parties like investors,lenders, creditors, shareholders etc. The shareholder would like to know the profitability of the firm while the creditors would like to know the solvency of the business. This is only possible if the financial statement discloses all relevant information in a complete, fair and an unprejudiced manner.

6. **Convention of conservatism**: This concept accentuates that profits should never be overstated or anticipated. However, if the business anticipates any loss in the near future, provision should be made for it in the books of accounts, for the same. For example, creating provision for doubtful debts, discount on debtors, writing off intangible assets like goodwill, patent and so on should be taken in to consideration. Traditionally, accounting follows the rule 'anticipate no profit and provide for all possible losses. ‘For example, the closing stock is valued at cost price or market price, whichever is lower. The effect of the above is that in case market price has come down then provides for the 'anticipated loss', but if the market price has increased then ignore 'anticipated profits'. The convention of conservatism is a valuable tool in situation of ambiguity and qualms.

Introduction To Accounting Standards:

ACCOUNTING STANDARDS ISSUED BY ICAI

AS-1 DISCLOSURE OF ACCOUNTING POLICIES

AS-2 VALUATION OF INVENTORIES

AS-3 CASH FLOW STATEMENTS

AS-4 CONTINGECIES AND EVENTS OCCURING AFTER THE BALANCE SHEET DATE

AS-5 NET PROFIT OR LOSS FOR THE PERIOD, PRIOR PERIOD ITEMS AND CHANGES IN ACCOUNTING POLICIES

AS-6 DEPRECIATION ACCOUNTING

AS-7 ACCOUNTING FOR CONSTRUCTION CONTRACTS

AS-8 ACCOUNTING FOR RESEARCH AND DEVELOPMENT

AS-9 REVENUE RECOGNITION

AS-10 ACCOUNTING FOR FIXED ASSETS

AS-11 ACCOUNTING FOR THE EFFECTS OF CHANGES IN FOREIGN EXCHANGE RATES

AS-12 ACCOUNTING FOR GOVERNMENT GRANTS

AS-13 ACCOUNTING FOR INVESTMENTS

AS-14 ACCOUNTING FOR AMALGAMATIONS

AS-15 ACCOUNTING FOR RETIREMENT BENEFITS IN THE FINANCIAL STATEMENTS OF EMPLOYERS

AS-16 BORROWING COSTS

Basic Concepts of Accounting Standards:

Basic concepts of (GAAP) Accounting Standards :

In order to achieve the aforesaid objectives of GAAP and implement fundamental qualities, the set accounting standards feature four basic assumptions as listed below:

- **Going concern**

This assumption of GAAP accounting standards presumes that the business stays in operation indefinitely thus validating the techniques of asset capitalization, amortization, and depreciation. This assumption is, however, not applicable in case of liquidation. The business is believed to continue in the unforeseeable future.

- **Monetary Unit Principle**

This assumption presumes a stable currency going to be the unit of record.

- **Accounting Entity**

This assumption presumes the business to individually exist from its owners or other business entities. Also, revenue and expense need to be kept separate from personal expenses.

- **Time-period principle**

This assumption states that an entity's economic activities can be divided into simulated time-periods.

Benefits of Accounting Standards:

As per the accounting standards presented by GAAP, the financial reports should provide info which is:

- Useful in being presented to potential creditors and investors in addition to other users in making cogent investment, credit, and similar financial decisions.
- Helpful for the potential creditors and investors in addition to other users in evaluating the timing, amounts, and uncertainty of probable cash receipts.
- Related to economic resources, the claims to these resources, and the changes occurring in them.
- Helpful in taking financial decisions. ·

Helpful in taking long-term decisions.

· Helpful in improving the business' performance. ·

Useful in maintaining records.

Procedure for Issuing Accounting Standards in India

A summarized extract of the text of the "Preface to the Statements of Accounting Standards (Revised 2004)," issued by the council of the Institute of Chartered Accountants of India, explains the procedure of issuing Accounting Standards. They are:

1. The ASB determines the broad areas requiring formulation of Accounting Standards and lists them according to priority.
2. In the preparation of Accounting Standards, the ASB is assisted by a Study Group, constituted for this purpose. Views of government, public sector undertakings, industry and other organizations are obtained before formulating the Exposure Draft.
3. The Exposure Draft comprises the following:

· Objective and scope of the standard.

· Definition of the terms used in the standard.

· The manner in which the accounting principles have been applied for formulating the standard.

· The presentations and disclosure requirements of it comply with the standard. ·

Class of enterprises to which the standard will apply.

· Date from which the standard will be effective.

Need and Significance of IFRS:

International Financial Reporting Standards (IFRS) are principles-based Standards, Interpretations and the Framework adopted by the International Accounting Standards Board (IASB). IFRS represent a set of internationally accepted accounting principles used by companies to prepare financial statements.

The goal with IFRS is to make international comparisons as easy as possible. More than 100 countries around the world currently require or permit IFRS reporting. Approximately 85 of those countries require IFRS reporting for all domestic, listed companies. All member states of the EU are required to use IFRS as adopted by the EU for listed companies since 2005. The US is also gearing towards IFRS. While some countries require all companies to adhere to IFRS, others merely allow it or try to coordinate their own country's standards to be similar.

IFRS include the following Standards:

· IFRS 1 First-time Adoption of International Financial Reporting Standards. ·
IFRS 2 Share-based Payment.

· IFRS 3 Business Combinations. ·
IFRS 4 Insurance Contracts.

· IFRS 5 Non-current Assets Held for Sale and Discontinued Operations. ·
IFRS 6 Exploration for and Evaluation of Mineral Resources.

· IFRS 7 Financial Instruments: Disclosures. ·
IFRS 8 Operating Segments.

· IFRS 9 Financial Instruments.

· IAS 1 Presentation of Financial Statements. ·
IAS 2 Inventories.

· IAS 7 Statement of Cash Flows.

· IAS 8 Accounting Policies, Changes in Accounting Estimates and Errors. ·
IAS 10 Events after the Reporting Period.

· IAS 11 Construction Contracts. ·
IAS 12 Income Taxes.

· IAS 16 Properties, Plant and Equipment.

· IAS 17 Leases. ·
IAS 18 Revenue.

· IAS 19 Employee Benefits.

· IAS 20 Accounting for Government Grants and Disclosure of Government Assistance.

· IAS 21 The Effects of Changes in Foreign Exchange Rates. ·
IAS 23 Borrowing Costs.

· IAS 24 Related Party Disclosures.

· IAS 26 Accounting and Reporting by Retirement Benefit Plans. ·
IAS 27 Consolidated and Separate Financial Statements.
· IAS 28 Investments in Associates.
· IAS 29 Financial Reporting in Hyperinflationary Economies. ·
IAS 31 Interests in Joint Ventures.
· IAS 32 Financial Instruments: Presentation. ·
IAS 33 Earnings Per Share.
· IAS 34 Interim Financial Reporting. ·
IAS 36 Impairment of Assets.
· IAS 37 Provisions, Contingent Liabilities and Contingent Assets. ·
IAS 38 Intangible Assets.
· IAS 39 Financial Instruments: Recognition and Measurement. ·
IAS 40 Investment Property.
· IAS 41 Agriculture.

International Accounting Standards (IAS) are the older standards that IFRS are gradually replacing (IAS were issued from 1973 to 2000).

XBRL:

XBRL(Extensive Business Reporting Language) is a freely available and global standard for exchanging business information. XBRL allows the expression of semantic meaning commonly required in business reporting. The language is XML-based and uses the XML syntax and related XML technologies such as XML Schema, XLink, XPath, and Namespaces. One use of XBRL is to define and exchange financial information, such as a financial statement. The XBRL Specification is developed and published by XBRL International, Inc. (**XII**).

XBRL is a standards-based way to communicate and exchange business information between business systems. These communications are defined by metadata set out in taxonomies, which capture the definition of individual reporting concepts as well as the relationships between concepts and other semantic meaning. Information being communicated or exchanged is provided within an XBRL instance.

Early users of XBRL included regulators such as the U.S. Federal Deposit Insurance Corporation and the Committee of European Banking Supervisors (CEBS). Common functions in many countries that make use of XBRL include regulators of stock exchanges and securities, banking regulators, business registrars, revenue reporting and tax-filing agencies, and national statistical agencies.

UNIT-2 JOURNALISING TRANSACTIONS

Rules of Debit and Credit

Any account that obtains a benefit is Debit.

OR

Anything that will provide benefit to the business is Debit.

Both these statements may look different but in fact if we consider that whenever an account benefits as a result of a transaction, it will have to return that benefit to the business then both the statements will look like different sides of the same picture. For credit, Any account that provides a benefit is **Credit.**

OR

Anything to which the business has a responsibility to return a benefit in future is **credit**. As explained in the case of Debit, whenever an account provides benefit to the business the business will have a responsibility to return that benefit at some time in future and so it is Credit.

***Rules of Debit and Credit for Assets**

Similarly we have established that whenever a business transfers a value / benefit to an account and as a result creates something that will provide future benefit; the `thing' is termed as **Asset**. By combining both these rules we can devise following rules of Debit and Credit for Assets: When an asset is created or purchased, value / benefit is transferred to that account, so it is debited.

I. Increase in Asset is Debit

Reversing the above situation if the asset is sold, which is termed as disposing off, for say cash, the asset account provides benefit to the cash account. Therefore, the asset account is credited.

II. Decrease in Asset is Credit

***Rules of Debit and Credit for Liabilities**

Anything that transfers value to the business, and in turn creates a responsibility on part of the business to return a benefit, is a **Liability**. Therefore, liabilities are the exact opposite of the assets.

When a liability is created the benefit is provided to business by that account so it is Credited

III. Increase in Liability is Credit

When the business returns the benefit or repays the liability, the liability account benefits from the business. So it is debited

IV. Decrease in Liability is Debit

***Rules of Debit and Credit for Expenses**

Just like assets, we have to pay for expenses. From assets, we draw benefit for a long time whereas the benefit from expenses is for a short run. Therefore, Expenditure is just like Asset but for a short run.

Using our rule for Debit and Credit, when we pay cash for any expense that expense account benefits from cash, therefore, it is debited.

Now we can lay down our rule for Expenditure:

V. Increase in Expenditure is Debit

Reversing the above situation, if we return any item that we had purchased, we will receive cash in return. Cash account will receive benefit from that Expenditure account. Therefore, Expenditure account will be credited

VI. Decrease in Expenditure is Credit

***Rules of Debit and Credit for Income**

Income accounts are exactly opposite to expense accounts just as liabilities are opposite to that of assets.

Therefore, using the same principle we can draw our rules of Debit and Credit for Income.

VII. Increase in Income is Credit

VIII. Decrease in Income is Debit

<u>Posting and Preparation of Trial Balance:</u>

<u>Trial Balance:</u>

Trial Balance is a list of closing balances of ledger accounts on a certain date and is the first step towards the preparation of financial statements. It is usually prepared at the end of an accounting period to assist in the drafting of financial statements. Ledger balances are segregated into debit balances and credit balances. Asset and expense accounts appear on the debit side of the trial balance whereas liabilities, capital and income accounts appear on the credit side. If all accounting entries are recorded correctly and all the ledger balances are accurately extracted, the total of all debit balances appearing in the trial balance must equal to the sum of all credit balances.

How to prepare a Trial Balance:

Following Steps are involved in the preparation of a Trial Balance:

1. All Ledger Accounts are closed at the end of an accounting period.
2. Ledger balances are posted into the trial balance.
3. Trial Balance is cast and errors are identified.
4. Suspense account is created to agree the trial balance totals temporarily until corrections are accounted for.
5. Errors identified earlier are rectified by posting corrective entries.
6. Any adjustments required at the period end not previously accounted for are incorporated into the trial balance.

Closing Ledger Accounts:

Ledger accounts are closed at the end of each accounting period by calculating the totals of debit and credit sides of a ledger. The difference between the sum of debits and credits is known as the closing balance. This is the amount which is posted in the trial balance.

How closing balances are presented in the ledger depends on whether the account is related to income statement (income and expenses) or balance sheet (assets, liabilities and equity). Balance sheet ledger accounts are closed by writing 'Balance c/d' next to the balancing figure since these are to be rolled forward in the next accounting period. Income statement ledger accounts on the other hand are closed by writing 'Income Statement' next to the residual amount because it is being transferred to the income statement as revenue or expense incurred for the period.

The steps involved in closing a ledger account may be summarized as below:

1. Add the totals of both sides of a ledger

2. The higher of the totals among the debit side and credit side must be inserted at the end of **BOTH** sides. Closing balance is the balancing figure on the side with the lower balance.
3. In case of ledger accounts of assets, liabilities and equity, 'balance c/d' is written next to the closing balance whereas in case of income and expenses ledger accounts, 'Income Statement' is written next to the closing balance.
4. The closing balances of all ledger accounts are posted into the trial balance.

Closing Balance of all ledger accounts are posted into the trial balance. It is important to remember that a debit closing balance in the ledger account appears on the credit side but in the trial balance it is presented in the debit column and vice versa.

Posting of closing balances should be done carefully as many errors may occur during the posting process such as Posting Error, Transposition Errors and Slide error.

Following is an example of a trial balance prepared from the closing balances of the ledgers detailed above.

ABC LTD **Trial Balance as at 31 December 2011**		
Account Title	**Debit (in rupees)**	**Credit (in rupees)**
Share Capital		10,000
Bank Loan		10,000
Cash	30,000	
Salaries Expense	5,000	
Sales Revenue		15,000
Total	**35,000**	**35,000**

<u>Capital and Revenue:</u>

<u>Capital:</u> The amount invested in the firm by the owner is referred to as the Capital. The expenditure borne/incurred in purchase of fixed assets, investments is known as Capital Expenditure.

<u>Revenue:</u> The profit/gain arising from the business from the sale proceeds is referred to as the revenue. Revenue includes the total cost/expenses (whether direct or indirect) and the total income earned during the accounting period by the firm.

<u>Classification of Income:</u>

- Gain on sale of Investments.
- Capital Income.
- Dividends Received.
- Interest Received.
- Commission Received.

<u>Classification of Expenditure:</u>

in Fixed Assets.

- Purchase of Land & Building.
- Repairs and Installation of Tools & Equipments.
- Payment of direct and indirect expenses.
- Deferred Revenue Expenditure.
- Depreciation on Fixed Assets.

<u>Capital and Revenue Expenditure</u>

Expenditure on fixed assets may be classified into Capital Expenditure **and** Revenue Expenditure. **The distinction between the nature of capital and revenue expenditure is important as only capital expenditure is included in the cost of fixed asset.**

<u>Capital Expenditure:</u>

Capital expenditure includes costs incurred on the acquisition of a fixed asset and any subsequent expenditure that increases the earning capacity of an existing fixed asset.

The cost of acquisition not only includes the cost of purchases but also any additional costs incurred in bringing the fixed asset into its present location and condition (e.g. delivery costs).

Capital expenditure, as opposed to revenue expenditure, is generally of a one-off kind and its benefit is derived over several accounting periods. Capital Expenditure may include the following:

- Purchase costs (less any discount received)
- Delivery costs
- Legal charges
- Installation costs
- Up gradation costs
- Replacement costs

Revenue Expenditure

Revenue expenditure incurred on fixed assets include costs that are aimed at 'maintaining' rather than enhancing the earning capacity of the assets. These are costs that are incurred on a regular basis and the benefit from these costs is obtained over a relatively short period of time. For example, a company buys a machine for the production of biscuits. Whereas the initial purchase and installation costs would be classified as capital expenditure, any subsequent repair and maintenance charges incurred in the future will be classified as revenue expenditure. This is so because repair and maintenance costs do not increase the earning capacity of the machine but only maintains it (i.e. machine will produce the same quantity of biscuits as it did when it was first put to use).

Revenue costs therefore comprise of the following:

- Repair costs
- Maintenance charges
- Repainting costs
- Renewal expenses

As revenue costs do not form part of the fixed asset cost, they are expensed in the income statement in the period in which they are incurred.

Journal and subsidiary Books:

JOURNAL

The word ‘Journal’ has been derived from the French word ‘JOUR’ means daily records. **Journal** is a book of original entry in which transactions are recorded as and when they occur in chronological order (in order of date) from source documents. Recording in journal is made showing the accounts to be debited and credited in a systematic manner. Thus, the journal provides a date-wise record of all the transactions with details of the accounts and amounts debited and credited for each transaction with a short explanation, which is known as narration. Firms having limited number of transactions record those in journal and from there post these to the concerned ledger accounts. Firms having large number of transactions, maintain some special purpose journals such as, Purchase Book, Sales Books, Returns books, Bills Book, Cash Book, Journal proper etc.

COMPOUND JOURNAL ENTRY

A compound journal entry is an accounting entry which effects more than two account heads. A simple journal entry has one debit and one credit whereas a compound journal entries includes one or more debits and/or credits than a simple journal entry. A compound journal entry may combine two or more debits and a credit, or a debit and two or more credits, or two or more of both debits and credits.

OPENING ENTRY

In the case of continuing business we are required to pass an entry in the journal for bringing in the new books all assets and liabilities as appearing in the books on the last day of the previous year. This entry is known as 'opening entry'. Rule of passing opening entry is to debit each asset account; credit each liability account; excess of debits over credits represents capital balance.

SUB DIVISION OF JOURNAL

CASH JOURNAL

The number of cash transactions in a firm is generally larger, therefore, it becomes inconvenient to record all cash transactions in the journal. Since all cash transactions are recorded for the first time in the cash book, it is therefore called a book of original entry. Only cash transactions are recorded in the cash book.

PETTY CASH BOOK

It saves the time of chief cashier. Maintenance of petty cash book does not require any specialized knowledge of accounting. It provides control over small payments. It minimizes the chances of fraud.

PURCHASE JOURNAL

It is a subsidiary book which records transactions of credit purchases of goods. Cash purchases are not recorded in the purchases book since they are recorded in the cash book.

SALES JOURNAL

It is a subsidiary book which records transactions of credit sales of goods. Cash sales will be recorded in the cash book and not in the sales book. Sale of assets is not recorded in the sales books.

SALES RETURN JOURNAL

Sales returns journal is a subsidiary book in which seller records all the sales that have been returned to him by his customers. Sales returns journal is also known as returns inwards book and sales returns day book.

VOUCHER SYSTEM

Type of internal system used to control the cash (checks) being spent (written). The voucher system consists of vouchers, voucher files (paid and unpaid), voucher register that takes the place of the purchase journal, cash register that takes the place of the cash disbursement journal, and the general journal.

Unit-3 Depreciation

Depreciation is systematic allocation the cost of a fixed asset over its useful life. It is a way of matching the cost of a fixed asset with the revenue (or other economic benefits) it generates over its useful life. Without depreciation accounting, the entire cost of a fixed asset will be recognized in the year of purchase. Depreciation is the measure of wearing out of a fixed asset. All fixed assets are expected to be less efficient as time goes on.

Depreciation is calculated as the estimate of this measure of wearing out and is charged to the Profit & Loss account either on a monthly or annual basis. The cost of the asset less the total depreciation will give you the Net Book Value of the asset. This will give a misleading view of the profitability of the entity. The observation may be explained by way of an example.

Example

ABC LTD purchased a machine costing $1000 on 1st January 2001. It had a useful life of three years over which it generated annual sales of $800. ABC LTD's annual costs during the three years were $300.

Causes for Depreciation:

The major causes of depreciation are as follows:

1. Wear And Tear

wear and tear refer to a decline in the efficiency of asset due to its constant use. When an asset losses its efficiency, its value goes down and depreciation arises. This is true in case of tangible assets like plant and machinery, building, furniture, tools and equipment used in the factory.

2. Effusion of Time

The value of asset may decrease due to the passage of time even if it is not in use. There are some intangible fixed assets like copyright, patent right, and lease hold premises which decrease its value as time elapse.

3. Exhaustion

An asset may loss its value because of exhaustion too. This is the case with wasting assets such as mines, quarries, oil-wells and forest-stand. On account of continuous extraction, a stage will come where mines and oil-wells get completely exhausted.

Objectives for providing Depreciation:

1. Ascertainment of True Profits

When an asset is purchased, it is nothing more than a payment in advance for for the use of asset. Depreciation is the cost of using a fixed asset. To determine true and correct amount of profit or loss, depreciation must be treated as revenue expenses and debited to profit and loss account.

2. Reporting of True And Fair Financial Position Of A Business

The value of assets decrease over a period of time on account of various factors. In order to present a true state of affairs of the business, the assets should be shown in the balance sheet, at their true and fair values. If the depreciation is not provided then the asset will appear in the balance sheet at the original value. So, in order to show the true financial position of a business, depreciation is required to be charged on the assets.

3.**<u>Replacement of Assets</u>** Assets used in the business need to be replaced after the expiry of their useful life. Depreciation can be taken as a source of fund for replacing worn out asset by a new asset. Thus, depreciation charges help in accumulating funds for the replacement of an asset.

4. **<u>aving In Taxes</u>**
The profit and loss account will show more profits if depreciation is not charged on asset. So, the business needs to pay more income tax to the government. Depreciation charges on assets save the amount of tax equivalent to tax rate. Since it is shown as expense in the profit and loss account, it reduces the amount of the profit.

4<u>. Obsolescence</u> Changes in fashion are external factors which are responsible for throwing out of assets even if those are in good condition. For example black and white televisions have become obsolete with the introduction of color TVs, the users have discarded black and white TVs although they are in good condition. Such as loss on account of new invention or changed fashions is termed as obsolescence.

<u>Methods of Calculating Depreciation:</u>
There are two basic methods of depreciation to choose from when depreciating an asset. These methods include **<u>Straight-line and Written Down Value Method.</u>**.

The **Straight-Line** method is generally the most commonly used method due to its simplicity and consistency of allocating depreciation evenly over the useful life of the asset. To calculate depreciation under this method, the Cost of the Asset is reduced by the salvage or residual value to arrive at the depreciable basis. The resulting depreciable basis is then divided by the estimated useful life.

<u>Straight Line Method (SLM)</u>

This is the simple method of depreciation.
It charges equal amount of depreciation each year over useful life of asset.

It first add up all the costs incurred to bring the asset in use and then it divides that by the useful life of asset in years to calculate the depreciation expense.

E.g.: Say a Computer costs Rs. 30,000 and Rs. 11,000 (as additional set-up/installation/maintenance expenses) = Rs 41,000 and it is anticipated that its scrap value will be Rs. 1,000 at the end of its useful life, of say, 5 yrs.

Total Cost = Cost of Computer + Installation Exp. + Other Direct Costs
Depreciable Amount over No. of years = Total Cost - Salvage Value (At end of useful life)
30,000 +11,000 =41,000 (Total cost)
41,000 – 1,000 = 40,000 as the **Depreciable Amount**

Depreciable Amount = Rs. 40,000, Spread out over 5 years = Rs. 40,000/5(Yrs) = Rs. 8000/-depreciation per annum

Advantages:

- The straight-line method offers simplicity
- Write off the same amount each year and don't have to keep recalculating.
- Easy to determine profits for future years easily; at least as far as how much you will save because of depreciation. In other words, as your profits grow, your depreciation costs stay the same.
- This allows you to make financial forecasts for several years.
- You receive the benefit of depreciation evenly over the life of the asset.

Disadvantages:

- Assets tend to lose value more quickly in their early years.
- Straight-line depreciation does not take this fact into account. If you use assets as collateral for loans, the lender will assume your assets lose more of their value in their first years, so your straight-line method will make the asset look more valuable on your books than it really is for the lender.
- The efficiency of a machine declines in its later years and you will be writing off

the same value in a later year as you did in the first year. The machine actually will have less value for you than your depreciation amount indicates.

Written Down Value Method (WDV)

- This method involves applying the depreciation rate on the Net Book Value (NBV) of asset. In this method, depreciation of the asset is done at a constant rate.
- In this method depreciation charges reduces each successive period.
- This method should be used in those assets, where high depreciation should be charged in initial years.
- Assume the price of a depreciable asset i.e. computer is Rs. 40,000 and its salvage value after 10 years is 0. In this method NBV will never be zero.

Depreciation Per year = (1/N) Previous year's value, Where N= No. of years

So in our example, the depreciation amount during the first year is

[Rs. 40,000*1/10] =Rs. 4,000

NBV of computer after 1st year= Rs 40,000- 4,000 = Rs. 36,000

Depreciation for 2nd year is

[Rs. 36,000*1/10] =Rs. 3,600

Advantages:

1. Since under this method higher depreciation is charged in early years it takes into account that asset is more efficient in early years and therefore it is more realistic way of depreciation.
2. Since in early years machines requires less repairs and as the year passes by repair cost began to rise, therefore this method by charging more depreciation in early years and less in later years make sure that total cost of repairs and depreciation is same every year.

Disadvantages:

1. Since the rate of depreciation is fixed by not following formula chances of subjectivity in fixing rate of depreciation becomes high.

2. The value of asset will never be zero in books of account under this even if asset is of no use to company.

Change in method of charging Depreciation:

CHOICE OF DEPRECIATION METHOD

Depreciation expenses differ from method to method. Choice of selecting a suitable depreciation method is not easy. The decision is based on the inherent characteristic features of an asset. Accelerated depreciation methods may be of much use in case of the following:

Quality of the asset decreases with its age years roll, assets may loose its effective working capacity – Maintenance costs grow.

· Introduction of new equipment due to Research and Development may adversely affect the effective usage of existing equipment.

· The other Straight Line Method may be suitable for assets like buildings, furniture, patents, leases, etc. and for assets which do not warrant frequent repairs and renewals.

· Choice of a method of depreciation affects the amount of net income because quite often the management employs depreciation as an instrument of financial policy of the entity. Hence, selection of a method depends on the management too.

Change in depreciation amount due to change in method is to be given retrospective effect but in all other cases (like Change in Cost, Life, Revaluation etc.) Change in depreciation is given prospective effect

1. The depreciation method selected should be applied consistently from period to period.
2. A change from one method of providing depreciation to another should be made only if the adoption of the new method is required by statute or for compliance with an accounting standard or if it is considered that the change would result in a more appropriate preparation or presentation of the financial statements of the enterprise.
3. When a change in the method of depreciation is made, depreciation should be recalculated in accordance with the new method from the date of the asset coming into use. The deficiency or surplus arising from the retrospective recomputation of depreciation in accordance with the new method would be adjusted in the accounts in the year in which the method of depreciation is changed.

4. In case the change in the method results in deficiency in depreciation in respect of past years, the deficiency should be charged to the profit and loss account. In case the change in the method results in surplus, it is recommended that the surplus be initially transferred to the 'Appropriations' part of the profit and loss account and thence to General Reserve through the same part of the profit and loss account. Such a change should be treated as a change in accounting policy and its effects should be quantified and disclosed.

Salient Features of Accounting Standards(AS-6)(ICAI)Revised:

The Institute of Chartered Accountants of India, keeping in view with international accounting principles, revised (AS)–6.This standard AS–6 deals with the concept: Depreciation

Depreciation is defined as "a measure of the wearing out, consumption or other loss of value of a depreciable asset arising from use, effusion of time or obsolescence through technology and market changes. Depreciation is allocated so as to charge a fair proportion of the depreciable amount in each accounting period during the Expected Useful Life of the Asset. Depreciation includes amortization of assets whose useful life is predetermined".

Salient features of AS–6 (Revised) Accounting for Depreciation:

i. Existing Assets: The depreciable amount of existing assets = Cost of the asset (historical not market value) – salvage (scrap value) value.

ii. Revision of estimate useful life of an asset: In case, if there is a necessity to revise the estimated life of an asset, the unamortized depreciable amount will have to be charged over the remaining useful life.

iii. Addition (or) extension to an existing asset of capital nature: In such cases, two factors will have to be considered:

    a) such an addition should retain separate identity,
    b) It can still be used after the disposal of existing assets.

Then, depreciation is to be determined independently on the basis of an estimate of its own useful life. In other cases, the depreciation has to be determined on the basis of remaining useful life of the existing asset plus addition or extension as an integral part.

INVENTORIES:

The raw materials, work-in-process goods and completely finished goods that are considered to be the portion of a business's assets that is ready or will be ready for sale. Inventory represents one of the most important assets that most businesses possess, because the turnover of inventory represents one of the primary sources of revenue generation and subsequent earnings for the company's shareholders/owners.

Inventories are assets:

(a) held for sale in the ordinary course of business; (b) In the process of production for such sale; or

(c) In the form of materials or supplies to be consumed in the production process or in the rendering of services.

Inventories encompass goods purchased and held for resale, for example, merchandise purchased by a retailer and held for resale, computer software held for resale, or land and other property held for resale. Inventories also encompass finished goods produced, or work in progress being produced, by the enterprise and include materials, maintenance supplies, consumables and loose tools awaiting use in the production process. Inventories do notinclude machinery spares which can be used only in connection with an item of fixed asset and whose use is expected to be irregular; such machinery spares are accounted for in accordance with Accounting Standard (AS) 10, Accounting for Fixed Assets.

Valuation of Inventories:

Inventory is valued on the basis of the following factors:

1. **Measurement**

Inventories should be valued at the lower of cost and net realizable value.

The cost of inventories should comprise all costs of purchase, costs of conversion and other costs incurred in bringing the inventories to their present location and condition.

2. **Costs of Purchase**

The costs of purchase consist of the purchase price including duties and taxes (other than those subsequently recoverable by the enterprise from the taxing authorities), freight inwards and other expenditure directly attributable to the acquisition. Trade discounts, rebates, duty drawbacks and other similar items are deducted in determining the costs of purchase.

3. **Costs of Conversion**

The costs of conversion of inventories include costs directly related to the units of production, such as direct labour. They also include a systematic allocation of fixed and variable production overheads that are incurred in converting materials into finished goods. Fixed production overheads are those indirect costs of production that remain relatively constant regardless of the volume of production, such as depreciation and maintenance of factory buildings and the cost of factory management and administration. Variable production overheads are those indirect costs of production that vary directly, or nearly directly, with the volume of production, such as indirect materials.

4. **Other Costs**

Other costs are included in the cost of inventories only to the extent that they are incurred in bringing the inventories to their present location and condition. For example, it may be appropriate to include overheads other than production overheads or the costs of designing products for specific customers in the cost of inventories. Interest and other borrowing costs are usually considered as not relating to bringing the inventories to their present location and condition and are, therefore, usually not included in the cost of inventories.

Method of valuation of Inventories:

There are three basis approaches to valuing inventory that are allowed by GAAP - (a**) First-in, First-out (FIFO):** Under FIFO, the cost of goods sold is based upon the cost of material bought earliest in the period, while the cost of inventory is based upon the cost of material bought later in the year. This results in inventory being valued close to current replacement cost. During periods of inflation, the use of FIFO will result in the lowest estimate of cost of goods sold among the three approaches, and the highest net income.

(b) **<u>Last-in, First-out (LIFO)</u>:** Under LIFO, the cost of goods sold is based upon the cost of material bought towards the end of the period, resulting in costs that closely approximate current costs. The inventory, however, is valued on the basis of the cost of materials bought earlier in the year. During periods of inflation, the use of LIFO will result in the highest estimate of cost of goods sold among the three approaches, and the lowest net income.

(c) **<u>Weighted Average</u>:** Under the weighted average approach, both inventory and the cost of goods sold are based upon the average cost of all units bought during the period. When inventory turns over rapidly this approach will more closely resemble FIFO than LIFO.

Firms often adopt the LIFO approach for the tax benefits during periods of high inflation, and studies indicate that firms with the following characteristics are more likely to adopt LIFO - rising prices for raw materials and labor, more variable inventory growth, an absence of other tax loss carry forwards, and large size. When firms switch from FIFO to LIFO in valuing inventory, there is likely to be a drop in net income and a concurrent increase in cash flows (because of the tax savings). The reverse will apply when firms switch from LIFO to FIFO.

Given the income and cash flow effects of inventory valuation methods, it is often difficult to compare firms that use different methods. There is, however, one way of adjusting for these differences. Firms that choose to use the LIFO approach to value inventories have to specify in a footnote the difference in inventory valuation between FIFO and LIFO, and this difference is termed the LIFO reserve. This can be used to adjust the beginning and ending inventories, and consequently the cost of goods sold, and to restate income based upon FIFO valuation.

<u>Periodic Inventory System:</u>

Periodic inventory is a system of <u>inventory</u> in which updates are made on a periodic basis. This differs from <u>perpetual inventory</u> systems, where updates are made as seen fit.In a periodic inventory system no effort is made to keep up-to-date records of either the inventory or the cost of goods sold. Instead, these amounts are determined only periodically - usually at the end of each year. This physical count determines the amount of inventory appearing in the balance sheet. The cost of goods sold for the entire year then is determined by a short computation.

<u>Perpetual Inventory System:</u>

Under perpetual inventory system, inventory and cost of goods sold are updated for each sale/purchase and return transaction. We have already discussed the basic concept of perpetual inventory system in the comparison of perpetual-periodic inventory.

The perpetual inventory system is intended as an aid to material control. It is a system of stock control followed by stores department. The system follows a method of recording stores by which information about each receipt, issue and current balance of stock is always available.

Perpetual inventory system may be defined as a method of recording stores balances after every receipt and issue to facilitate regular checking and to obviate closing down for stock taking." So perpetual inventory system implies continuous maintenance of stock records and in its broad sense it covers both continuous stock taking as well as up to date recording of stores books. The balance of the same item of store in bin card should correspond with that shown in the materials or store ledger card and a frequent checking of these two records should be made and compared with the actual or physical quantity of materials in stock.

Final Accounts:

The financial statements of an organization made up at the end of an accounting period, usually the fiscal year. For a manufacturer, the final accounts consist of (1) manufacturing account, (2) trading account, (3) profit and loss account, and (4) profit and loss appropriation account. A commercial company's final accounts will include all of the above except the manufacturing account. Together, these accounts show the gross profit, net income, and distribution of net income figures of the company.

Form of Final Accounts: There is a standard format of final accounts only in the case of a limited company. There is no fixed prescribed format of financial accounts in the case of a proprietary concern and partnership firm.

- Transactions
- Journal
- Ledger
- Trial Balance
- Trading & Profit & Loss Account
- Balance Sheet

MEANING

The Trading and Profit & loss account and Balance Sheet prepared at the end of a year is known as **Final accounts**. While preparing the final accounts, there may be some items so far not adjusted. These items are to be adjusted in the final accounts for calculating the correct profit or loss of the business. The usual adjustments in the final accounts are:-

a. Expenses owing: - These are the expenses incurred during the year but not paid in cash. This amount will be paid in the near future (next year). The owing expense is to be added with the amount of same expense already paid given in the trial balance and it should be shown in the balance sheet as a current liability.

The double entry for recording the expenses owing is

Debit Expenses account

Credit Expenses owing account

This expense is also known as outstanding expenses, expenses payable or expense payable.

b. Prepaid expense. :- This is the expense paid during the year for the benefit of the next year. The portion of the expense which is prepaid is to be deducted from the total expenses already paid during the year (given in the trial balance) and shown as current asset in the balance sheet.

The double entry for recording the prepaid expense is

Debit Prepaid expense account and

Credit Expense account

This expense is also known as expense paid in advance or unexpired expense

c. Accrued income: - The income earned during the year but not received in cash is known as accrued income. The amount of accrued income is to be considered as current year's income and added

With the concerned income received during the year (given in the trial balance) and shown as a current asset in the balance sheet.

The double entry for recording the accrued income is:

Debit Accrued income account and

Credit Income account

The accrued income is also known as outstanding income.

d. Income received in advance: - This is the income received during the year for the services to be rendered during the next year. Since this income is not related to the current year, it should be deducted from the concerned income (given in the trial balance) and shown as a current liability in the balance sheet.

The double entry for recording the income received in advance is:

Debit   Income account and

Credit   Income received in advance

This is also known as unexpired income.

e. Depreciation: - The part of the cost of a fixed asset that is consumed by a business during the period of its use is known as depreciation. It is considered as an expense in the business therefore shown as an expense in the profit & loss account and deducted from the cost price of the concerned fixed asset in the balance sheet.

The double entry for recording depreciation is:
Debit   Profit & loss account and
Credit   Depreciation account

f. Bad debt: - The part of the amount of debtors which cannot be recovered is known as bad debt. It is an expense to be shown in the profit & loss account. If the bad debt appears in the trial balance, it is known as bad debt written off and shown in the profit & loss account only. If bad debt information appears among the adjustment points below the trial balance, then it should be shown as an expense in the profit & loss account and shown as a deduction from the debtors in the balance sheet under the heading “current assets”.

The double entry for recording the bad debt is:

Debit   Bad debt account and

Credit Debtors account

g. Goods drawings by the owner for his personal use:-

The amount of goods withdrawn by the owner for his personal use is to be considered as drawing. The double entry for recording the goods drawings is:

Debit Drawings account and

Credit Purchase account or sales account

The amount of goods drawings should be deducted from purchases and capital in the Balance Sheet.

From the following Trial Balance and additional information, you are required to prepare profit and loss account and Balance Sheet.

TRIAL BALANCE as on 31st March 2012

Particulars	**Debit**	**Credit**
Capital		**2,90,000**
Sundry Debtors	**65,000**	
Drawings	**7,600**	
Building	**2,20,000**	
Sundry Creditors		**12,000**
Wages	**8,000**	
Purchases	**89,000**	
Opening Stock	**12,000**	
Cash in Hand	**1,900**	
Cash at Bank	**12,000**	
Carriage Charges	**20,000**	
Salaries	**8,000**	
Rent, Taxes & Insurance	**1300**	
Sales		**1,50,000**
Purchase Returns		**4,500**
Sales Returns	**2,800**	
Bills Receivable	**15,000**	
Bils Payable		**7,000**
Interest		**3,500**
Advertisement	**2,400**	
Trade Expenses	**2,000**	
	4,67,000	**4,67,000**

Additional Information:

i. Closing Stock as on 31st March2012 was Rs.15,000.

ii. Prepaid Insurance Rs.400

iii. Outstanding Salaries Rs.2,000;Outstanding Rent & Taxes Rs.1,300 iv. Depreciation charged on Building @2%p.a.

Solution:

Dr. Trading and Profit & Loss Account as on 31st March2012 Cr.

Particulars	**Amount**	Particulars	**Amount**
To Opening Stock	12,000	By Closing Stock	15,000
To Purchases 89,000 (-)Returns 4,500	84,500	By Sales 1,50,000 (-)Returns 2,800	147200
To Wages	8,000		
To Carriage	20,000		
To Gross profit c/d	37,700		
	1,62,200		**1,62,200**
To Trade Expenses	2,000	By Gross Profit b/d	**37,700**
To Advertisement	2,400	By Interest	**3,500**
To Salaries 8,000	10,000		
(+)Outstanding 2,000			
To Rent,Taxes & Insurance 1,300 (-)Prepaid 400	2,200		
To Depreciation on Building	4,400		
To Net Profit transferred to Capital A/C	20,200		
	41,200		**41,200**

Balance-Sheet as on 31st March2012

Liabilities	**Amount**	**Assets**	**Amount**
Capital 2,90,000 (+)Net Profit 20,200 (-)Drawings 7,600	3,02,600	Building 2,20,000 (-)Depreciation 4,400	2,15,600
Creditors	12,000	Cash at Bank	12,000
Bills Payable	7,000	Cash in Hand	1,900
Outstanding Rent & Taxes	1,300	Debtors	65,000

Outstanding Salaries	2,000	Bills Receivable	15,000
		Prepaid Insurance	400
		Closing Stock	15,000
	3,24,900		**3,24,900**

PREPARATION OF FINAL ACCOUNTS OF NON-PROFIT ORGANIZATION

By preparing Receipts & Payments Account, Income and Expenditure Account and a Balance sheet. Being a "non-profit" organization does not mean it doesn't make any profit, but rather that the profits are not distributed to investors as dividends. Many non-profit organizations make billions of dollars of profits per year, and often donate these corporate profits to charitable causes.

BANK RECONCILIATION STATEMENT

A **Bank reconciliation** is a process that explains the difference between the bank balance shown in an organisation's bank statement, as supplied by the bank, and the corresponding amount shown in the organization's own accounting records at a particular point in time.

Such differences may occur, for example, because a cheque or a list of cheques issued by the organization has not been presented to the bank, a banking transaction, such as a credit received, or a charge made by the bank, has not yet been recorded in the organisation's books, or either the bank or the organization itself has made an error.

It may be easy to reconcile the difference by looking at very recent transactions in either the bank statement or the organisation's own accounting records (cash book) and seeing if some combination of them tallies with the difference to be explained. Otherwise it may be necessary to go through and match every single transaction in both sets of records since the last reconciliation, and see what transactions remain unmatched. The necessary adjustments should then be made in the cash book, or any timing differences recorded to assist with future reconciliations.

A bank reconciliation statement is prepared to reconcile the two balances of Cash Book and Pass Book. So, when you will prepare a bank reconciliation statement you will start it with one balance make adjustments and then you will reach to the other balance. This way both the balances will agree.

Unit-IV

CONSIGNMENT:

An arrangement whereby goods are left in the possession of another party to sell. Typically, the consignor receives a percentage of the sale (sometimes a very large percentage). Consignment deals are made on a variety of products - from artwork, to clothing, to books. In recent years, consignment shops have become rather trendy, especially those offering specialty products, infant wear and high-end fashion items.

It is also defined as a quantity of goods that are sent to a person or place to be sold the act or process of sending goods to a person or place to be sold.

Features of consignment are:

- The relation between the two parties is that of consignor and consignee and not that of buyer and seller
- The consignor is entitled to receive all the expenses in connection with consignment
- The consignee is not responsible for damage of goods during transport or any other procedure
- Goods are sold at the risk of consignor. The profit or loss belongs to consignor only

A consignor who consigns goods to a consignee transfers possession but not ownership of the goods to the consignee. The consignor retains title to the goods. The consignee takes possession of the goods subject to a trust. If the consignee converts the goods to a use not contemplated in the consignment agreement, for example selling them and keeping the proceeds of the sale for himself, then the consignee commits the crime of embezzlement.

Accounting Entries in the Books of Consignor:

(1)**On dispatch of goods:-**

Consignment account

To Goods sent on consignment account (With the cost of goods)

---

(2)**On payment of expenses on dispatch:-**

Consignment account

To Bank account (With the amount spent as expenses)

---

(3) **On receiving advance:**

Cash or bills receivable account

To Consignee's personal account (With the amount cash or bill)

---

(4)**On the consignee reporting sale (as per A/S):-**

Consignee's personal account

To Consignment account (With gross proceeds of sales)

---

(5) **For expenses incurred by the consignee (as per A/S):-**

Consignment account

To Consignee's personal account (With the amount of expenses)

---

(6) **For commission payable to the consignee:-**

Consignment account

To Consignee's personal account (With the amount of expenses)

<u>Difference between Consignment and Joint Venture</u>

The main differences between joint venture and consignment are as under:

1. Nature

Joint venture: It is a temporary partnership business without a firm name.

Consignment: It is an extension of business by principal through agent.

2. Parties

Joint venture: The parties involving in joint venture are known as co-ventures.

Consignment: Consignor and consignee are involving parties in the consignment.

3. Relation

Joint venture: The relation between co-ventures is just like the partners in partnership firm.

Consignment: The relation between the consignor and consignee is 'principal and agent'.

4. Sharing Profit

Joint venture: The profits ans losses of joint venture are shared among the co-ventures in their agreed proportion.

Consignment: The profits and losses are not shared between the consignor and consignee. Consignee gets only the commission.

5. Rights

Joint venture: The co-ventures in a joint venture have equal rights.

Consignment: In consignment, the consignor enjoys principal's right whereas consignee enjoys the right of agent.

6. Exchange Of Information

Joint venture: The co-ventures exchange the required information among them regularly. **Consignment:** The consignee prepares an account sale which contains a details of business activities carried on and is being sent to the consignor.

7. Ownership

Joint Venture: All the co-ventures are the owners of the joint venture.

Consignment: The consignor is the owner of the business.

8. Method Of Maintaining Accounts

Joint venture: There are different methods of maintaining accounts in joint venture.As per agreement the co-ventures maintain their account.

Consignment: In consignment, there is only one method of maintaining account.

9. Basis of Account

Joint venture: Cash basis of accounting is applicable in joint venture.

Consignment: Actual basis is adopted in consignment.

10. Continuity

Joint venture: As soon as the particular venture is completed, the joint venture is terminated. **Consignment:** The continuity of business exists according to the willingness of both consignor and consignee.

Loss of Goods on Consignment

The goods are consigned from one place to another. After receiving the goods by consignee, the goods are stored by the consignee before selling them to customers. It is natural that some loss to the goods may take place within that period. The goods may be lost, destroyed or damaged either in transit or in consignee's store. Such loss can be divided into two parts.

1. Normal Loss

The loss which is caused by unavoidable reasons is known as normal loss. For examples shrinkage, evaporation, leakage and pilferage. Such losses form part of cost of goods and no additional adjustment is required for this purpose. The normal loss is borne by goods units. The quantity of such loss is to be deducted from the total quantity sent by the consignor. The following formula may be used for the valuation of unsold stock.

Value of closing stock= (Total value of goods sent/Net quantity received by consignee) X unsold quantity

Net quantity received = Goods consigned quantity - Normal loss quantity.

2. Abnormal Loss

The loss which could be avoided by proper planning and care are abnormal loss. They are like theft, riots, accidents, fire, earthquake etc. These losses could occur in transit or in consignee's store and solely to be borne by consignor.

The abnormal loss should be adjusted before ascertaining the result of the consignment. The valuation of abnormal loss is done on the same basis as the unsold stock is valued. The journal entries for abnormal loss in different cases are as under:

If goods are not insured

For recording abnormal loss:

Abnormal loss A/C ....Dr.

To consignment A/C

For abnormal loss transferred:

Profit and loss A/C........Dr.

To abnormal loss A/C

If goods are insured and claim admitted in full

Bank/Consignee's/Insurance company A/C...........Dr.

To consignment A/C

If goods are insured and claim admitted in partial

Profit and loss A/C...........Dr. (Net loss amount)

Insurance Co./bank/consignee's A/c.......Dr. (Claim admitted)

To consignment A/c (total loss amount)

The following method should be followed while valuing abnormal loss:

A) Goods sent on consignment (at cost price)…............$ XXX B) Add: Non-recurring expenses:

Consignor's expenses…...$ XXX

Consignee's expenses…...$ XXX

Total cost before abnormal loss A+B…..........................$ XXX

Value of abnormal loss = (Total cost/Total units consigned) X abnormal loss units.

Question: A& Co. of Kolkata sent on consignment account goods to B& Co. of Mumbai at an invoice price of Rs.29675 and paid for freight Rs.762,Cartage rs.232 and insurance rs.700.Half the goods were sold by agents for Rs.17,500,subject to the agent's commission of Rs.875,storage expenses of Rs.200 and other selling expenses of Rs.350.One-fourth of the consignment was lost by fire and a claim of Rs.5000 was recovered. Draw up the necessary accounts in the books of A & Co. and ascertain the profit or loss made on the consignment. The consignor received a two months bill of exchange from the agents in satisfaction of the dues.

Solution:

Consignment of Mumbai Account

Particulars	Amount	Particulars	Amount

To Goods sent on consignment	29,675	By B& Co.(Sale proceeds)	17500
To Cash: Freight 762 Cartage 232 Insurance 700	1694	By Abnormal Loss A/C	7843
To B & Co. Commission 875	1425	By Consignment Stock A/C	7843
Other Selling [illegible] 350			
To Net Profit (transferred to P&L A/C)	392		
	33186		**33186**

B & Co. Mumbai

Particulars	**Amount**	**Particulars**	**Amount**
To Consignment to Mumbai (sale proceeds**)**	17500	By Consignment to Mumbai(expense and commission)	1425
		By Bills Receiveble A/C	16075
	17500		**17500**

Abnormal Loss Account

Particulars	**Amount**	**Particulars**	**Amount**
To Consignment A/C	7843	By Bank(received from Insurance Co.)	5000
		By P&L A/C	2843
	7843		**7843**

Working Notes:

Calculation of Abnormal Loss

Particulars	Rupees
¼ of invoice Price of goods	**7419(approx)**
Add :¼ of Freight, Cartage and Insurance	**424**
Total Abnormal Loss	**7843**
Less: Recovered from insurance Co.	**5000**
Net Abnormal Loss	**2843**

Joint Venture:

An association of two or more individuals or companies engaged in a solitary business enterprise for profit without actual partnership or incorporation; also called a joint adventure.

A joint venture is a contractual business undertaking between two or more parties. It is similar to a business partnership, with one key difference: a partnership generally involves an ongoing, long-term business relationship, whereas a joint venture is based on a single business transaction. Individuals or companies choose to enter joint ventures in order to share strengths, minimize risks, and increase competitive advantages in the marketplace. Joint ventures can be distinct business units (a new business entity may be created for the joint venture) or collaborations between businesses. In a collaboration, for example, a high-technology firm may contract with a manufacturer to bring its idea for a product to market; the former provides the know-how, the latter the means.

Difference between a Joint venture and Partnership

Partnership

A partnership is a legal arrangement where two or more people own a business together. This means that the entire business is shared for as long as the business exists. Both partners contribute money, time and expertise to making a profitable enterprise, and that enterprise lasts until the partnership is dissolved.

Joint Venture

You enter a joint venture for a specific project. There is a time limit on joint ventures, and they have clearly stated limits on their purposes. You might enter a joint venture in order to make a product that neither partner can afford to make on her own. An example is developing new software. You do not give up half of your business in a joint venture; you share the profits and expenses for a particular venture.

<u>Difference between Consignment and Joint Venture</u>

The main differences between joint venture and consignment are as under:

1. Nature

Joint venture: It is a temporary partnership business without a firm name.

Consignment: It is an extension of business by principal through agent.

2. Parties

Joint venture: The parties involving in joint venture are known as co-ventures.

Consignment: Consignor and consignee are involving parties in the consignment.

3. Relation

Joint venture: The relation between co-ventures is just like the partners in partnership firm.

Consignment: The relation between the consignor and consignee is 'principal and agent'.

4. Sharing Profit

Joint venture: The profits ans losses of joint venture are shared among the co-ventures in their agreed proportion.

Consignment: The profits and losses are not shared between the consignor and consignee. Consignee gets only the commission.

5. Rights

Joint venture: The co-ventures in a joint venture have equal rights.

Consignment: In consignment, the consignor enjoys principal's right whereas consignee enjoys the right of agent.

6. Exchange Of Information

Joint venture: The co-ventures exchange the required information among them regularly.

Consignment: The consignee prepares an account sale which contains a details of business activities carried on and is being sent to the consignor.

7. Ownership

Joint Venture: All the co-ventures are the owners of the joint venture.

Consignment: The consignor is the owner of the business.

8. Method Of Maintaining Accounts

Joint venture: There are different methods of maintaining accounts in joint venture.As per agreement the co-ventures maintain their account.

Consignment: In consignment, there is only one method of maintaining account.

9. Basis of Account

Joint venture: Cash basis of accounting is applicable in joint venture.

Consignment: Actual basis is adopted in consignment.

10. Continuity

Joint venture: As soon as the particular venture is completed, the joint venture is terminated.

Consignment: The continuity of business exists according to the willingness of both consignor and consignee.

FEATURES OF A JOINT VENTURE

The main features of a joint venture are specifically made clear. · Two or more person are needed.

- It is an agreement to execute a particular venture or a project. · The joint venture business may not have a specific name.
- It is of temporary nature. So the agreement regarding the venture automatically stands terminated as soon as the venture is complete.

· The co-ventures share profit and loss in an agreed ratio. The profits and losses are to be shared equally if not agreed otherwise.

- The co-ventures are free to continue with their own business unless agreed otherwise during the life of joint venture.

<u>**Accounting Entries in a Joint Venture**</u>

(1)When venture contributes cash to joint funds	Joint Bank A/C	Venturer's A/C(individual
(2) When amount is spent on account of expenses, for purchasing goods for the	Joint venture A/C	<u>Joint Bank A/C</u>
(3)If any expenses are paid by the	Joint Venture A/C	Venturer's A/C
(4)For Sales: i. Cash	Joint Bank A/C Sundry Debtors A/C	Joint Venture A/C Joint Venture A/C
(7) Balance of the Joint venture A/C will be either profit or Loss.	Joint Venture A/C(if profit)	Venturer's A/C
(8)Joint Bank Account and personal account of the Venturer's A/C will be automatically closed by the introduction		

<u>Hire Purchase:</u>

<u>Hire Purchase</u> is defined asa system for purchasing merchandise, such as cars or furniture, in which the buyer takes possession of the merchandise on payment of a deposit and completes the purchase by paying a series of regular installments while the seller retains ownership until the final installment is paid

A method of buying goods through making installment payments over time. The term hire purchase originated in the U.K., and is similar to what are called "rent-to-own" arrangements in the United States. Under a hire purchase contract, the buyer is leasing the goods and does not obtain ownership until the full amount of the contract is paid.

Parties in a Hire-Purchase System:

1. **Hirer-**who buy the goods at any time by giving notice to the owner and paying the balance of the HP price less a rebate (each jurisdiction has a different formula for calculating the amount of this rebate).He also to pay the hire installments and to take reasonable care of the goods (if the hirer damages the goods by using them in a non-standard way, he or she must continue to pay the installments and, if appropriate, recompense the owner for any loss in asset value).

2. **Seller-** a person who has the resources and the legal right to sell the goods on credit (which usually depends on a licensing system in most countries), the seller and the owner will be the same person. But most sellers prefer to receive a cash payment immediately He also has the right to terminate the agreement where the hirer defaults in paying the installments or breaches any of the other terms in the agreement.

Hire Purchase agreements must be in writing and signed by both [parties].They must clearly lay out the following information in a print that all can read without effort:

1. A clear description of the goods
2. The cash price for the goods
3. The Hire Purchase price, i.e., the total sum that must be paid to hire and then purchase the goods
4. The deposit
5. The monthly installments (most states require that the applicable interest rate is disclosed and regulate the rates and charges that can be applied in HP transactions) and
6. A reasonably comprehensive statement of the parties' rights (sometimes including the right to cancel the agreement during a "cooling-off" period).
7. The right of the hirer to terminate the contract when he feels like doing so with a valid reason.

Characteristics of Hire-Purchase System

The characteristics of hire-purchase system are as under

- Hire-purchase is a credit purchase.
- The price under hire-purchase system is paid in installments.
- The goods are delivered in the possession of the purchaser at the time of commencement of the agreement.
- Hire vendor continues to be the owner of the goods till the payment of last installment.
-

The hire-purchaser has a right to use the goods as a bailer.

- The hire-purchaser has a right to terminate the agreement at any time in the capacity of a hirer.
- The hire-purchaser becomes the owner of the goods after the payment of all installments as per the agreement.
- If there is a default in the payment of any installment, the hire vendor will take away the goods from the possession of the purchaser without refunding him any amount.

  Thus," **hire- purchase agreement**" means an agreement under which goods are let on hire and under which the hirer has an option to purchase them in accordance with the terms of the agreement and includes an agreement under which-
- possession of goods is delivered by the owner thereof to a person on condition that such person pays the agreed amount in periodical installments, and
- the property in the goods is to pass to such person on the payment of the last of such installments, and (iii) such person has a right to terminate the agreement at any time before the property so passes;
- " hirer" means the person who obtains or has obtained possession of goods from an owner under a hire- purchase agreement, and includes a person to whom the hirer' s rights or liabilities under the agreement have passed by assignment or by operation of law;
- “owner" means the person who lets or has let, delivers or has delivered possession of goods, to a hirer under a hire- purchase agreement and includes a person to whom the owner' s property in the goods or any of the owner' s rights or liabilities under the

agreement has passed by assignment or by operation of law.

Difference between Hire-Purchase System and Installment System

Installment Payment System is system of purchase and sale of goods in which title of goods is immediately transferred to the purchaser at the time of sale of goods and the sale price of the goods is paid in installments. In the event of default in payment of any installment, the seller has no right to take back goods from the possession of the purchaser. He can file a suit for the recovery of the outstanding balance of the price of goods sold.

The followings are the differences between Hire-purchase system and Installment payment system:

- In **Hire-purchase system**, the transfer of ownership takes place after the payment of all installments while in case of **Installment payment system**, the ownership is transferred immediately at the time of agreement.
- In **Hire-purchase system**, the hire-purchase agreement is like a contract of hire though later on it may become a purchase after the payment of last installment while in **Installment payment system**, the agreement is like a contract of credit purchase.
- In case of default in payment, in **Hire-purchase system** the vendor has a right to back goods from the possession of the hire-purchaser while in case of **Installment payment system**, the vendor has no right to take back the goods from the possession of the purchaser; he can simply sue for the balance due.
- In **Hire-purchase system**, if the purchaser sells the goods to a third party before the payment of last installment, the third party does not get a better title on the goods purchased. But in case of **Installment payment system**, the third party gets a better title on the goods purchased.
- In **Hire-purchase system** the provisions of the Hire-purchase Act apply to the transaction while in case of **Installment payment system**, the provisions of Sale of Goods Act apply to the transaction.

Accounting Entries in the Books of Hire purchaser

Journal Entries in the Books Of Hire Purchaser

There are two methods of recording hire purchase transactions in the books of the hire purchaser:

i. When the asset is recorded in full cash price-full cash price method

ii. When the asset is recorded at cash price actually paid in each installment-: Actual cash price method.

1. **For the purchase of asset**:

First Method

Asset A/C (full cash price)...........Dr.

To vendor A/C

Second Method

No entry

2. **For the payment made for 'down payment'**

First Method

Vendor A/C.............Dr.

To bank A/C

Second Method

Asset A/C............Dr.

To Bank A/C

3**. For installment due**

First Method

Interest A/C............Dr.

To vendor A/C

Second Method

Asset A/C (part of cash value)............Dr.

To Interest A/C

4. **For the payment of installment (both method)**

Vendor A/C............Dr.

To Bank A/C

5. **For charging depreciation(on the basis of cash value) (both methods)**

Depreciation A/C...................Dr.

To Asset A/C

6. **For transfer of interest and depreciation (both methods)**

Profit and loss A/C............Dr.

To depreciation A/C

To interest A/C

Journal Entries In The Books Of Vendor

1. **For selling goods on hire purchase**

Hire purchase A/C...........Dr. (full cash price)

To sales/hire purchase sales A/C

2. **For receiving down payment**

Cash/bank A/C.................Dr.

To hire purchaser A/C

3. **For installment due**

Hire purchaser A/C............Dr.

To Interest A/C

4. **For receiving the installment**

Cash/bank A/C ....Dr.

To hire purchaser A/C

5. **For transferring interest**

Interest A/C............Dr.

To profit and loss A/C

Posting in Ledger Accounts: After passing journal entries under any of the methods discussed above, the following ledger accounts are opened in the ledger and the postings are made accordingly.

(i) Asset A/c. (e.g. Trucks A/c, Machinery A/c. etc.)

(ii) Vendor's A/c.

(iii) Interest A/c.

(iv) Depreciation A/c.

Note: Before recording the entries the amounts of interest and depreciation will be calculated in two separate tables showing the calculations of interest and depreciation.

Calculation of Interest

The total payment made under hire-purchase system is more than cash price. In fact, this excess of payment over the cash price is interest. It is very essential to calculate interest because the amount paid for interest is charged to revenue and the asset is capitalized at cash price. Thus normally all installments will include a part of cash price and a part of interest on the outstanding balance. However the amount paid at the time of agreement (down payment) will not include any interest. The calculation of interest is made under two conditions:

(a) When interest is included in amount of installment: Where the hire-purchase price i.e. payment made in the form of down payment and all installments is more than the cash price, it is regarded that the interest is included in installments.

Illustration: On Ist April, 2005 Mr. X purchased from M/s Y & Co. one 'Motor Truck' under hire-purchase system, Rs. 5,000 being paid on delivery and the balance in five annual installments of Rs. 7,500 each payable on 31st March each year. The cash price of the motor truck is Rs. 37,500 and vendors charge interest at the rate of 5 per cent per annum on yearly balances. Find out the amounts of principal and interest included in each installment.

(b) When interest is not included in installments: Where the total amount paid in the form of down payment and all installments is exactly equal to the cash price, it is regarded that the interest is not included in installments. It means that interest is payable in addition to the agreed amount of installment.

Question: On April 1,2005, A Transport Company purchased a Motor Lorry from Motor Supply Co. Ltd. on hire-purchase basis, the cash price being Rs. 60,000. Rs. 15,000 on signing of the contract and balance in three annual installments of Rs. 15,000 each on 31st

March every year. In addition to it, interest at 5 per cent per annum was also payable to vendors on outstanding balances.

Branch Accounting

An accounting system in which separate accounts are maintained for each branch of a corporate entity or organization. The primary objectives of branch accounting are better accountability and control, since profitability and efficiency can be closely tracked at the branch level.

Branch accounting may involve added expenses for an organization in terms of accounting and infrastructure. This is because it may be necessary to appoint branch accountants to ensure accurate financial reporting and compliance with head office procedures and processes.

Types of Branches

The branches opened in the different parts of the nation, where the original undertaking being registered are called inland branches. These types of branches are also called home branches or national branches. There are two types of inland branches, which are:

a) Dependent branch

b) Independent branch

a) Dependent Branch:

Dependent branches are the branches that do not keep their records but all the records are maintained by head office. They are not authorized to act solely without the prior permission of the head office. All the plans, policies, rules and regulations of these branches are totally formulated and executed by the head office. In other words, all the functions of dependent branch are totally controlled by head office.

b). Independent Branch:

The branches that can keep their accounts themselves and sell goods that are sent by the head office as well as those purchased by themselves are known as independent branches. These are the branches which can sell the goods to head office too. They can pay their own expenses and can deposit their collection in their own name in the bank. These branches record separately and independently all the transactions which are even recorded by the head office.

Systems of Accounting:

Stock and Debtors system is generally used when the goods are sent to the branch at pro-forma invoice price and the size of the branch is large. Under this system, the branch maintains a few central accounts to exercise greater control over the branch stock and other related expenses. These accounts are as follows:

1.Branch Stock Account
2.Branch Debtors Account
3.Branch Expenses Account
4.Branch Adjustment Account
5.Goods Sent to Branch Account
6.Branch Stock Reserve Account

Branch Stock Account

This account is on the pattern of a stock account. The account helps the Head Office in maintaining an effective control over the Branch Stock and tells about shortage and surplus in the branch stock because of the difference between the pro-forma invoice price and the selling price.

Unlike traditional accounting practice, branch stock a/c is always maintained on the selling price or pro-forma invoice price. Selling price is used to record the goods sold by the branch to its customer and goods returned by the branch customers.

Branch Debtors Account

Branch debtors' a/c is maintained in the traditional manner to record transactions in between branch and its credit customers

Branch Expense Account

The purpose of maintaining this account is nothing but the compile all branch expenses at one place. This will include all types of expenses i.e. cash based expenses and receivables based expenses

Branch Adjustment Account

Branch adjustment a/c replaces the branch income statement (profit & loss a/c). This is the account in which all expenses and losses are closed along with the margin that is a difference between cost and the selling price. This difference is split into two; one is termed as "surplus" that comes from the branch stock a/c representing the difference between selling price and pro-forma invoice price, the second is termed as "loading" that represents the difference between pro-forma invoice price and cost. This loading is calculated on opening and closing stock balances and also on the net of the goods sent branch.

Goods Sent to Branch Account

This is a supporting account, which is maintained to show second effects of the goods sent to branch and the goods returned from branch at pro-forma invoice price. Although the goods sent to and returned form the branch should be adjusted in the purchases a/c of the head office, but as we know that the branch stock a/c is not maintained at cost price, therefore, second effect of goods sent to and returned from branch is not recorded directly into the purchases a/c instead this second effect is recorded into the goods sent to branch a/c which after adjustment of the loading is finally closed into the purchases a/c.

Branch Stock Reserve Account

This is contra to branch stock account. In this account opening and closing balance of loading on branch stock is maintained.

Accounting Entries in books of head office under Debtors system of Branch Accounting

1) For Goods sent by Head Office to the Branch

Branch A/C Dr.

To Goods sent to the Branch A/C

2) For Goods returned by the Branch to H.O.

Goods sent to the Branch A/C Dr.

To Branch A/C

3) For Goods sent by the Branch to another Branch at instructions from the H.O.

Goods sent to Branch A/C Dr.

To Branch A/C

4) For Goods returned by the Branch Debtors to H.O. directly

Goods sent to Branch A/C Dr.

To Branch A/C

5) For Expenses at the Branch met by the H.O.

Branch A/C Dr.

To Bank A/C

6) For Assets at the Branch at end of accounting period

Branch Assets A/C Dr.

To Branch A/C

7) For Liabilities at the end of the accounting period

Branch A/C Dr.

To Branch Liabilities/C

8) For Profit/Loss

If Profit: Branch A/C Dr.

To General P&L A/C

If Loss: General P&L A/c

To Branch A/C

9) For transfer of Balance in Goods sent to Branch account

Goods sent to Branch A/C Dr.

To Purchases/Trading A/C

10) For remittance of cash or cheque to the branch

Branch A/C Dr.

To Cash/Bank A/C

Question: Excellent Garments of Multan has a branch at Lahore. Goods are supplied to the branch at cost. The expenses of the branch are paid from Multan and the branch keeps a sales journal and the debtors' ledger only. From the following information supplied by the branch, prepare a Branch Account in the books of the head office. Goods are sent to branch at pro-forma invoice price which is cost plus 20%.(All figures in rupees).

Opening Stock (at Pro-forma invoice) 28,800, Closing Debtors9,150

Closing Stock (at Pro-forma invoice) 21,600

Opening Debtors 6200

Goods received from HO(at Pro-forma invoice)

Bad Debt 140

Credit Sales 41,000

Expenses paid by Head office 10,400

Cash Sales 17,500

Cash received from Debtors 37,900

Pilferage of goods by the employees(Normal Loss) 2,000

Solution:

(Debtors System)

In the books of H.O. (Multan)

Lahore Branch Account

Particulars	Amount	Particulars	Amount
Opening Stock	24,000	Cash Recd.fromBranch	17500
Opening Debtors	6200	Cash Recd. From Debtors	37900
Cash sent to Branch	10400	Goods sent to Branch	6720
Goods sent to Branch	40,320	Closing Stock	18000
To general P&L A/C	8360	Closing Debtors	9160
	89280		**89280**

Debtors Account

Particulars	Amount	Particulars	Amount
Opening		Cash recd.from	37900
Credit Sales	41000	Bad Debt	140
		Closing Debtors (c/.f.)	9160
	47200		**47200**

Question: On 1st January, 2008 goods costing Rs. 132,000 were invoiced by Multan head office to its new branch at Lahore and charged at selling price to produce a gross profit of 25% on the selling price. At the end of the year, the return from Lahore Branch showed that the credit sales were Rs. 150,000. Goods invoiced at Rs. 2,000 to Lahore branch have been returned to Multan head office. The closing stock at Lahore branch was Rs. 24,000 at selling price. Record the above transactions in the books of

(i)Lahore Branch Stock Account; (ii) Goods Sent to Lahore Branch Account; (iii)Lahore Branch Adjustment Account; and (iv) Lahore Branch Debtors Account in the head office book and close the said accounts on 31st December 2008.

In the Books of the Head Office, Multan

Lahore Branch Stock Account

Particulars	Amount	Particulars	Amount
To Goods sent to Lahore	176000	By goods sent to Lahore	2000
		By a/c returns	150000
		By Branch	24,000
	176000		**176000**

Goods Sent to Lahore Branch Account

Particulars	Amount	Particulars	Amount
To Lahore Branch Stock A/c(Returns)	2000	By Lahore Branch stock a/c	176000
To Lahore branch adj.a/c	43500		
Topurchases(bal.fig.)	130500		
	176000		**176000**

Lahore Branch Debtors Account

Particulars	**Amount**	**Particulars**	**Amount**
To Lahore Branch Stock a/c	150000	By balance c/d	150000
	150000		150000

Lahore Branch adjustment Account

Particulars	**Amount**	**Particulars**	**Amount**
To Stock Reserve	6000	By goods sent to Delhi Branch	43500
To general P&L a/c	37500		
	43500		43500

Independent Branch Accounting:

Accounting Entries under Independent Branch:

A. For goods supplied by head office to branch:

Branch book:

[Type text]

Goods supplied by head office A/C.........Dr.

To Head office A/C

(Being receipt of goods)

Head office book:

Branch A/C.............Dr.

To goods supplied to branch A/C

(Being goods sent to branch)

B. **For cash remitted by head office to branch:**

Branch book:

Cash A/C.............Dr.

To head office A/C

(Being cash received)

Head office book:

Branch A/C ....Dr.

To cash A/C

(Being cash sent to branch)

C. **For goods returned by branch:**

Branch book:

Head office A/C...........Dr.

To goods supplied to head office A/C

(Being goods return to head office)

Head office book:

Goods supplied from branch A/C...........Dr.

To Branch A/C

(Being goods returned from branch)

D. **For cash remitted by branch to head office:**

Branch book:

Head office A/C............Dr.

To cash

(Being cash sent to head office)

Head office book:

Cash A/C.................Dr.

To Branch A/C

(Being cash received from branch)

E. **For assets purchased by branch on behalf of head office:**

Branch book:

Head office A/C ....Dr.

To cash A/C

(Being purchase of assets)

Head office book:

Branch assets A/C.............Dr.

To branch A/C

(Being assets purchased by branch)

F. **For depreciation charged:**

Branch book:

Depreciation A/C ....Dr.

To Head office A/C

(Being depreciation on branch fixed assets)

Adjustment in Profit sharing Ratio

When a new partner is admitted he/she acquires his/her share in profit from the existing partners. As a result, the profit sharing ratio in the new firm is decided mutually between the existing partners and the new partner. The incoming partner acquires his/her share of future profits either incoming from one or more existing partner. The existing partners sacrifice a share of their profit in the favour of new partner, hence the calculation of new profit sharing ratio becomes necessary.

Sacrificing Ratio

At the time of admission of a partner, existing partners have to surrender some of their share in favour of the new partner. The ratio in which they agree to sacrifice their share of profits in favour of incoming partner is called sacrificing ratio. Some amount is paid to the existing partners for their sacrifice. The amount of compensation is paid by the new partner to the existing partner for acquiring the share of profit which they have surrendered in the favour of the new partner.

Sacrificing Ratio is calculated as follows:

Sacrificing Ratio = Existing Ratio – New Ratio

Following cases may arise for the calculation of new profit sharing ratio and sacrificing ratio:

Only the new partner's share is given

In this case, it is presumed that the existing partners continue to share the remaining profit in the same ratio in which they were sharing before the admission of the new partner. Then, existing partner's new ratio is calculated by dividing remaining share of the profit in their existing ratio. Sacrificing ratio is calculated by deducting new ratio from the existing ratio.

Question:

Deepak and Vivek are partners sharing profit in the ratio of 3 : 2. They admit Ashu as a new partner for 1/5 share in profit. Calculate the new profit sharing ratio and sacrificing ratio.

Solution:

Calculation of new profit sharing ratio:

Let total Profit = 1

New partner's share = 1/5

Remaining share = 1 – 1/5 = 4/5

Deepak's new share = 3/5 of 4/5 i.e. 12/25

Vivek's new share = 2/5 of 4/5 i.e. 8/25

Ashu's Share = 1/5

The new profit sharing ratio of Deepak, Vivek and Ashu is :

= 12/25 : 8/25 : 1/5 = 12 : 8 : 5/25 = 12 : 8 : 5

So Deepak Sacrificed = 3/5 – 12/25 = 15 – 12/25 = 3/25

Vivek Sacrificed = 2/5 – 8/25 = 10 – 8/25 = 2/25

Sacrificing Ratio = 3 : 2

Sacrificing ratio of the existing partners is same as their existing ratio.

(ii) The new partner purchases his/her share of the profit from the Existing partner in a particular ratio.

In this case : the new profit sharing ratio of the existing partners is to be ascertained after deducting the sacrifice agreed from his share. It means the incoming partner has purchased some share of profit in a particular ratio from the existing partners.

Only the new partner's share is given

In this case, it is presumed that the existing partners continue to share the remaining profit in the same ratio in which they were sharing before the admission of the new partner. Then, existing partner's new ratio is calculated by dividing remaining share of the profit in their existing ratio. Sacrificing ratio is calculated by deducting new ratio from the existing ratio.

When a new partner is admitted in the firm, the existing/old partners have to sacrifice, what is given to the new partner, from their future profits, the reputation they have gained in their past efforts and the side of capital they have taken before. The new partner when admitted, has to compensate for all these sacrifices made by the old ones. The compensation for such sacrifice can be termed as 'goodwill'. Hence, at the time of admission of the new partner, it is necessary to account the valuation of goodwill in the firm. If the new partner brings in cash for his share of goodwill, in addition to his capital, it is known as premium method. When the new partner brings nothing but only the capital, and the value of goodwill is erected or raised, this method of treatment is called Revaluation Method. However, once creating the value of goodwill and writing of the same after admission is done, it can be said to be Memorandum Revaluation Method. Thus, keeping in mind, all these methods, the various ways of treating goodwill in the books of the firm at the time of admission of the new partner, are as follows:

1. Share of goodwill brought by the new partner in cash.
2. Share of goodwill brought by the new partner in kind.
3. Nothing is brought by the new partner as his share of goodwill.
4. Share of goodwill brought by the new partner in cash only a portion not as a whole.
5. Hidden goodwill

1. When the new partner brings his share of goodwill in cash

When the new partner brings his share of goodwill in cash, the payment ma be made to the old partners, as if outside/private transaction. It may be retained in the business or after recording the same in the firm, the old partners may withdraw the whole amount or some portion only,

a. When the amount of goodwill brought by the new partner is not recorded in the books and the payment is made to the old partners as outside or private transaction, it does not affect in the transaction of the firm and hence no entry is passed in the books of the firm

b.When the amount of goodwill brought in by the new partner is retained in the business to increase cash resources, and if there exists already no-goodwill:

i) Cash/Bank A/C.......................Dr.

To Goodwill A/c

(Being goodwill brought in by the new partner)

ii) Goodwill A/C...........................Dr.

To old partners' capital A/C

(Being goodwill credited to old partners in the sacrificing ratio)

c. When there is no-goodwill already appeared in the books and the amount of goodwill brought in by the new partner, is fully or partially withdrawn by the old partners:

i) Cash A/C.......................Dr.

To Goodwill A/C

(Being goodwill brought by the new partner)

ii) Goodwill A/C................Dr.

To old partners' capital A/C

(Being goodwill divided among old partners)

iii) Old partners' capital A/C...............Dr.

To Cash/Bank A/C

(Being the amount withdrawn)

d. When there is goodwill already appeared in the books and even then if the new partner brings his share of goodwill in cash, the amount may be retained or withdrawn by the old partners. If the amount of goodwill brought in by the new partner is retained in the business:

i) Old partners' capital A/C.............................Dr.

To Goodwill A/C

(Being goodwill appearing in the book written off in the old ratio)

ii) Cash/Bank A/C.......................Dr.

To Goodwill A/C

(Being goodwill brought in by the new partner)

iii) Goodwill A/C.......................Dr.

To old partners' capital A/C

(Being goodwill brought in by new partner shared by the old partners)

If the goodwill amount brought by the new partner is withdrawn by the old partners, the following extra entry should also be passed:

Old partners' capital A/C.............Dr.

To Cash/Bank

(Being amount withdrawn)

If they agree to show the original value of goodwill in the books, it is raised by passing the entry:

Goodwill A/C ....Dr.

To All partners capital A/C

(Being goodwill raised)

When the new partner brings his share of goodwill in kind

The new partner may bring his share of goodwill and capital in kind i.e. the form of assets instead of cash. Again, new partner may have an established name in the market among the customers. In such case, he may be recognized for his goodwill. As a result he will bring a lesser amount of assets than the amount of credited to him. This requires two journal entries:

i) All assets A/C............................Dr.

Goodwill A/C/New partner's capital A/C

(Being goodwill brought in kind by the new partner)

ii) Goodwill A/C/New partner's capital A/C................Dr.

To old partners' capital A/C

(Being goodwill shared by the old partners)

3. When the new partner is unable to bring his share of goodwill in cash or kind

When the new partner cannot bring anything for his share of goodwill, first of all we have to see if there exists goodwill already or not. If there is no-goodwill already appearing in the books of the firm, goodwill is raised at its full value. If goodwill already appears in the books, it is compared to the full value of goodwill raised or created and the adjustment is done accordingly.

a. When the new partner is unable to bring his share of goodwill and if there is no-goodwill already appearing in the books, goodwill is raised at its full value:

i) Goodwill A/C.................Dr.

To old partners' capital A/C

(Being goodwill is created at its full value and credited to the old partners in old ratio)

* By this entry, goodwill A/C then appears as an asset in the balance sheet of the firm.

b. If the new partner cannot bring his share of goodwill and there appears goodwill already in the books, even then goodwill is raised at its full value. If the raised value of goodwill is equal to the existing value of goodwill, no entry what so ever is needed. If the raised goodwill is more than the existing goodwill, then goodwill will be credited to the old partner's capital A/C by the excess amount only:

Goodwill A/C......................Dr. (excess value)

To old partners' capital A/C

(Being the value of goodwill increased to..../increased by......)

* Goodwill then appears at its full value in the balance sheet of the firm

c. If the raised value of goodwill is less than the existing value of goodwill, then excess over raised value of goodwill is written off:

Old partners' capital A/C................Dr.

To Goodwill A/C

(Being the goodwill written off by the reduction in value)

d. Whatever the case may be stated in a,b,c, the partners may not wish goodwill in the books for an indefinite period after the admission of new one, as the value of goodwill changes constantly. They may write off the whole or some portion of the value of goodwill. For writing off the goodwill:

All partners' capital A/C.............Dr.

To Goodwill A/C

(Being goodwill written off)

4. When the new partner can bring only a portion of his share of goodwill

When the new partner cannot bring the entire amount of his share of goodwill and he brings only a part of this, it is shared by the old partners in sacrificing ratio. Then goodwill A/C is raised in the books for the portion not brought by the new partner which is also credited to the old partners in their sacrificing ratio. Goodwill raised for the part of goodwill not brought in by the new partner is calculated as under:

= (Full value of goodwill/share of goodwill of new partner) X goodwill not brought in

But it should be remembered that , if there exists any goodwill in the books, first it should be written off by crediting to the old partners in old ratio. Therefore, the entries are:

i) Old partners' capital A/C..................Dr.

To Goodwill A/C

(Being goodwill written off)

ii Cash/Bank A/C........................Dr.

To Goodwill A/C

(Being the portion of goodwill brought in by new partner)

iii) Goodwill A/C.......................Dr.

To old partners' capital A/C

(Being the goodwill brought in by new partner credited to old partners)

iv) When the goodwill is raised for the part of goodwill not brought in by the new partner, the amount of goodwill is calculated as said above. The entry would be the same as in iii), only the amount being different, which is shared by the old partners in their old profit sharing ratio.

5. Hidden Goodwill

When the value of goodwill is not given in the question, the value of goodwill has to be calculated on the basis of total capital/net worth of the firm and profit sharing ratio.

A. New partner's capital X Reciprocal of the share of new partner....XXX

B. Less net worth(excluding goodwill) of new firm...............................XXX

C. A-B = Value of goodwill..XXX

<u>Retirement of a Partner:</u>

When one or more partners leaves the firm and the remaining partners continue to do the business of the firm, it is known as retirement of a partner. Amit, Sunil and Ashu are partners in a firm. Due to some family problems, Ashu wants to leave the firm. The other partners decide to allow him to withdraw from the partnership. Thus, due to some reasons like old age, poor health, strained relations etc., an existing partner may decide to retire from the partnership. Due to retirement, the existing partnership comes to an end and the remaining partners form a new agreement and the partnership firm is reconstituted with new terms and conditions. At the time of retirement the retiring partner's claim is settled.

A partner retires either :

(i) with the consent of all partners, or

(ii) as per terms of the agreement; or

(iii) at his or her own will.

The terms and conditions of retirement of a partner are normally provided in the partnership deed. If not, they are agreed upon by the partners at the time of retirement. At the time of retirement the following accounting issues are dealt :

(a) New profit sharing ratio and gaining ratio.

(b) Goodwill

(c) Adjustment of changes in the value of Assets and liabilities

(d) Treatment of reserve and accumulated profits.

(e) Settlement of retiring partners dues,

(f) New capital of the continuing partners.

New profit sharing ratio and gaining ratio

As soon as a partner retires the profit sharing ratio of the continuing partners get changed. The share of the retiring partner is distributed amongst the continuing partners. In the absence of information, the continuing partners take the retiring partner's share in their profit sharing ratio or in an agreed ratio. The ratio in which retiring partner's share is distributed amongst continuing partners is known as gaining ratio. It is

Gaining Ratio = New Ratio – Existing Ratio

Various cases of new ratio and gaining ratio are illustrated as follows:

(i) Retiring partner's share distributed in Existing Ratio :

In this case, retiring partner's share is distributed in existing ratio amongst the remaining partners. The remaining partners continue to share profits and losses in the existing ratio.

The following example illustrates this :

Tanu, Manu and Rena are partners sharing profits and losses in the ratio of = 4 : 3 : 2. Tanu retires and remaining partners decide to take Tanu's share in the existing ratio i.e. 3 : 2. Calculate the new ratio of Manu and Rena.

Existing Ratio between Manu and Rena = 3/9 and 2/9

Tanu's Ratio (retiring partner) = 4/9

Tanu's share taken by the Manu and Rena in the ratio of 3 : 2

Manu's gets = 4/9 × 3/5 = 12/45

Manu's New Share = 3/9 + 12/45 = 27/45

Rena's gets = 4/9 × 2/5 = 8/45

Rena's New Share = 2/9 + 8/45 = 18/45

New ratio between Manu and Rena is 27/45 : 18/45 = 27 : 18 = 3 : 2.

Gaining Ratio = New Ratio – Existing Ratio

Manu Gain = 27/45 – 3/9 = 12/45

Rena Gain = 18/45 – 2/9 = 8/45

12/45 : 8/45

3 : 2

You may note that the new ratio is similar to existing ratio that existed between Manu and Rena before Tanu's retirement.

Note: In absence of any information in the question, it will be presumed that retiring partner's share has been distributed in existing ratio.

(ii) Retiring partner's share distributed in Specified proportions: Sometimes the remaining partners purchase the share of the retiring partner in specified ratio. The share purchased by them is added to their old share and the new ratio is arrived at. The following example illustrates this:

A, B and C are partners in the firm sharing profits in the ratio of 3 : 2 : 1. B retired and his share was divided equally between A and C. Calculate the new profit sharing ratio of A and C.

B's Share = 2/6

B's share is divided between A and C in the ratio of 1 : 1.

A gets 1/2 of 2/6 = 2/6 × 1/2 = 1/6

A's New Share = 3/6 + 1/6 = 4/6

C's gets 1/2 of 2/6 = 2/6 × 1/2 = 1/6

C's New share = 1/6+1/6 = 2/6

Gaining Ratio = New Ratio – Existing, Ratio

(iii)Retiring Partner's share is taken by one of the partners

The retiring partner's share is taken up by one of the remaining partners. In this case, the retiring partner's share is added to that of partner's existing share. Only his/her share changes. The other partners continue to share profit in the existing ratio. An example illustrating this point is given below: Anuj, Babu and Rani share profit in the ratio of 5 : 4 : 2. Babu retires and his share is taken by Rani, So Rani's share is 2/11 + 4/11 = 6/11, Anuj share will remain unchanged i.e, 5/11. Thus, the new profit sharing ratio of Anuj and Rani is 5 : 6.

Treatment of Goodwill

The retiring partner is entitled to his/her share of goodwill at the time of retirement because the goodwill is the result of the efforts of all partners including the retiring one in the past. The retiring partner is compensated for his/her share of goodwill. As per Accounting Standard 10 (AS-10), goodwill is recorded in the books only when some consideration in money is paid for it. Therefore, goodwill is recorded in the books only when it is purchased and the goodwill account cannot be raised on its own. Therefore, in case of retirement of a partner, the goodwill is adjusted through partner's capital accounts. The retiring partner's capital account is credited with. his/her share of goodwill and remaining partner's capital account is debited in their gaining ratio. The journal entry is made as under:

Remaining Partners' Capital A/c Dr. (individually)

To Retiring Partner's Capital A/c

(Retiring partner's share of goodwill adjusted to remaining partners in the gaining ratio).

When the Goodwill Account already appears in the Books

Normally the goodwill is not shown in the books of the firm. If at the time of retirement/death of a partner, goodwill appears in the Balance Sheet of the firm, it will be written off by debiting all the partners' capital account in their existing profit sharing ratio and crediting the goodwill account. In such a case, the following journal entry is made:

Partners' Capital A/c Dr (including retiring partner's capital A/c)

To Goodwill A/c

(Existing goodwill written-off)

Revaluation of Assets and Liabilities:

At the time of retirement of a partner the assets and liabilities of the firm are revalued and Revaluation Account is prepared in the same way as in case of admission of a partner. This is done to adjust the changes in value of assets and liabilities at the time of retirement/death of a partner. Any profit or loss due to revaluation is divided amongst all the partners including Retiring/deceased in their existing profit sharing ratio. Following journal entries are made for this purpose :

(i) **For increase in value of assets:**

Assets A/c Dr. [Individually]

To Revaluation A/c

(Increase in the value of assets)

(ii) **For decrease in value of assets:**

Revaluation A/c Dr.

To Assets A/c (Individually)

(decrease in the value of asset)

(iii) **For increase in value of Liabilities**:

Revaluation A/c Dr.

To Liabilities A/c [Individually]

(Increase in the value of liabilities)

(iv) **For decrease in value of Liabilities:**

Liabilities A/c Dr. [Individually]

To Revaluation A/c

(decrease in the value of liabilities)

Revaluation account is prepared to record the change in the value of assets or liabilities. It will reveal profit or loss on revaluation. This profit or loss is divided amongst all partners including the retiring/deceased partner in existing profit sharing ratio.

(iv) **For Profit on Revaluation :**

Revaluation A/c Dr. (Individually)

To Partner’s Capital A/c

(Profit on revaluation divided amongst all partners in their existing profit sharing ratio)

(v**) For loss on Revaluation:**

Partner’s Capital A/c Dr. (Individually)

To Revaluation A/c

(Loss on revaluation borne by all partners in their existing profit sharing ratio.

(vi**) For distribution of undistributed profit and reserve.**

Reserves A/c Dr

Profit & Loss A/c (Profit) Dr.

To Partners’ Capital A/c (individually)

(Reserves and Profit & Loss (Profit) transferred to all partners capitals A/c in existing profit sharing ratio)

(vii)**For distribution of undistributed loss** Partners’ Capital A/c Dr. (individually)

To Profit & Loss A/c (Loss) [Profit & Loss (loss) transferred to all partners Capitals A/c in old profit sharing ratio]

Death of a Partner:

The death of a partner dissolves the partnership. On the date of death, the accounts are closed and the net income for the year to date is allocated to the partners' capital accounts. Most agreements call for an audit and revaluation of the assets at this time. The balance of the deceased partner's capital account is then transferred to a liability account with the deceased's estate.

The surviving partners may continue the business or liquidate. If the business continues, the procedures for settling with the estate are the same as those described earlier for the withdrawal of a partner.

On the death of a partner, the accounting treatment regarding goodwill, Revaluation of assets and reassessment of liabilities, accumulated reserves and undistributed profit are similar to that of the retirement of a partner, When the partner dies the amount payable to him/her is paid to his/her legal representatives. The representatives are entitled to the followings :

(a) The amount standing to the credit to the capital account of the deceased partner

(b) Interest on capital, if provided in the partnership deed upto the date of death:

(c) Share of goodwill of the firm;

(d) Share of undistributed profit or reserves;

(e) Share of profit on the revaluation of assets and liabilities; (f) Share of profit upto the date of death;

(g) Share of Joint Life Policy.

The following amounts are debited to the account of the deceased partner's legal representatives: (i) Drawings

(ii) Interest on drawings

(iii) Share of loss on the revaluation of assets and liabilities;

(iv) Share of loss that have occurred till the date of his/her death.

The above adjustments are made in the capital account of the deceased partner and then the balance in the capital account is transferred to an account opened in the name of his/her executor.

The payment of the amount of the deceased partner depends on the agreement. In the absence of an agreement the legal representative of adeceased partner is entitled to interest @ 6% p.a. on the amount due from the date of death till the date of final payment.

Calculation of profit upto the date of death of a partner.

If the death of a partner occurs during the year, the representatives of the deceased partner are entitled to his/her share of profits earned till the date of his/her death. Such profit is ascertained by any of the following methods: (i) Time Basis

(ii) Turnover or Sales Basis

(i) Time Basis

In this case, it is assumed that profit has been earned uniformly throughout the year. For example:

The total profit of previous year is Rs. 2,25,000 and a partner dies three months after the close of previous year, the profit of three months is Rs. 31,250 i.e. 1,25,000 × 3/12, if the deceased partner took 2/10 share of profit, his/her share of profit till the date of death is Rs. 6,250 i.e. Rs. 31,250 × 2/10.

(ii) Turnover or Sales Basis

In this method, we have to take into consideration the profit and the total sales of the last year. Thereafter the profit upto the date of death is estimated on the basis of the sale of the last year. Profit is assumed to be earned uniformly at the same rate.

Illustration 12

Arun, Tarun and Neha are partners sharing profits in the ratio of 3 : 2 : 1 Neha dies on 31st May 2006. Sales for the year 2005-2006 amounted to Rs.4,00,000.and the profit on sales is Rs.60,000. Accounts are closed on 31 March every year. Sales from lst April 2006 to 31st May 2006 is Rs.1,00,000.Calculate the deceased partner's share in the profit upto the date of death.

Solution :

Profit from 1st April 2006 to 31st May 2006 on the basis of sales: If sales are Rs.4,00,000, profit is Rs.60,000

If the sales are Rs.1,00,000 profit is : 60,000/4,00,000 × 1,00,000 = Rs.15,000

Neha's share= 15,000 × 1/6 = Rs.2,500

Alternatively profit is calculated as

Rate of profit =60000/400000×100 = 15%

Sale up to date of death = 1,00,000

Profit = 1 00 000X15/100 = Rs 15000.

BUSINESS ECONOMICS (107)

UNIT – 1

INTRODUCTION TO BUSINESS ECONOMICS & FUNDAMENTAL CONCEPTS INTRODUCTION: -

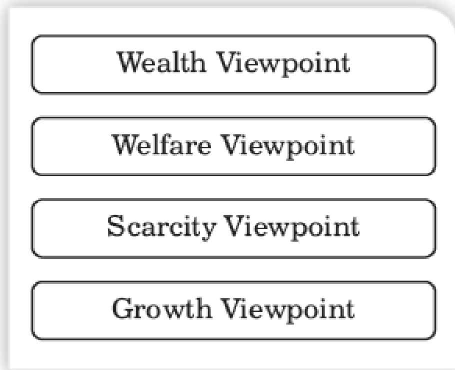
In simple terms, economics can be defined as a discipline that studies the behaviour patterns of human beings. The main aim of economics is to analyse how individuals, households, organisations, and nations use their scarce resources to satisfy their wants and needs or to achieve maximum profit. Economics is broadly classified into two parts, namely microeconomics and macroeconomics. Microeconomics is a branch of economics that studies the behaviour of individual consumers and organisations in the market. It focuses on the demand and supply, pricing, and output of individual organisations. On the other hand, macroeconomics examines the economy as a whole and deals with issues related to national income, employment pattern, inflation, recession, and economic growth.

With the advent of globalisation and rise in competition, it is of paramount importance for managers to make rational decisions. For this, managers should have a clear understanding of different economic concepts, theories, and tools. Business economics or managerial economics is a specialised discipline of economics that undertakes a study of various economic theories, logics, and tools used in business decision making. It applies various economic concepts, such as demand and supply, competition, allocation of resources, and economic trade-offs, to help managers in making better decisions.

MEANING OF ECONOMICS: -

In simple terms, economics can be defined as the study of how individuals, households, organisations, and nations make optimum utilization of scarce resources to satisfy their wants and needs. It is the study of mankind's attempt to satisfy their unlimited wants with the help of limited resource The word economics has originated from a Greek word oikonomikos, which can be divided into two parts: oikos means home and nomos means management. Thus, in earlier times, economics was referred to as home management where the head of a family managed the needs of family members from his limited income. However, over the years, the scope of economics has broadened to society and how it satisfies the needs of people by using limited resources.

Different economists have different viewpoints on economics. Some economists had a viewpoint that economics is a study of money, while others believed that economics deals with problems, such as



inflation and unemployment. Therefore, to simplify the concept, economics is defined by taking four viewpoints,

Let us study these viewpoints in detail.

‰‰Wealth viewpoint: This is a classical viewpoint on economics that was given by Adam Smith, who is also considered as the father of modern economics. According to him, Economics is "the study of the nature and causes of nations' wealth or simply as the study of wealth." He stated that the main purpose of all economic activities is to gain maximum wealth as possible. In Smith's view, the citizens of wealthy nations are happy; thus, economics shows nations to be wealthy.

‰‰Welfare viewpoint: It is a neo-classical viewpoint on economics that was given by Alfred Marshall. According to Alfred Marshall, "Economics is a study of man in the ordinary business of life. It enquires how he gets his income and how he uses it. Thus, it is on the one side, the study of wealth and on the other and more important side, a part of the study of man." He associated economics with the welfare of men, who are responsible for generating wealth

‰‰Scarcity viewpoint: It is a pre-Keynesian thought of economics that was given by Lionel Robins in his book 'Essays on the Nature and Significance of the Economic Science' (1932). According to Robins, "Economics is a science which studies human behavior as a relationship between ends and scarce means which have alternative uses". The definition focused on human behaviour in the optimum utilisation of scarce resources. It provides three basic features of human existence, which are unlimited wants, limited resources, and alternative uses of limited resources

‰‰Growth viewpoint: This is the modern perspective of economics mainly given by Paul Samuelson. He provided the growth-oriented definition of economics. According to him, "Economics is a study of how men and society choose with or without the use of money, to employ scarce productive uses resource which could have alternative uses, to produce various commodities over time and distribute them for consumption, now and in the future among the various people and groups of society.” The definition outlines three main aspects, namely human behaviour, allocation of resources, and alternative uses of resources.

NATURE OF ECONOMICS

Let us now understand the true nature of economics.

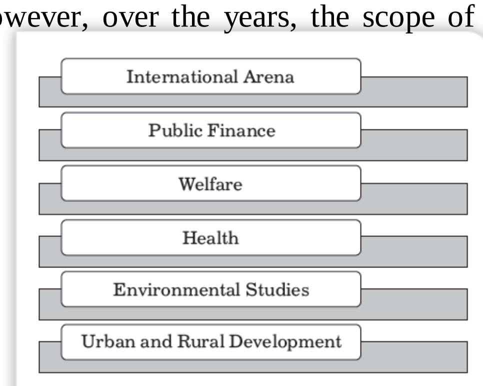
‰‰Economics as a science: Science is a branch of knowledge that defines the relationship between cause and effect. As results observed in science are measurable and based on facts, economics also endeavours to find a relationship between cause and effect and provides measurable results. Similar to science, in economics, emphasis is laid on collecting relevant information, which is categorized and analysed to reach conclusions.

‰‰Economics as a social science: Economics is also considered as social science as it deals with studying the behaviour of human beings and their relationships in a society. This is because the exchange of goods takes place within the society and among different societies to satisfy the needs and wants of people.

‰‰Economics is an art: Art is a branch of study that deals with expressing or applying the creative skills and imagination of humans to perform a certain activity. Similarly, economics also requires human imagination for the practical application of scientific laws, principles, and theories to perform a particular activity.

SCOPE OF ECONOMICS

Earlier, the scope of economics was limited to the utilisation of scarce resources to meet needs and wants of people and society. However, over the years, the scope of economics has been broadened to many areas,



Let us study about the scope in detail.

‰‰International arena: With the advent of globalisation and cross-border integration, economic concepts are applied in order to conduct successful business dealings between countries. Economic concepts can be used in areas, such as foreign trade (exports and imports), foreign exchange (trading currency), balance of payments, and balance of trade.

‰‰Public finance: Economic concepts are also applied to assess the government's collection of taxes from the users of public goods as well as expenditure on production and distribution of these goods to the general public.

‰‰Welfare: Economic theories and concepts are used to analyse the growth and development of low-income countries. This helps in improving the living standard of people in less developed and developing societies by understanding their needs for various facilities and utilities, such as health and education facilities and good working conditions.

‰‰Health: Economic concepts are also applicable in assessing the problems faced in promoting health in different countries. These concepts help the government in making decisions for defining appropriate health packages and programs for the general public.

‰‰Environmental studies: Economic concepts are used to analyse the utilisation and depletion of natural resources. Moreover, they are applied to study the impact of increasing ecological imbalance on society.

‰‰Urban and rural development: In urban development, the scope of economics covers the analysis of different urban issues such as crime, education, public transit, housing, and local government finance. On the other hand, in rural development, economics can be used to analyse the shortage of natural resources, obtain the best price for production, study constraints of productivity, adapt to climate change, etc.

DEFINING BUSINESS ECONOMICS

Organisations face many problems on a day to day basis. These problems require careful analysis and thoughtful consideration. For example, organisationsare always concerned with producing maximum output in the most economical way. To solve problems of such nature, managers are required to apply various economic concepts and theories. The application of economic theories, economic principles, economic tools, economic laws, equations economic concepts in business decision making is called business economics or managerial economics.

Business economics is the study of how scarce resources are directed most efficiently to achieve managerial goals. It is a valuable tool for analyzing business situations to take better decisions.

Prof. Evan J Douglas defines Managerial Economics as "Managerial Economics is concerned with the application of economic principles and methodologies to the decision making process within the firm or organization under the conditions of uncertainty. It seeks to establish rules and principles to facilitate the attainment of the desired economic goals of management"

According to Milton H Spencer and Louis Siegelman "Managerial Economics is the integration of economic theory with business practices for the purpose of facilitating decision making and forward planning by management"

Siegel man has defined managerial economic (or business economic) as "the integration of economic theory with business practice for the purpose of facilitating decision-making and forward planning by management." We may, therefore, define business economic as that discipline which deals with the application of economic theory to business management. Business economic thus lies on the borderline between economic and business management and serves as a bridge between the two disciplines.

From the aforementioned definitions, it can be concluded that managerial economics is a link between two disciplines, which are management and economics. The management discipline focuses on a number of principles that aid the decision-making process of organisations. On the other hand, economics is related to an optimum allocation of limited resources for attaining the set objectives of organisations.

Therefore, it can be said that managerial economics is a special discipline of economics that can be applied in business decision making of organisations.

The purpose of Managerial Economics is to show how economic analysis can be used in formulating Business Policies.

Economics can be divided into two broad categories: micro economics and macro economics. Macro economics is the study of the economic system as a whole. It is related to issues such as determination of national income, savings, investment, employment at aggregate levels, tax collection, government expenditure, foreign trade, money supply etc.,

Micro economics focuses on the behavior of the individuals, firms and their interaction in markets.

Managerial economics is an application of the principles of micro and macro economics in managerial decision making. The economic way of thinking about business decision making provides all managers with a powerful set of tools and insights for furthering the goals of their organization. Successful managers take good decisions, and one of their most useful tools is the methodology of managerial economics.

SCOPE OF BUSINESS ECONOMICS

Business economics involves the application of various economic tools, theories, and methodologies for analyzing and solving different business problems. These business problems can be related to demand and supply prospects of an organisation, level of production, pricing, market structure, and degree of competition.

Following are the scope of Business Economics:

1. Demand analysis and forecasting: Demand refers to the willingness or capability of individuals to buy a product at a specific price. Demand analysis is a process of identifying potential consumers, the amount of goods they want to purchase, and the price they are willing to pay for it. This process is important for an organization to analyse the demand for its products and produce accordingly. In business economics, demand forecasting occupies an important place by helping organisations in business planning and deciding on strategic issues.
2. Cost and benefit analysis (CBA): By analysing costs, management can estimate costs required for running the organization successfully. Cost analysis helps firms in determining hidden and uncontrollable costs and taking measures for effective cost control. It further enables the organisation to determine the return on investment (ROI). In a nutshell, CBA is a process of comparing the costs and benefits of a particular project
or activity. Business economics involves various aspects of cost and benefit analysis, such as cost-output relationships and cost control.
3. Production Analysis: Production analysis is narrower, in scope than cost analysis. Production analysis frequently proceeds in physical terms while cost analysis proceeds in monetary terms. The main topics covered under cost and production analysis are: Cost concepts and classification, Cost-output Relationships, Economies and Diseconomies of scale, Production function and Cost control.
4. Pricing decisions, policies, and practices: Pricing is one of the key areas of business economics. It is a process of finding the value of a product or service that an organisation receives in exchange for its product/service. The profit of an organisation depends a great deal on its pricing

strategies and policies. Business economics includes various pricing- related concepts, such as pricing methods, product-line pricing, and price forecasting.

5. Profit maximisation: Profit generation and maximisation is the main aim of every organisation (except for non-profit organisations). In order to maximise profit, organisations need to have complete knowledge about various economic concepts, such as profit policies and techniques, and break-even analysis.

6. Capital management: Organisations often find it difficult to make decisions related to capital investment. These decisions require sound knowledge and expertise on various economic aspects. To make sound capital investment decisions, an organisation needs to determine various aspects, such as cost of capital and rate of return.

Conclusion: The various aspects outlined above represent major uncertainties which a business firm has to reckon with viz., demand uncertainty, cost uncertainty, price uncertainty, profit uncertainty and capital uncertainty. We can therefore, conclude that the subject matter of business economic consists of applying economic principles and concepts to dea1 with various uncertainties faced by a business firm.

DIFFERENCE BETWEEN BUSINESS ECONOMICS AND ECONOMICS

Economics and business economics are different from each other in various aspects. As discussed earlier, economics is a study of human behaviour in making decisions related to the allocation of resources. Business economics, on the other hand, deals with managerial decision making in organisations. The following points distinguish between economics and business economics:

‰‰Economics is a traditional subject that has prevailed from a long time, while business economics is a modern concept and is still developing.

‰‰ Economics focuses primarily with the theoretical aspect whereas Business Economics devotes with the practical aspect. The former is associated with concepts, theories, models and building theoretical framework. The latter is associated with the applications of the selected theories and concepts to solve business problems and help the business decision making process.

Business Economics is essentially normative in nature. But, the Economics is concerned with both positive and normative economics. Positive Economics explains the economic phenomena as they are, while normative economics discusses as to what they ought to be. Business Economics explains what objectives and avenues a business should pursue and how they are to be. Therefore, it is normative in nature.

‰‰ Economics studies the complex economic phenomena and rational human behaviour by developing certain meaningful and consistent assumptions, hypothesis and developing models. Business Economics endeavors to solve real life complex business problems. It selectively applies economic models with required modifications to solve the business problems.

‰‰ Economics concentrates only the economic aspect of the problems but Business Economics deals with some non-economic aspects of the problems along with the economic aspects.

‰‰ Business Economics is fundamentally micro-economic in nature. It studies the activities of an individual firm or unit. There is an extensive application of the concepts and theories of microeconomics in it. The Economics has both micro and macro aspects within its purview.

Business Economics focuses on the theory of profit only. Whereas, the Economics has within its ambit not only profit maximization but also other aspects like Utility maximization, distribution theories of wage, rent interest and welfare economics as well.

‰‰Economics has a wider scope and covers the economic issues of nations, whereas business economics is a part of economics and is limited to the economic problems of organisations.

The scope of Business Economics is restricted as compared to the scope of the Economics.

Thus, it can be stated that economics is a wide concept that can be applied to various fields, whereas business economics is a narrow approach that can be applied in selected areas.

CONTRIBUTION & APPLICATION OF BUSINESS ECONOMICS IN BUSINESS

Setting business goals: Forecasting from marketing models are used to set revenue and profit goals. These objectives can often become metrics for performance evaluations of employees and managers.

Creating a pricing strategy: Managerial economics uses supply/demand curves to predict how consumers will react to price changes.

Deciding how much product to produce: Depending on the projections from sales forecasts, managers have to decide how much of each product to produce and at what price points.

Creating an internet strategy: Developing an effective internet strategy is about understanding SEO, driving traffic and monetizing a website. Economics gets applied to define the demographics of the visitors to the site and creating a content marketing strategy to develop those consumers.

Hiring policies needed to attract labor: Workers want to receive reasonable pay and benefits and have some assurance of long-term stability in their jobs. Managers must balance the marginal cost of labor with the incremental revenues received from product expansions or introductions of new products.

Evaluating investments and capital budgets: Long-term investments in plant and equipment are typically assessed and prioritized using a type of discounted cash flow technique.

Marketing and promotional strategies: Marketing strategies rely on the level of consumer demand for goods and services. Marketing managers try to estimate the size of the market for existing or new products. However, the market size depends on non-economic and economic factors which are represented by the price/demand curves for a product. Managerial economics applies income and price elasticity to make projections of demand.

Introducing new products: Managers use statistical forecasting and supply/demand curves to gauge the potential success of launching a new product. Discounted cash flow projections analyze future cash expenditures for the cost of a new plant and equipment and cash inflows from revenues.

Planning production schedules: Sales forecasts from marketing must be translated into production schedules, inventory quantities and number of workers needed on a production line. Managerial economics analyzes labor performance and provides insights into labor productivity and effects of the law of diminishing returns.

Micro v/s Macro Economics

Micro Economics talks about the actions of an individual unit, i.e. an individual, firm, household, market, industry, etc. On the other hand, the

Macro Economics studies the economy as a whole, i.e. it assesses not a single unit but the combination of all i.e. firms, households, nation, industries, market, etc.

Economics' is defined as the study of how humans work together to convert limited resources into goods and services to satisfy their wants (unlimited) and how they distribute the same among themselves. Economics has been divided into two broad parts i.e. Micro Economics and Macro Economics. There are two broad categories into which Economics is classified, i.e. Micro Economics and Macro Economics.

Opportunity Costs

Opportunity cost in economics can be defined as benefits or value missed out by business owners, small businesses, organization, investors, or an individual because they choose to accomplish or achieve anything else. It helps organizations in better decision-making by showing the lost opportunity because of investing over an alternative which can be anything like shares, stock market, real estate, land, services, etc. Generally, the financial report does not show the opportunity cost because it is not only about money or monetary cost. It is also associated with the lost time invested somewhere else which is providing utility. In simple terms, it is a concept in microeconomics that tells you about the output and potential opportunities foregone.

It shows the relation between choice and scarcity.

Definition of Opportunity Cost in Economics

In modern economic analysis, the factors of production are scarce as compared to the wants.

Therefore, when society uses a certain factor in the production of a specific commodity, then it forgoes other commodities for which it could use the same factor. This led to the idea of an opportunity cost (OC).

Let’s say that a certain kind of steel is needed to manufacture weapons for war. Therefore, society

has to give up the number of utensils that it could produce using the same amount of steel. Hence, the opportunity cost of producing weapons for war is the number of utensils forgone.

In other words, opportunity costs are the costs of the next best alternative forgone. Therefore, we can deduce two important aspects:

1. The opportunity costs of a product are only the best alternative forgone and not any other alternative.
2. These costs are viewed as the next-best alternative goods that we can produce with the same value of factors which are more or less the same.

How to Calculate Opportunity Cost

Formula of Opportunity cost = Return of Investment from the best option available – Return of investment from the chosen option.

Examples of Opportunity Cost

Let's understand these costs with the help of an illustration.

Let's say that a farmer has a piece of land on which he can grow wheat or rice. Therefore, if he chooses to grow wheat, then he cannot grow rice and vice-versa.

Hence, the opportunity cost for rice is the wheat crop that he forgoes. The following diagram explains this:

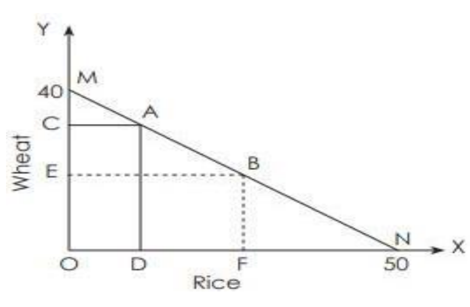
Example 1: If a person is having cash in hand Rs. 100000/-, he may think of two alternatives to increase cash.

Option 1: Investing in bank. We will get returns amount 10000/-

Option2: Investing in business. We get returns amount 17000/-

Generally we chose the option 2 because we will get more returns than the option 1. Here the option 1 is the opportunity cost, that what we have not chosen.

Opportunity Cost Graph –



Let’s assume that the farmer can produce either 50 quintals of rice (ON) or 40 quintals of wheat (OM) using this land. Now, if he produces rice, then he cannot produce wheat.

Therefore, the OC of 50 quintals of rice (ON) is 40 quintals of wheat (OM).

Further, the farmer can choose to produce any combination of the two crops along the curve MN (production possibility curve). Let’s say that he chooses the point A as shown above.

Therefore, he produces OD amount of rice and OC amount of wheat. Subsequently, he decides to shift to point B. Now, he has to reduce the production of wheat from OC to OE in order to increase the production of rice from OD to OF.

Therefore, the OC of DF amount of rice is CE amount of wheat.

Applications of Opportunity Cost

- Determining factor prices
- Determining economic rent
- Consumption pattern decisions
- Determining factor prices
- Product plan decisions
- Decisions about national priorities

Determining factor prices

The factors for production need a price equal to or greater than what they command for alternative uses. If the factor price is less than the factor’s opportunity cost, then the said factor moves to the better-paying alternative.

Determining economic rent

Many modern economists use this concept for determining economic rent. As per them, economic rent = The factor’s actual earning – Its opportunity cost or transfer earning

Consumption pattern decisions

According to this concept, if with a given amount of money a consumer chooses to have more of one thing, then he needs to have less of the other.

Further, he cannot increase the consumption of all the goods at the same time. Therefore, he decides his consumption pattern using the concept of opportunity cost.

Product plan decisions

Let’s say that a producer has fixed resources and technology. If he wants to produce a greater amount of one commodity, then he must sacrifice the quantity of another commodity.

Therefore, he uses this concept to make decisions about his production plan.

Decisions about national priorities

Every country has certain resources at its command and needs to plan the production of a wide range of commodities. This decision depends on the national priorities which are based on opportunity costs.

For example, if a country is at war, then it will use its resources to produce more war-related goods as compared to civilian goods.

This concept helps the country in making these decisions.

What is the Difference between Sunk Cost and Opportunity Cost?

The sunk cost can be defined as the financial cost which is already invested and now it cannot be incurred or money you cannot get back.

For example, if a company purchases 1000s of laptops for $1000000, then that money is sunk i.e. the company cannot get the money back for those laptops. To get that money back we need to get the amount higher than the purchase price.

Opportunity cost is how much less return of investment a company received because of investing capital somewhere else.

Types of Opportunity Cost in Production

- Explicit Cost
- Implicit Cost
- Marginal Opportunity Cost

What is Explicit Cost?

Explicit costs are the cost which includes the monetary payment from the producers. For example, if the company is paying $1000 per month in food by providing free lunch and breakfast, then its explicit OC is $1000. The expenditure on food could have been used somewhere else.

What is Implicit Cost?

Implicit cost aka notional cost can be defined as the OC which a company used in order to produce something. For example, a company purchased small electronic devices to produce mobile phones, laptops, etc. This cost is used to produce something, the electronic devices are not sold or rented.

What is Marginal Opportunity Cost?

Marginal opportunity cost is a cost required to produce something extra. For example, currently a company is producing 1000 burgers per day, but due to heavy demand, they are running out of the burgers. So, the company decided to hire more people and cook more burgers.

Now marginal opportunity cost will include – payment of new employees, cost required for ingredients required to cook more burgers, profit company was missing before and many other extra costs required for producing additional burgers.

Time Value of Money

The time value of money (TVM) is the concept that money available at the present time is worth more than the identical sum in the future due to its potential earning capacity. This core principle of finance holds that provided money can earn interest, any amount of money is worth more the sooner it is received. TVM is also sometimes referred to as present discounted value.

Understanding Time Value of Money (TVM)

The time value of money draws from the idea that rational investors prefer to receive money today rather than the same amount of money in the future because of money's potential to grow in value over a given period of time. For example, money deposited into a savings account earns a certain interest rate and is therefore said to be compounding in value.

Further illustrating the rational investor's preference, assume you have the option to choose between receiving $10,000 now versus $10,000 in two years. It's reasonable to assume most people would choose the first option. Despite the equal value at the time of disbursement, receiving the $10,000 today has more value and utility to the beneficiary than receiving it in the future due to the opportunity costs associated with the wait. Such opportunity costs could include the potential gain on interest were that money received today and held in a savings account for two years.

Time Value of Money Formula

Depending on the exact situation in question, the time value of money formula may change slightly. For example, in the case of annuity or perpetuity payments, the generalized formula has additional or less factors. But in general, the most fundamental TVM formula takes into account the following variables:

- FV = Future value of money
- PV = Present value of money
- i = interest rate
- n = number of compounding periods per year
- t = number of years

Based on these variables, the formula for TVM is:

FV = PV x [1 + (i / n)] (n x t)

Time Value of Money Examples

Assume a sum of $10,000 is invested for one year at 10% interest. The future value of that money is:

FV = $10,000 x (1 + (10% / 1) ^ (1 x 1) = $11,000

The formula can also be rearranged to find the value of the future sum in present day dollars. For example, the value of $5,000 one year from today, compounded at 7% interest, is:

PV = $5,000 / (1 + (7% / 1) ^ (1 x 1) = $4,673

Effect of Compounding Periods on Future Value

The number of compounding periods can have a drastic effect on the TVM calculations. Taking the $10,000 example above, if the number of compounding periods is increased to quarterly, monthly or daily, the ending future value calculations are:

- Quarterly Compounding: FV = $10,000 x (1 + (10% / 4) ^ (4 x 1) = $11,038
- Monthly Compounding: FV = $10,000 x (1 + (10% / 12) ^ (12 x 1) = $11,047
- Daily Compounding: FV = $10,000 x (1 + (10% / 365) ^ (365 x 1) = $11,052

This shows TVM depends not only on interest rate and time horizon, but also on how many times the compounding calculations are computed each year.

Marginalism

Marginalism generally includes the study of marginal theories and relationships within economics. The key focus of marginalism is how much extra use is gained from incremental increases in the number of goods created, sold, etc. and how these measures relate to consumer choice and demand.

Marginalism covers such topics as marginal utility, marginal gain, marginal rates of substitution, and opportunity costs, within the context of consumers making rational choices in a market with known prices. These areas can all be thought of as popular schools of thought surrounding financial and economic incentives.

Understanding Marginalism

The idea of marginalism and its use in establishing market prices, as well as supply and demand patterns, was popularized by British economist Alfred Marshall in a publication dating back to 1890.
Marginalism is sometimes criticized as one of the "fuzzier" areas of economics, as much of what is proposed is hard to accurately measure, such as an individual consumers' marginal utility. Also, marginalism relies on the assumption of (near) perfect markets, which do not exist in the practical world. Still, the core ideas of marginalism are generally accepted by most economic schools of thought and are still used by businesses and consumers to make choices and substitute goods.

Modern marginalism approaches now include the effects of psychology or those areas that now encompass behavioral economics. Reconciling neoclassic economic principles and marginalism with the evolving body of behavioral economics is one of the exciting emerging areas of contemporary economics.

Examples of Marginalism

One of the key foundations of marginalism is the concept of marginal utility. The utility of a product or service is its usefulness in satisfying our needs. Marginal utility extends the concept to the additional satisfaction derived from the same product or service.

Marginal utility is used to explain the discrepancy between products that should be considered valuable but are not and products that are rare and expensive. For example, water is essential to human existence and, as such, should be considered more precious than a diamond. However, an average human being is willing to pay more for an additional diamond than a glass of water. The theory of marginal utility claims that this is so because we derive more satisfaction from owning an additional diamond than another glass of water.

Within the context of consumption, there is the law of diminishing marginal utility, which states that consumption is inversely proportional marginal utility. This means that as consumption increases, the marginal utility derived from a product or service declines. Thus, the satisfaction that a consumer derives from a new product is highest when he or she is first introduced to it. Subsequent use of the product or service diminishes the satisfaction derived from it.

Incrementalism

The incremental concept is probably the most important concept in economics and is certainly the most frequently used in Managerial Economics. Incremental concept is closely related to the marginal cost and marginal revenues of economic theory.

The two major concepts in this analysis are incremental cost and incremental revenue. Incremental cost denotes change in total cost, whereas incremental revenue means change in total revenue resulting from a decision of the firm.

The incremental principle may be stated as follows:

A decision is clearly a profitable one if

(i) It increases revenue more than costs.

(ii) It decreases some cost to a greater extent than it increases others.

(iii) It increases some revenues more than it decreases others.

(iv) It reduces costs more than revenues.

Illustration:

Some businessmen hold the view that to make an overall profit, they must make a profit on every job. The result is that they refuse orders that do not cover full costs plus a provision of profit. This will lead to rejection of an order which prevents short run profit. A simple problem will illustrate this point. Suppose a new order is estimated to bring in an additional revenue of Rs. 10,000. The costs are estimated as under:

Labour Rs. 3,000

Materials Rs. 4,000 Overhead charges Rs. 3,600

Selling and administrative expenses Rs. 1,400 Full Cost Rs.12, 000

The order appears to be unprofitable. For it results in a loss of Rs. 2,000. However, suppose there is idle capacity which can be utilised to execute this order. If order adds only Rs. 1,000 to overhead charges, and Rs. 2000 by way of labour cost because some of the idle workers already on the pay roll will be deployed without added pay and no extra selling and administrative costs, then the actual incremental cost is as follows:

Labour Rs. 2,000

Materials' Rs. 4,000 Overhead charges Rs. 1,000

Total Incremental Cost Rs. 7,000

Thus there is a profit of Rs. 3,000. The order can be accepted on the basis of incremental reasoning. Incremental reasoning does not mean that the firm should accept all orders at prices which cover merely their incremental costs.

The concept is mainly used by the progressive concerns. Even though it is a widely followed concept, it has certain limitations:

(a) The concept cannot be generalised because observed behaviour of the firm is always variable.

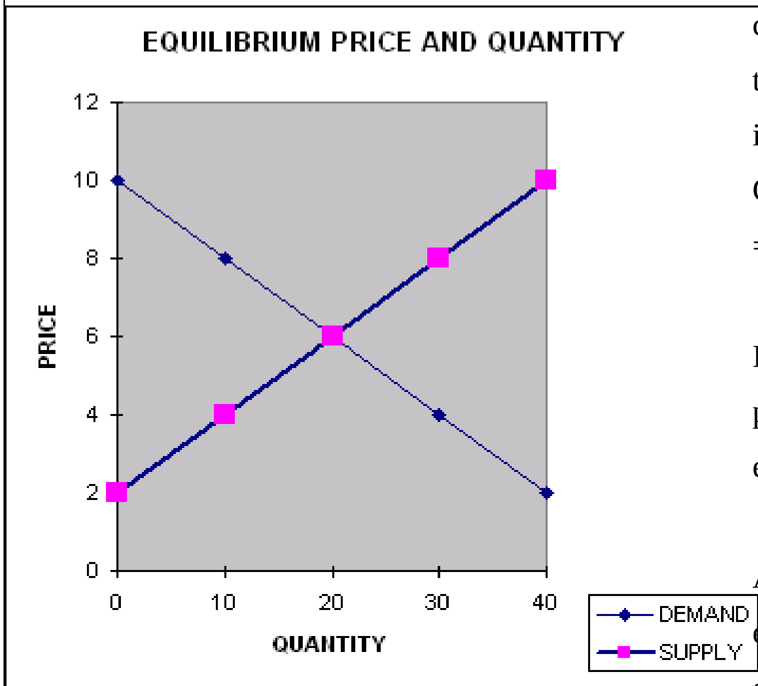
(b) The concept can be applied only when there is excess capacity in the concern.

(c) The concept is applicable only during the short period.

Market forces & Equilibrium

Market forces refer to supply and demand, which determine the allocation of scarce resources and the relative prices of goods, services, and assets in a market economy. A free market system is one in which there is no government intervention

When the supply and demand curves intersect, the market is in equilibrium. This is where the quantity demanded and quantity supplied are equal. The corresponding price is the equilibrium price or market-clearing price, the quantity is the equilibrium quantity.



Putting the supply and demand curves from the previous sections together. These two curves will intersect at Price = $6, and Quantity = 20.

In this market, the equilibrium price is $6 per unit, and equilibrium quantity is 20 units.

At this price level, market is in equilibrium. Quantity supplied is equal to quantity demanded (Qs = Qd).

Market is clear.

Surplus and shortage:

If the market price is above the equilibrium price, quantity supplied is greater than quantity demanded, creating a surplus. Market price will fall.

Example: if you are the producer, you have a lot of excess inventory that cannot sell. Will you put them on sale? It is most likely yes. Once you lower the price of your product, your product's quantity demanded will rise until equilibrium is reached. Therefore, surplus drives price down.

If the market price is below the equilibrium price, quantity supplied is less than quantity demanded, creating a shortage. The market is not clear. It is in shortage. Market price will rise because of this shortage.

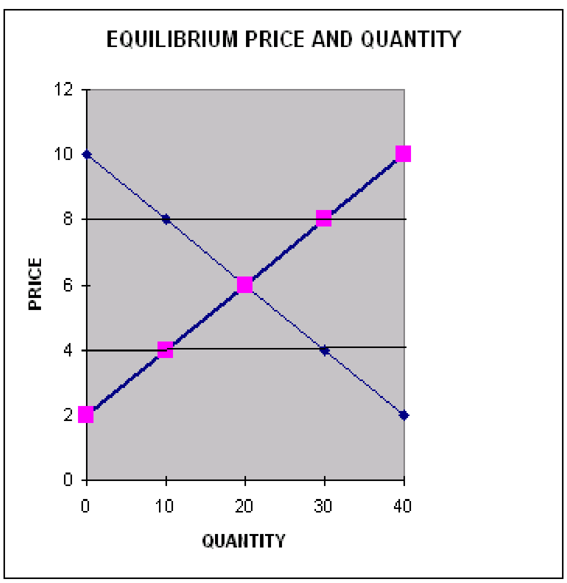
Example: if you are the producer, your product is always out of stock. Will you raise the price to make more profit? Most for-profit firms will say yes. Once you raise the price of your product, your product's quantity demanded will drop until equilibrium is reached. Therefore, shortage drives price up.

If a surplus exist, price must fall in order to entice additional quantity demanded and reduce quantity supplied until the surplus is eliminated. If a shortage exists, price must rise in order to entice additional supply and reduce quantity demanded until the shortage is eliminated.

Government regulations will create surpluses and shortages in the market. When a price ceiling is set, there will be a shortage. When there is a price floor, there will be a surplus.

Price Floor: is legally imposed minimum price on the market. Transactions below this price is prohibited.

•Policy makers set floor price above the market equilibrium price which they believed is too low.



• Price floors are most often placed on markets for goods that are an important source of income for the sellers, such as labor market. •Price floor generate surpluses on the market. •Example: minimum wage.

Price Ceiling: is legally imposed maximum price on the market. Transactions above this price is prohibited. •Policy makers set ceiling price below the market equilibrium price which they believed is too high. •Intention of price ceiling is keeping stuff affordable for poor people. •Price ceiling generates shortages on the market. •Example: Rent control.

Changes in equilibrium price and quantity:

Equilibrium price and quantity are determined by the intersection of supply and demand. A change in supply, or demand, or both, will necessarily change the equilibrium price, quantity or both. It is highly unlikely that the change in supply and demand perfectly offset one another so that equilibrium remains the same.

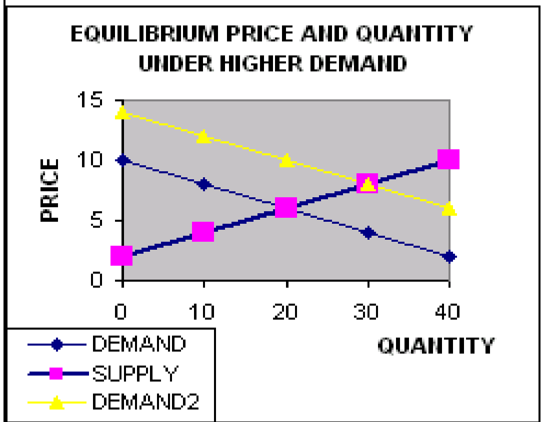
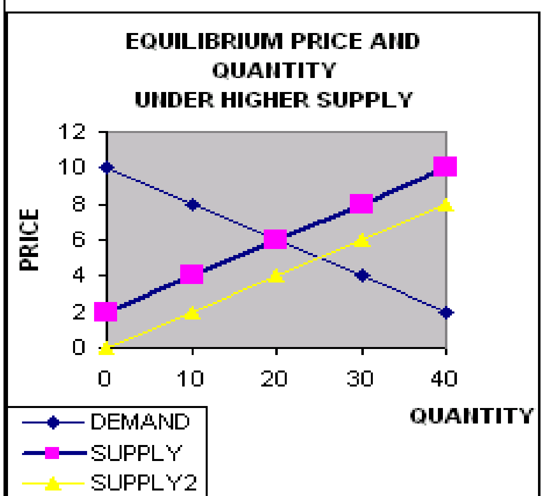
Example: This example is based on the assumption of Ceteris Paribus.

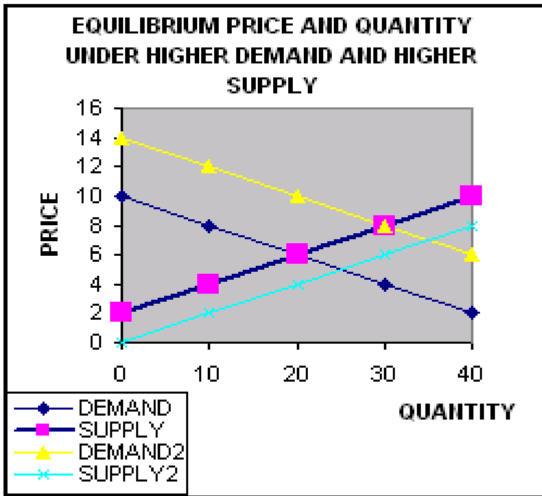
1) If there is an exporter who is willing to export oranges from Florida to Asia, he will increase the demand for Florida's oranges. An increase in demand will create a shortage, which increases the equilibrium price and equilibrium quantity.

2) If there is an importer who is willing to import oranges from Mexico to Florida, he will increase the supply for Florida's oranges. An increase in supply will create a surplus, which lowers the equilibrium price and increase the equilibrium quantity.

3) What will happen if the exporter and importer enter the Florida's orange market at the same time? From the above analysis, we can tell that equilibrium quantity will be higher. But the import and exporter's impact on price is opposite. Therefore, the change in equilibrium price cannot be determined unless more details are provided. Detail information should include the exact quantity the exporter and importer is engaged in. By comparing the quantity between importer and exporter, we can determine who has more impact on the market.

In the following table, an example of demand and supply increase is illustrated.

	In this graph, supply is constant, demand increases. As the new demand curve (Demand 2) has shown, the new curve is located on the right hand side of the original demand curve. The new curve intersects the original supply curve at a new point. At this point, the equilibrium price (market price) is higher, and equilibrium quantity is higher also.
	In this graph, demand is constant, and supply increases. As the new supply curve (SUPPLY 2) has shown, the new curve is located on the right side of the original supply curve. The new curve intersects the original demand curve at a new point. At this point, the equilibrium price (market price) is lower, and the equilibrium quantity is higher.
	In this graph, the increased demand curve and increased supply were drawn together. The new intersection point is located on the right hand side of the original intersection point. This new equilibrium point indicated an equilibrium quantity which is higher than the original equilibrium quantity. The equilibrium price is also

	higher. It is because demand has increased relatively more than supply in this case.

Risk, Return and Profits

The risk-return profits/tradeoff states that the potential return rises with an increase in risk. Using this principle, individuals associate low levels of uncertainty with low potential returns, and high levels of uncertainty or risk with high potential returns. According to the risk-return tradeoff, invested money can render higher profits only if the investor will accept a higher possibility of losses.

Understanding Risk-Return Tradeoff

The risk-return tradeoff is the trading principle that links high risk with high reward. The appropriate risk-return tradeoff depends on a variety of factors including an investor's risk tolerance, the investor's years to retirement and the potential to replace lost funds. Time also plays an essential role in determining a portfolio with the appropriate levels of risk and reward. For example, if an investor has the ability to invest in equities over the long term, that provides the investor with the potential to recover from the risks of bear markets and participate in bull markets, while if an investor can only invest in a short time frame, the same equities have a higher risk proposition.

Investors use the risk-return tradeoff as one of the essential components of each investment decision, as well as to assess their portfolios as a whole. At the portfolio level, the risk-return tradeoff can include assessments of the concentration or the diversity of holdings and whether the mix presents too much risk or a lower-than-desired potential for returns.

Special Considerations

Measuring Singular Risk in Context

When an investor considers high-risk-high-return investments, the investor can apply the risk- return tradeoff to the vehicle on a singular basis as well as within the context of the portfolio as a whole. Examples of high-risk-high return investments include options, penny stocks and leveraged exchange-traded funds (ETFs). Generally speaking, a diversified portfolio reduces the risks presented by individual investment positions. For example, a penny stock position may have a high risk on a singular basis, but if it is the only position of its kind in a larger portfolio, the risk incurred by holding the stock is minimal.

Risk-Return Tradeoff at the Portfolio Level

That said, the risk-return tradeoff also exists at the portfolio level. For example, a portfolio composed of all equities presents both higher risk and higher potential returns. Within an all- equity portfolio, risk and reward can be increased by concentrating investments in specific sectors or by taking on single positions that represent a large percentage of holdings. For investors, assessing the cumulative risk-return tradeoff of all positions can provide insight on whether a portfolio assumes enough risk to achieve long-term return objectives or if the risk levels are too high with the existing mix of holdings.

UNIT – 2

CONSUMER BEHAVIOUR AND DEMAND ANALYSIS

Cardinal Utility Approach:

Definition and Explanation:

Human wants are unlimited and they are of different intensity. The means at the disposal of a man are not only scarce but they have alternative uses. As a result of scarcity of recourses, the consumer cannot satisfy all his wants. He has to choose as to which want is to be satisfied first and which afterward if the recourses permit. The consumer is confronted in making a choice.

For example, a man is thirsty. He goes to the market and satisfy his thirst by purchasing coca cola instead of tea. We are here to examine the economic forces which make him purchase a particular commodity. The answer is simple. The consumer buys a commodity because it gives him satisfaction. In technical term, a consumer purchases a commodity because it has utility for him. We now examine the tools which are used in the analyzes of consumer behavior.

Concept of Utility:

Jevon (1835 -1882) was the first economist who introduces the concept of utility in economics. According to him:

"Utility is the basis on which the demand of a individual for a commodity depends upon". Utility is defined as:

"The power of a commodity or service to satisfy human want".

Utility is thus the satisfaction which is derived by the consumer by consuming the goods.

For example, cloth has a utility for us because we can wear it. Pen has a utility who can write with it. The utility is subjective in nature. It differs from person to person. The utility of a bottle of wine is zero for a person who is non drinker while it has a very high utility for a drinker.

Here it may be noted that the term ‘utility’ may not be confused with pleasure or unfulness which a commodity gives to an individual. Utility is a subjective satisfaction which consumer gets from consuming any good or service.

For example, poison is injurious to health but it gives subjective satisfaction to a person who wishes to die. We can say that utility is value neutral.

Assumptions of Cardinal Utility Analysis:

The main assumption or premises on which the cardinal utility analysis rests are as under.

(i) Rationality. The consumer is rational. He seeks to maximize satisfaction from the limited income which is at his disposal.

(ii) Utility is cardinally measurable. The utility can be measured in cardinal numbers such as 1, 3, 10, 15, etc. The utility is expressed in imaginary cardinal numbers tells us a great deal about the preference of the consumer for a good.

(iii) Marginal utility of money remains constant. Another important premise of cardinal utility of money spent on the purchase of a good or service should remain constant.

(iv) Diminishing marginal utility. It is also assumed that the marginal utility obtained from the consumption of a good diminishes continuously as its consumption is increased.

(v) Independent utilities. According to the Cardinalist school, the utility which is derived from the consumption of a good is a function of the quantity of that good alone. If does not depend at all upon the quantity consumed of other goods. The goods, we can say, possess independent utilities and are additive.

(vi) Introspection method. The Cardinalist school assumes that the behavior of marginal utility in the mind of another person can be judged with the help of self observation. For example, I know that as I purchase more and more of a good, the less utility I derived from the additional units of it. By applying the same principle, I can read other people mind and say with confidence that marginal utility of a good diminishes as they have more units of it.

Criticism:

Pareto, an Italian Economist, severely criticized the concept of cardinal utility. He stated that utility is neither quantifiable nor addible. It can, however be compared. He suggested that the concept of utility should be replaced by the scale of preference. Hicks and Allen, following the footsteps of Pareto, introduced the technique of indifference curves. The cardinal utility approach is thus replaced by ordinal utility function.

“During the course of consumption, as more and more units of a commodity are used, every successive unit gives utility with a diminishing rate, provided other things remaining the same; although, the total utility increases.”

Utils:

'Utils' is considered as the measurable 'unit' of utility.

Explanation for the Law of Diminishing Marginal Utility:

We can briefly explain Marshall’s theory with the help of an example. Assume that a consumer consumes 6 apples one after another. The first apple gives him 20 utils (units for measuring utility). When he consumes the second and third apple, the marginal utility of each additional apple will be lesser. This is because with an increase in the consumption of apples, his desire to consume more apples falls.

Therefore, this example proves the point that every successive unit of a commodity used gives the utility with the diminishing rate.

We can explain this more clearly with the help of a schedule and diagram.

Schedule for Law of Diminishing Marginal Utility:

In the above table, the total utility obtained from the first apple is 20 utils, which keep on increasing until we reach our saturation point at 5th apple. On the other hand, marginal utility keeps on diminishing with every additional apple consumed. When we consumed the 6th apple, we have gone over the limit. Hence, the marginal utility is negative and the total utility falls.

With the help of the schedule, we have made the following diagram:

Saturation Point: The point where the desire to consume the same product anymore becomes zero.

Disutility:

If you still consume the product after the saturation point, the total utility starts to fall. This is known as disutility.

When the first apple is consumed, the marginal utility is 20. When the second apple is consumed, the marginal utility increases by 15 utils, which is less than the marginal utility of the 1st apple – because of the diminishing rate. Therefore, we have shown that the utility of apples consumed diminishes with every increase of apple consumed.

Similarly, when we consumed the 5th apple, we are at our saturation point. If we consume another apple, i.e. 6th apple, we can see that the marginal utility curve has fallen to below X- axis, which is also known as 'disutility'.

Assumptions in the Law of Diminishing Marginal Utility:

For the law of diminishing marginal utility to be true, we need to make certain assumptions. Each assumption is quite logical and understandable. If any of the assumptions are not true in the case, the law of diminishing marginal utility will not be true.

Following are the assumptions in the law of diminishing marginal utility:

The quality of successive units of goods should remain the same. If the quality of the goods increase or decrease, the law of diminishing marginal utility may not be proven true. Consumption of goods should be continuous. If there comes a substantial break in the consumption of goods, the actual concept of diminishing marginal utility will be altered. Consumer's mental outlook should not change. Unit of good should not be very few or small. In such a case, the utility may not be measured accurately.

Exceptions for the Law of Diminishing Marginal Utility:

The law of diminishing marginal utility states that with the consumption of every successive unit of commodity yields marginal utility with a diminishing rate. However, there are certain things on which the law of diminishing marginal utility does not apply.

Law of Equi-Marginal Utility

The idea of equi-marginal principle was first mentioned by H.H.Gossen (1810-1858) of Germany. Hence it is called Gossen's second Law. Alfred Marshall made significant refinements of this law in his 'Principles of Economics'.

The law of equi-marginal utility explains the behaviour of a consumer when he consumers more than one commodity. Wants are unlimited but the income which is available to the consumers to satisfy all his wants is limited. This law explains how the consumer spends his limited income on various commodities to get maximum satisfaction. The law of equi-marginal utility is also known as the law of substitution or the law of maximum satisfaction or the principle of proportionality between prices and marginal utility.

Definition

In the words of Prof. Marshall, 'If a person has a thing which can be put to several uses, he will distribute it among these uses in such a way that it has the same marginal utility in all'.

Assumptions

1. The consumer is rational so he wants to get maximum satisfaction.
2. The utility of each commodity is measurable.
3. The marginal utility of money remains constant.
4. The income of the consumer is given.
5. The prices of the commodities are given.
6. The law is based on the law of diminishing marginal utility.

Explanation of the law

Suppose there are two goods X and Y on which a consumer has to spend a given income. The consumer being rational, he will try to spend his limited income on goods X and Y to maximise his total utility or satisfaction. Only at that point the consumer will be in equilibrium.

According to the law of equi-marginal utility, the consumer will be in equilibrium at the point where the utility derived from the last rupee spent on each is equal.

Limitations of the Law

The law of equi-marginal utility bristles with the following difficulties.

1. Indivisibility of Goods

The theory is weakened by the fact that many commodities like a car, a house etc. are indivisible. In the case of indivisible goods, the law is not applicable.

2. The Marginal Utility of Money is Not Constant

The theory is based on the assumption that the marginal utility of money is constant. But that is not really so.

3. The Measurement of Utility is not Possible

Marshall states that the price a consumer is willing to pay for a commodity is equal to its marginal utility. But modern economists argue that, if two persons are paying an equal price for given commodity, it does not mean that both are getting the same level of utility. Thus utility is a subjective concept, which cannot be measured, in quantitative terms.

4. Utilities are Interdependent

This law assumes that commodities are independent and therefore their marginal utilities are also independent. But in real life commodities are either substitutes or complements. Their utilities are therefore interdependent.

5. Indefinite Budget Period

According to Prof. K.E. Boulding, indefinite budget period is another difficulty in the law. Normally the budget period is assumed to be a year. But there are certain commodities which are available in several succeeding accounting periods. It is difficult to calculate marginal utility for such commodities.

In conclusion, we may say all prudent and rational persons are expected to act upon the law consciously or unconsciously. As Chapman puts it,

'We are not, of course compelled to distribute our incomes according to the law of substitution or equi-marginal expenditure, as a stone thrown into the air is compelled, in a sense to fall back to the earth, but as a matter of fact, we do in a certain rough fashion, because we are reasonable.'

Importance

According to Marshall, 'the applications of this principle extend over almost every field of economic activity.'

1. It applies to consumption

Every rational human being wants to get maximum satisfaction with his limited means. The consumer arranges his expenditure in such a way that, $MU_x/P_x = MU_y/P_y = MU_z/P_z$ so that he will get maximum satisfaction.

2. It applies to production

The aim of the producer is to get maximum output with least-cost, so that his profit will be maximum. Towards this end, he will substitute one factor for another till

$MP_l / P_l = MP_c / P_c = MP_n / P_n$

3. Distribution of Earnings Between Savings and Consumption

According to Marshall, a prudent person will endeavour to distribute his resources between his present needs and future needs in such a way that the marginal utility of the last rupee put in savings is equal to the marginal utility of the last rupee spent on consumption.

4. It applies to distribution

The general theory of distribution involves the principle of substitution. In distribution, the rewards to the various factors of production, that is their relative shares, are determined by the principle of equi-marginal utility.

5. It Applies to Public Finance

The principle of 'Maximum Social Advantage' as enunciated by Professors Hicks and Dalton states that, the revenue should be distributed in such a way that the last unit of expenditure on various programmes brings equal welfare, so that social welfare is maximised.

6. Expenditure of Time

Prof. Boulding relates Marshall's law of equi-marginal utility to the expenditures of limited time, i.e. twenty-four hours. He states that a person should spend his limited time among alternative uses such as reading; studying and gardening, in such a way that the marginal utility from all these uses are equal.

Ordinal Utility Approach

The Ordinal Utility approach is based on the fact that the utility of a commodity cannot be measured in absolute quantity, but however, it will be possible for a consumer to tell subjectively whether the commodity derives more or less or equal satisfaction when compared to another.

The modern economists have discarded the concept of cardinal utility and instead applied ordinal utility approach to study the behavior of the consumers. While the neo-classical economists believed that the utility can be measured and expressed in cardinal numbers, but the modern economists maintain that the utility being the psychological phenomena cannot be measured theoretically, quantitatively and even cardinally.

The modern economist, Hicks, in particular, have applied the ordinal utility concept to study the consumer behavior. He introduced a tool of analysis called “Indifference Curve” to analyze the consumer behavior. An indifference curve refers to the locus of points each showing different combinations of two substitutes which yield the same level of satisfaction and utility to the consumer.

Assumptions of Ordinal Utility Approach

1. Rationality: It is assumed that the consumer is rational who aims at maximizing his level of satisfaction for given income and prices of goods and services, which he wish to consume. He is expected to take decisions consistent with this objective.
2. Ordinal Utility: The indifference curve assumes that the utility can only be expressed ordinally. This means the consumer can only tell his order of preference for the given goods and services.
3. Transitivity and Consistency of Choice: The consumer’s choice is expected to be

either transitive or consistent. The transitivity of choice means, if the consumer prefers commodity X to Y and Y to Z, then he must prefer commodity X to Z. In other words, if X= Y, Y = Z, then he must treat X=Z. The consistency of choice means that if a consumer prefers commodity X to Y at one point of time, he will not prefer commodity Y to X in another period or even will not consider them as equal.

4. Nonsatiety: It is assumed that the consumer has not reached the saturation point of any commodity and hence, he prefers larger quantities of all commodities.

5. Diminishing Marginal Rate of Substitution (MRS): The marginal rate of substitution refers to the rate at which the consumer is ready to substitute one commodity (A) for another commodity (B) in such a way that his total satisfaction remains unchanged. The MRS is denoted as DB/DA. The ordinal approach assumes that DB/DA goes on diminishing if the consumer continues to substitute A for B.

Indifference Curve

A popular alternative to the marginal utility analysis of demand is the Indifference Curve Analysis. This is based on consumer preference and believes that we cannot quantitatively measure human satisfaction in monetary terms. This approach assigns an order to consumer preferences rather than measure them in terms of money.

What is an Indifference Curve?

It is a curve that represents all the combinations of goods that give the same satisfaction to the consumer. Since all the combinations give the same amount of satisfaction, the consumer prefers them equally. Hence the name Indifference Curve.

Here is an example to understand the indifference curve better. Peter has 1 unit of food and 12 units of clothing. Now, we ask Peter how many units of clothing is he willing to give up in exchange for an additional unit of food so that his level of satisfaction remains unchanged.

Peter agrees to give up 6 units of clothing for an additional unit of food. Hence, we have two combinations of food and clothing giving equal satisfaction to Peter as follows:

1. 1 unit of food and 12 units of clothing
2. 2 units of food and 6 units of clothing

Graphical Representation:

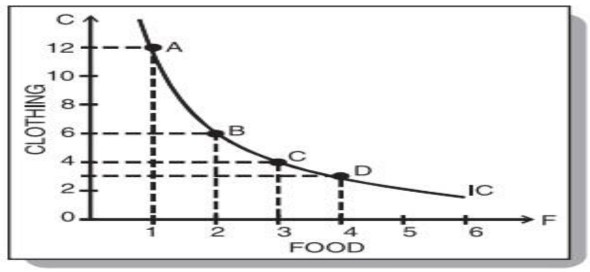


Fig. 1 : A Consumer's Indifference Curve

The diagram shows an Indifference curve (IC). Any combination lying on this curve gives the same level of consumer satisfaction. It is also known as Iso-Utility Curve.

The Assumptions are as follows,

- The consumer is rational. Also, he possesses full information about all the relevant aspects of the economic environment in which he lives.
- The consumer can rank combination of goods based on the satisfaction they yield. However, he can’t quantitatively express how much he prefers a certain good over the other.
- If a consumer prefers A over B and B over C, then he prefers A over C.
- If a combination X has more commodities than the combination Y, then X is preferred over Y.

Indifference Map

An Indifference Map is a set of Indifference Curves. It depicts the complete picture of a consumer’s preferences. The following diagram showing an indifference map consisting of three curves:

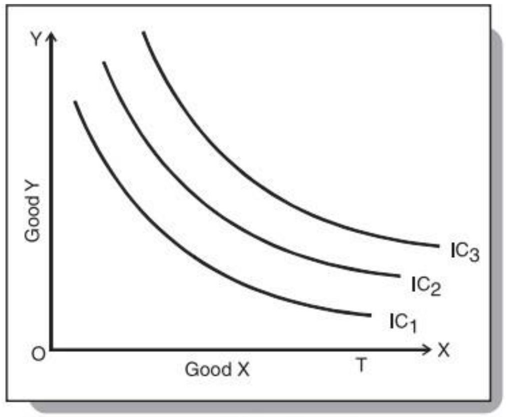


Fig. 2 : Indifference Map

We know that a consumer is indifferent among the combinations lying on the same indifference curve. However, it is important to note that he prefers the combinations on the higher indifference curves to those on the lower ones.

This is because a higher indifference curve implies a higher level of satisfaction. Therefore, all combinations on IC1 offer the same satisfaction, but all combinations on IC2 give greater satisfaction than those on IC1.

Marginal Rate of Substitution

An indifference curve is formed when one good is substituted for other. Marginal Rate of Substitution (MRS) refers to a rate at which one good is substituted for other, while keeping the level of satisfaction of a consumer constant. In other words, MRS between two goods X and Y is defined as the quantity of X which is required to replace Y or quantity of Y required to replace X, so that the total utility remains same.

It is expressed as:

MRS x,y = $\Delta Y/\Delta X$

MRS is called the slope of indifference curve.

Let us discuss the concept of MRS through a schedule and graph.

Table-4 shows the schedule of MRS between goods X and Y:

Table-4: MRS Schedule			
Combination	Good X	Good Y	MRS x,y
A	1	12	4
B	2	8	3
C	3	5	2
D	4	3	1
E	5	2	-

In Table-4, it can be seen the MRS is diminishing. This is because the basic assumption of the ordinal utility concept is that MRS diminishes.' This implies that a consumer sacrifices some unit of a good X or Y when substituting X for Y or Y for X.

The diminishing MRS x,y attained from the combination of good X and Y is shown in Figure-8:

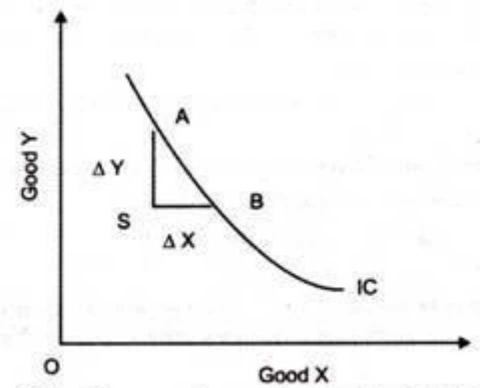


Figure-8: Indifference Curve showing Diminishing MRS

From Figure-8, it can be seen that the consumer is ready to sacrifice AS of Y to gain SB of X. Thus, MRS = AS/SB= Δ Y/Δ X. It can be noted from schedule (Table-4) and graph (Figure-8) that MRS falls as more and more of good X is consumed. The reason for convex shape of indifference curve is diminishing MRS. The MRS diminishes because of the two main reasons. Firstly, in indifference curve, the combination of two goods is such that quantity of one good is more than the quantity of another good. Thus, consumer is ready to sacrifice more of good whose quantity is large. When the consumer's stock of good Y is large than good X, he/she would be willing to give up large amount of good Y to gain more of X and vice versa; therefore, MRS falls.

MRS diminishes because two goods X and Y are not perfect substitutes of each other. If they are perfect substitutes, then indifference curve would be a straight line showing constant MRS. The increase in quantity of one and decrease in quantity of other would not make any difference in the marginal significance of goods. However, in case of imperfect substitutes, the consumer is required to sacrifice additional units of Y to gain increasing units of X to maintain the level of satisfaction. Thus, MRS decreases.

Properties of an Indifference Curve or IC

Here are the properties of an indifference curve:

An IC slopes downwards to the right

This slope signifies that when the quantity of one commodity in combination is increased, the amount of the other commodity reduces. This is essential for the level of satisfaction to remain the same on an indifference curve.

An IC is always convex to the origin

From our discussion above, we understand that as Peter substitutes clothing for food, he is willing to part with less and less of clothing. This is the diminishing marginal rate of substitution. The rate gives a convex shape to the indifference curve. However, there are two extreme scenarios:

1. Two commodities are perfect substitutes for each other – In this case, the indifference curve is a straight line, where MRS is constant.
2. Two goods are perfect complementary goods – An example of such goods would be gasoline and water in a car. In such cases, the IC will be L-shaped and convex to the origin.

Indifference curves never intersect each other

Two ICs will never intersect each other. Also, they need not be parallel to each other either. Look at the following diagram:

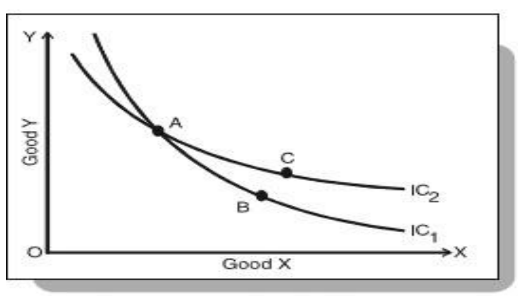


Fig. 3 : Intersecting Indifference Curves

Fig 3 shows tow ICs intersecting each other at point A. Since A and B lie on IC1, the give the same satisfaction level. Similarly, A and C give the same satisfaction level, as they lie on IC2. Therefore, we can imply that B and C offer the same level of satisfaction, which is logically absurd. Hence, no tow ICs can touch or intersect each other.

A higher IC indicates a higher level of satisfaction as compared to a lower IC

A higher IC means that a consumer prefers more goods than not.

An IC does not touch the axis

This is not possible because of our assumption that a consumer considers different combinations of two commodities and wants both of them. If the curve touches either of the axes, then it means that he is satisfied with only one commodity and does not want the other, which is contrary to our assumption.

Budget Line

A consumer prefers to reach the highest possible indifference curve on indifference map to attain satisfaction. However, he/she suffers from two constraints, namely, limited income and price of goods. The lack of income is called budgetary constraint.

The budget equation is expressed as:

$Px.Qx + Py.Qy = M$

Px and Py are the prices of goods X and Y

Qx and Qy are the quantities of goods X and Y

M= money income of the consumer

The budget equation states that the total expenditure cannot exceed the total income.

The quantities can be derived as:

$Qx = M/Px-Py/Px.\ Qy$ $Qy = M/Py-Px/Py.Qx$

When these quantities are plotted on a graph, a budget line is obtained, which is also called price line. The indifference curve shows preferences of combination of two goods where the actual choice of preferences depends on income.

Budget line is the combination of two goods that can be purchased with a given money income and prices of goods. The consumer behavior is well depicted by the budget line. The budget line is drawn as a continuous line that identifies alternatives from which a consumer selects an appropriate combination of goods.

Suppose the income of a consumer is Rs. 60. He/she wants to consume goods X and Y. The price of good X is Rs. 12 and the price of good Y is Rs. 6.

The various preferences of X and Y that can be purchased from the given income and prices are shown in

Table-5: Preferences of X and Y (Given Income and Prices)

Combination	Quantity of X	Quantity of Y
A	0	10
B	1	8
C	2	6
D	3	4
E	4	2
F	5	0

Table-5:

From Table-5, it can be seen in combination A the whole amount of Rs. 60 is spent on purchase of quantity Y. In such a ca.se, the consumer buys 10 units of good Y and nothing is Left for purchasing good X. Similarly, combination F shows that if the consumer spends the entire amount on good X, then he/she is able to purchase 5 units of good X and nothing of good Y. The combinations B to E show the combined quantities of good X and good Y. For instance, in combination B, the consumer would purchase 8 units of good Y and 1 unit of good X.

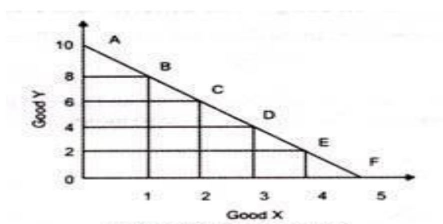


Figure-16: Budget Line

With the help of preferences, the budget line is drawn, which is shown in Figure-16:

In Figure-16, budget line AF shows various combinations of good X and Y that a consumer can purchase from his or her given budget. The combinations are shown by points A, B, C, D, E, and F that can be purchased with the given budget. If a consumer buys the combination of goods inside the budget line AF, then the total expenditure comes out to be less than the given budget.

Slope of Budget Line:

The slope of a budget line shows how many units of Y are sacrificed to get more units of X.

It can be expressed as:

Slope of the Budget Line = $\Delta Y/\Delta X$

For instance, from Figure-16, it can be seen that the slope at point B equals to 1 unit of X for 2 units of Y. It implies that 2 units of good X. It should be noted that the slope of budget line is negative. For instance, slope is -2 in case of preference B.

The slope of the budget line (Figure-16) would be given as:

$\Delta Qy/\Delta Qx = OA/OF$ When $X = 0$, $OA = M/Py$ When $Y = 0$, $OF = M/Px$ Thus, $OA/OF = Px/Py$

Therefore, it can be said that the slope of the budget line equals to the price ratio of two goods.

Shifts in Budget Line:

The budget line is determined by the income level of consumers and prices of goods in the market. The budget line shifts if there is a change in the income and prices. Let us take the both cases one by one.

Case 1: Change in Income:

Suppose there is a charge in the income of the consumer and the price of the goods remain same. The budget line would shift from the original position. If there is a rise in income, the budget line would shift upward to the right. On the other hand if there is a fell in income, the budget line would shift downward to the left.

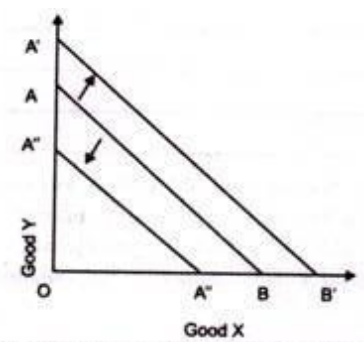


Figure-17: Effect on Budget Line when Income Changes

Figure-17 shows a shift in budget line due to change in income:

From Figure-17, it can be seen that a rise in income shifts the budget line from AB to A’B’ and a fall in income shifts the budget line from AB to A”B”.

Case 2: Change in Prices:

Suppose there is a change in the price of a good, say good X, and the income and price of good Y are constant. If there is a fall in the price of good X, then the consumer can buy more of good X with the same income.

Figure-18 shows the shift in the budget line due to change in the price of good X:

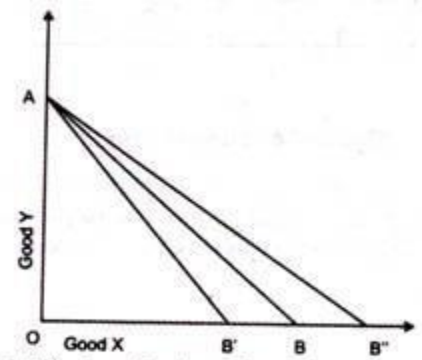


Figure-18: Effect on Budget Line when Price Changes

As shown in Figure-18, when the price of good X falls, this would make the budget line too move flatter and rightward from "AB to AB". In case, the price of good X increases, the budget line shifts to the left that is from AB to AB'.

Consumer's Equilibrium through Indifference Curve and Budget Line:

Consumer's equilibrium is the point at which consumer attains maximum satisfaction. A consumer is said to be in equilibrium when the budget line touches indifference curve, with given price and income.

The consumer's equilibrium through indifference curve and budget line is shown in Figure-19:

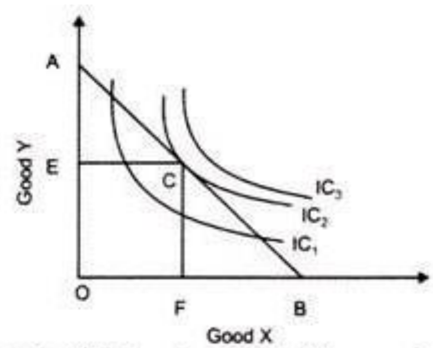


Figure-19: Consumer's Equilibrium through Indifference Curve and Budget Line

In Figure-19, there are three indifference curves IC1, IC2 and IC3. The budget line AB is tangent to IC2 at point C. At this level, a consumer attains maximum satisfaction level at OE units of good Y and OF units of good X. This is the first condition for the consumer to be in equilibrium that indifference curve should touch the budget line. The second condition is that the slope of budget line should be equal to the slope of indifference curve.

Slope of budget line =Px/Py

Slope of indifference curve= $\Delta Y/\Delta X$ = MRSxy Thus, Px/Py = MRSxy

Theory of Demand

Demand theory is an economic principle relating to the relationship between consumer demand for goods and services and their prices in the market. Demand theory forms the basis for the demand curve, which relates consumer desire to the amount of goods available. As more of a good or service is available, demand drops and so does the equilibrium price.

Demand theory highlights the role that demand plays in price formation, while supply-side theory favors the role of supply in the market.

Understanding Demand Theory

Demand is simply the quantity of a good or service that consumers are willing and able to buy at a given price in a given time period. People demand goods and services in an economy to satisfy their wants, such as food, healthcare, clothing, entertainment, shelter, etc. The demand for a product at a certain price reflects the satisfaction that an individual expects from consuming the product. This level of satisfaction is referred to as utility and it differs from consumer to consumer. The demand for a good or service depends on two factors: (1) its utility to satisfy a want or need, and (2) the consumer's ability to pay for the good or service. In effect, real demand is when the readiness to satisfy a want is backed up by the individual's ability and willingness to pay.

Demand theory is one of the core theories of microeconomics. It aims to answer basic questions about how badly people want things, and how demand is impacted by income levels and satisfaction (utility). Based on the perceived utility of goods and services by consumers, companies adjust the supply available and the prices charged.

Built into demand are factors such as consumer preferences, tastes, choices, etc. Evaluating demand in an economy is, therefore, one of the most important decision-making variables that a business must analyze if it is to survive and grow in a competitive market. The market system is governed by the laws of supply and demand, which determine the prices of goods and services. When supply equals demand, prices are said to be in a state of equilibrium. When demand is higher than supply, prices increase to reflect scarcity. Conversely, when demand is lower than supply, prices fall due to the surplus.

Law of Demand

Introduction to the Law of Demand:

The law of demand expresses a relationship between the quantity demanded and its price. It may be defined in Marshall's words as "the amount demanded increases with a fall in price, and diminishes with a rise in price". Thus it expresses an inverse relation between price and demand. The law refers to the direction in which quantity demanded changes with a change in price.

On the figure, it is represented by the slope of the demand curve which is normally negative throughout its length. The inverse price- demand relationship is based on other things remaining equal. This phrase points towards certain important assumptions on which this law is based.

Assumptions of the Law of Demand:

These assumptions are:

(i) There is no change in the tastes and preferences of the consumer;

(ii) The income of the consumer remains constant;

(iii) There is no change in customs;

(iv) The commodity to be used should not confer distinction on the consumer;

(v) There should not be any substitutes of the commodity;

(vi) There should not be any change in the prices of other products;

(vii) There should not be any possibility of change in the price of the product being used;

(viii) There should not be any change in the quality of the product; and

(ix) The habits of the consumers should remain unchanged. Given these conditions, the law of demand operates. If there is change even in one of these conditions, it will stop operating.

Given these assumptions, the law of demand is explained in terms of Table 3 and Figure 7.

Table 3.
Demand Schedule

Price (Rs)	Quantity Demanded
5	100 Units
4	200 Units
3	300 Units
2	400 Units
1	600 Units

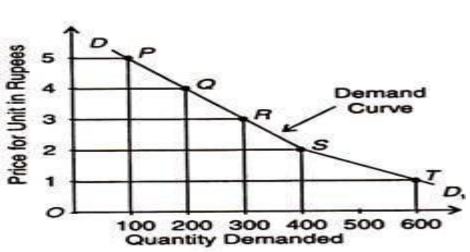


Fig. 7

The above table shows that when the price of say, orange, is Rs. 5 per unit, 100 units are de- manded. If the price falls to Rs.4, the demand increases to 200 units. Similarly, when the price declines to Re.1, the demand increases to 600 units. On the contrary, as the price increases from Re. 1, the demand continues to decline from 600 units.

In the figure, point P of the demand curve DD1 shows demand for 100 units at the Rs. 5. As the price falls to Rs. 4, Rs. 3, Rs. 2 and Re. 1, the demand rises to 200, 300, 400 and 600 units respectively. This is clear from points Q, R, S, and T. Thus, the demand curve DD1 shows increase in demand of orange when its price falls. This indicates the inverse relation between price and demand.

Exceptions to the Law of Demand:

In certain cases, the demand curve slopes up from left to right, i.e., it has a positive slope. Under certain circumstances, consumers buy more when the price of a commodity rises, and less when price falls, as shown by the D curve in Figure 8. Many causes are attributed to an upward sloping demand curve.

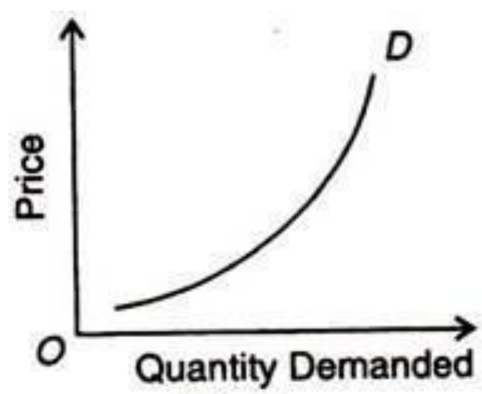


Fig. 8

(i) War:

If shortage is feared in anticipation of war, people may start buying for building stocks or for hoarding even when the price rises.

(ii) Depression:

During a depression, the prices of commodities are very low and the demand for them is also less. This is because of the lack of purchasing power with consumers.

(iii) Giffen Paradox:

If a commodity happens to be a necessity of life like wheat and its price goes up, consumers are forced to curtail the consumption of more expensive foods like meat and fish, and wheat being still the cheapest food they will consume more of it. The Marshallian example is applicable to developed economies.

In the case of an underdeveloped economy, with the fall in the price of an inferior commodity like maize, consumers will start consuming more of the superior commodity like wheat. As a result, the demand for maize will fall. This is what Marshall called the Giffen Paradox which makes the demand curve to have a positive slope.

(iv) Demonstration Effect:

If consumers are affected by the principle of conspicuous consumption or demonstration effect, they will like to buy more of those commodities which confer distinction on the possessor, when their prices rise. On the other hand, with the fall in the prices of such articles, their demand falls, as is the case with diamonds.

(v) Ignorance Effect:

Consumers buy more at a higher price under the influence of the "ignorance effect", where a commodity may be mistaken for some other commodity, due to deceptive packing, label, etc.

(vi) Speculation:

Marshall mentions speculation as one of the important exceptions to the downward sloping demand curve. According to him, the law of demand does not apply to the demand in a campaign between groups of speculators. When a group unloads a great quantity of a thing on to the market, the price falls and the other group begins buying it. When it has raised the price of the thing, it arranges to sell a great deal quietly. Thus when price rises, demand also increases.

(vii) Necessities of Life:

Normally, the law of demand does not apply on necessities of life such as food, cloth etc. Even the price of these goods increases, the consumer does not reduce their demand. Rather, he purchases them even the prices of these goods increase often by reducing the demand for comfortable goods. This is also a reason that the demand curve slopes upwards to the right.

Determinants of Demand

When price changes, quantity demanded will change. That is a movement along the same demand curve. When factors other than price changes, demand curve will shift. These are the determinants of the demand curve.

1. Income: A rise in a person's income will lead to an increase in demand (shift demand curve to the right), a fall will lead to a decrease in demand for normal goods. Goods whose demand varies inversely with income are called inferior goods (e.g. Hamburger Helper).

2. Consumer Preferences: Favorable change leads to an increase in demand, unfavorable change lead to a decrease.

3. Number of Buyers: the more buyers lead to an increase in demand; fewer buyers lead to decrease.

4. Price of related goods:

a. Substitute goods (those that can be used to replace each other): price of substitute and demand for the other good are directly related.

Example: If the price of coffee rises, the demand for tea should increase.

b. Complement goods (those that can be used together): price of complement and demand for the other good are inversely related.

Example: if the price of ice cream rises, the demand for ice-cream toppings will decrease.

5. Expectation of future:

a. Future price: consumers' current demand will increase if they expect higher future prices; their demand will decrease if they expect lower future prices.

b. Future income: consumers' current demand will increase if they expect higher future income; their demand will decrease if they expect lower future income.

Movement along the Demand Curve and Shift of the Demand Curve

While understanding the meaning and analysis of a demand curve in the study of Economics, it is also important to be able to make a distinction between the movement and shift of the demand curve. In this article, we will look at ways by which you can understand the difference between a movement along a demand curve and shift of the demand curve.

Every firm faces a certain demand curve for the goods it supplies. There are many factors that affect the demand and these effects can be seen by observing the changes in the demand curve. Broadly speaking, the factors can be categorized into two types:

- Change in demand
- Change in the quantity demanded Movement of the Demand Curve

When there is a change in the quantity demanded of a particular commodity, because of a change in price, with other factors remaining constant, there is a movement of the quantity demanded along the same curve.

The important aspect to remember is that other factors like the consumer's income and tastes along with the prices of other goods, etc. remain constant and only the price of the commodity changes.

In such a scenario, the change in price affects the quantity demanded but the demand follows the same curve as before the price changes. This is Movement of the Demand Curve. The movement can occur either in an upward or downward direction along the demand curve.

We know that if all other factors remain constant, then an increase in the price of a commodity decreases its demand. Also, a decrease in the price increases the demand. So, what hapen to the demand curve?

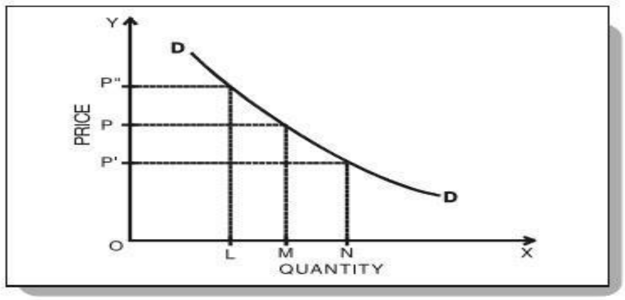


Fig. 1

In Fig. 1 above, we can see that when the price of a commodity is OP, its demand is OM (provided other factors are constant). Now, let's look at the effect of an increase and decrease in price on the demand:

- When the price increases from OP to OP", the quantity demanded falls to OL. Also, the demand curve moves UPWARD.
- When the price decreases from OP to OP', the quantity demanded rises to ON. Also, the demand curve moves DOWNWARD.

Therefore, we can see that a change in price, with other factors remaining constant moves the demand curve either up or down.

The shift of the Demand Curve

When there is a change in the quantity demanded of a particular commodity, at each possible price, due to a change in one or more other factors, the demand curve shifts. The important aspect to remember is that other factors like the consumer's income and tastes along with the prices of other goods, etc., which were expected to remain constant, changed.

In such a scenario, the change in price, along with a change in one/more other factors, affects the quantity demanded. Therefore, the demand follows a different curve for every price change.

This is the Shift of the Demand Curve. The demand curve can shift either to the left or the right, depending on the factors affecting it.

Let's look at an example which captures the effect of a change in consumer's income on the quantity demanded.

The demanded quantities are plotted as demand curves DD and D'D' as shown below:

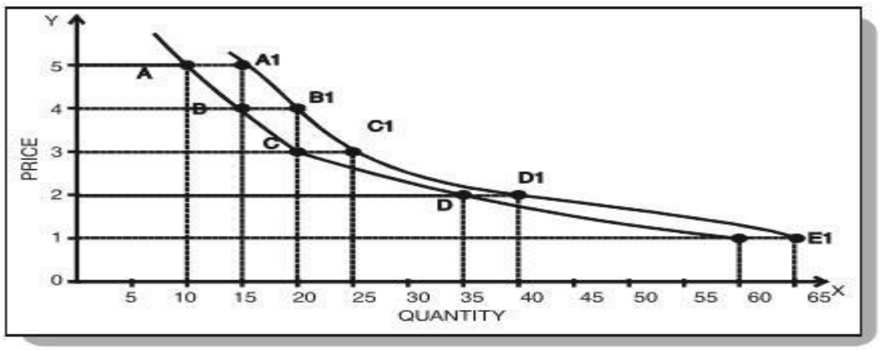


Fig. 2 : Figure showing two demand curves with different incomes

From Fig. 2 above, we can clearly see that if the income changes, then a change in price shifts the demand curve. In this case, the shift is to the right which indicates that there is an increase in the desire to purchase the commodity at all prices.

Hence, we can conclude that with an increase in income the demand curve shifts to the right. On the other hand, if the income falls, then the demand curve will shift to the left decreasing the desire to purchase the

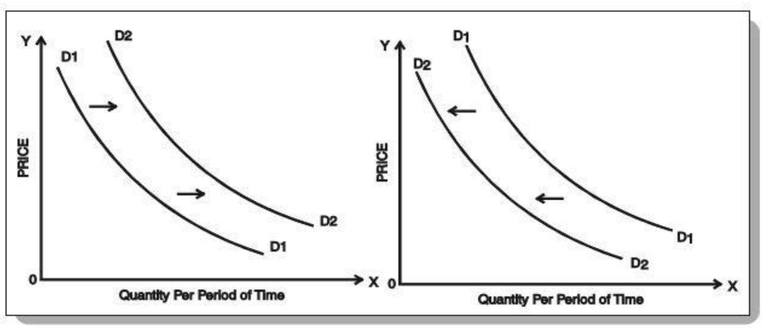


Fig. 3(a) : Rightward shift in the demand Curve

Fig. 3(b) : Leftward shift in the demand curve.

commodity.

Elasticity of Demand

A change in the price of a commodity affects its demand. We can find the elasticity of demand, or the degree of responsiveness of demand by comparing the percentage price changes with the quantities demanded. In this article, we will look at the concept of elasticity of demand and take a quick look at its various types.

“Elasticity of demand is the responsiveness of the quantity demanded of a commodity to changes in one of the variables on which demand depends. In other words, it is the percentage change in quantity demanded divided by the percentage in one of the variables on which demand depends.”

The variables on which demand can depend on are:

- Price of the commodity
- Prices of related commodities
- Consumer’s income, etc. Let’s look at some examples:

a. The price of a radio falls from Rs. 500 to Rs. 400 per unit. As a result, the demand increases from 100 to 150 units.

b. Due to government subsidy, the price of wheat falls from Rs. 10/kg to Rs. 9/kg. Due to this, the demand increases from 500 kilograms to 520 kilograms.

In both cases above, you can notice that as the price decreases, the demand increases. Hence, the demand for radios and wheat responds to price changes.

Types of Elasticity of Demand

Based on the variable that affects the demand, the elasticity of demand is of the following types. One point to note is that unless otherwise mentioned, whenever the elasticity of demand is mentioned, it implies price elasticity.

Price Elasticity

The price elasticity of demand is the response of the quantity demanded to change in the price of a commodity. It is assumed that the consumer’s income, tastes, and prices of all other goods are steady. It is measured as a percentage change in the quantity demanded divided by the percentage change in price. Therefore,

$$\text{Price Elasticity} = E_p = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}$$

Or,

Where,

Ec

is the cross elasticity,

Δqx

is the original demand of commodity X,

Δqx

is the change in demand of X,

Δpy

is the original price of commodity Y, and Δpy

is the change in price of Y.

Factors affecting Elasticity of Demand

A change in price does not always lead to the same proportionate change in demand. For example, a small change in price of AC may affect its demand to a considerable extent/whereas, large change in price of salt may not affect its demand. So, elasticity of demand is different for different goods.

Various factors which affect the elasticity of demand of a commodity are:

1. Nature of commodity:

Elasticity of demand of a commodity is influenced by its nature. A commodity for a person may be a necessity, a comfort or a luxury.

i. When a commodity is a necessity like food grains, vegetables, medicines, etc., its demand is generally inelastic as it is required for human survival and its demand does not fluctuate much with change in price.

ii. When a commodity is a comfort like fan, refrigerator, etc., its demand is generally elastic as consumer can postpone its consumption.

iii. When a commodity is a luxury like AC, DVD player, etc., its demand is generally more elastic as compared to demand for comforts.

iv. The term 'luxury' is a relative term as any item (like AC), may be a luxury for a poor person but a necessity for a rich person.

2. Availability of substitutes:

Demand for a commodity with large number of substitutes will be more elastic. The reason is that even a small rise in its prices will induce the buyers to go for its substitutes. For example, a rise in the price of Pepsi encourages buyers to buy Coke and vice-versa.

Thus, availability of close substitutes makes the demand sensitive to change in the prices. On the other hand, commodities with few or no substitutes like wheat and salt have less price elasticity of demand.

3. Income Level:

Elasticity of demand for any commodity is generally less for higher income level groups in comparison to people with low incomes. It happens because rich people are not influenced much by changes in the price of goods. But, poor people are highly affected by increase or decrease in the price of goods. As a result, demand for lower income group is highly elastic.

4. Level of price:

Level of price also affects the price elasticity of demand. Costly goods like laptop, Plasma TV, etc. have highly elastic demand as their demand is very sensitive to changes in their prices. However, demand for inexpensive goods like needle, match box, etc. is inelastic as change in prices of such goods do not change their demand by a considerable amount.

5. Postponement of Consumption:

Commodities like biscuits, soft drinks, etc. whose demand is not urgent, have highly elastic demand as their consumption can be postponed in case of an increase in their prices. However,

commodities with urgent demand like life saving drugs, have inelastic demand because of their immediate requirement.

6. Number of Uses:

If the commodity under consideration has several uses, then its demand will be elastic. When price of such a commodity increases, then it is generally put to only more urgent uses and, as a result, its demand falls. When the prices fall, then it is used for satisfying even less urgent needs and demand rises.

For example, electricity is a multiple-use commodity. Fall in its price will result in substantial increase in its demand, particularly in those uses (like AC, Heat convector, etc.), where it was not employed formerly due to its high price. On the other hand, a commodity with no or few alternative uses has less elastic demand.

7. Share in Total Expenditure:

Proportion of consumer's income that is spent on a particular commodity also influences the elasticity of demand for it. Greater the proportion of income spent on the commodity, more is the elasticity of demand for it and vice-versa.

Demand for goods like salt, needle, soap, match box, etc. tends to be inelastic as consumers spend a small proportion of their income on such goods. When prices of such goods change, consumers continue to purchase almost the same quantity of these goods. However, if the proportion of income spent on a commodity is large, then demand for such a commodity will be elastic.

8. Time Period:

Price elasticity of demand is always related to a period of time. It can be a day, a week, a month, a year or a period of several years. Elasticity of demand varies directly with the time period. Demand is generally inelastic in the short period.

It happens because consumers find it difficult to change their habits, in the short period, in order to respond to a change in the price of the given commodity. However, demand is more elastic in long rim as it is comparatively easier to shift to other substitutes, if the price of the given commodity rises.

9. Habits:

Commodities, which have become habitual necessities for the consumers, have less elastic demand. It happens because such a commodity becomes a necessity for the consumer and he continues to purchase it even if its price rises. Alcohol, tobacco, cigarettes, etc. are some examples of habit forming commodities.

Finally it can be concluded that elasticity of demand for a commodity is affected by number of factors. However, it is difficult to say, which particular factor or combination of factors determines the elasticity. It all depends upon circumstances of each case.

Price Elasticity of Demand

Economists define elasticity of demand as to how reactive the demand for a product is to changes in factors such as price or income. Let us learn more about the price elasticity of demand. However, before we go further, let us briefly revisit the laws of supply and demand.

Laws of Demand and Supply

The law of demand states that all conditions being equal, as the price of a product increases, the demand for that product will decrease. Consequently, as the price of a product decreases, the demand for that product will increase. Therefore, the law of demand defines an inverse relationship between the price and quantity factors of a product.

The law of supply, on the other hand, states that all factors being constant, an increase in price will cause an increase in the quantity supplied. That is the quantity being supplied will move in the same direction as the price. Production units will invest more in production and supply more products for sale at an increased price. Therefore, the law of supply defines a direct relationship between the price and quantity.

Price Elasticity of Demand

Now as mentioned earlier, the elasticity of demand measures how factors such as price and income affect the demand for a product. Price elasticity of demand measures how the change in a product's price affects its associated demand. Now you can measure the price elasticity of demand (PED) mathematically as follows:

Price Elasticity of Demand (PED) = % change in quantity demanded / % change in price

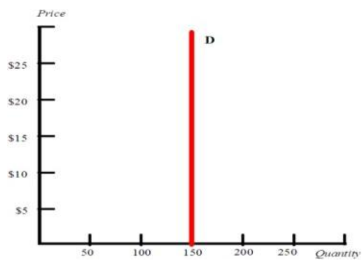
Next, let us look at how we can measure PED. Coefficient of Price Elasticity

Economists measure the price elasticity of demand (PED) in coefficients. In response to the change in price, demand for a product can be elastic, perfectly elastic, inelastic, or perfectly inelastic based on the coefficient.

Now, you need to understand that since price and demand move in opposite directions, the coefficient will have a negative value. However, in most cases, economists do not use the negative sign and focus on the coefficient itself. Let us now look at the numerical values of the coefficient of PED.

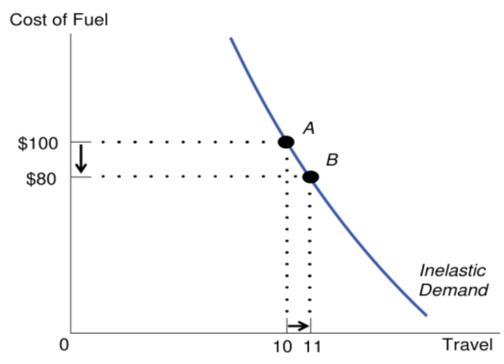
Perfectly Inelastic (PED = 0)

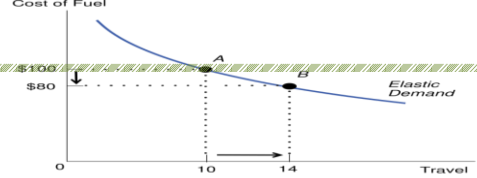
When the price elasticity of demand or PED is zero, then the demand is perfectly inelastic. That is, there is no change in the quantity demanded in response to the change in price. The demand curve remains vertical. Demand is completely unresponsive to the change in price.



Inelastic (PED is between 0 and 1)

If the percentage of change in demand is less than the percentage of change in price, then the demand is inelastic. For instance, let us say that the price of a chocolate increases from Rs.10 to Rs.20 and the associated demand decreases from ten chocolates to five chocolates. So now the PED will be 50% divided by 100%, which is 0.5. Hence, the demand here is inelastic.



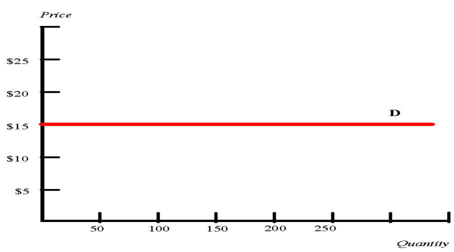


Elastic or Unit Elastic (PED = 1)

When the percentage of change in demand is the same as the percentage of change in price, then the demand is unit elastic. For example, let us say that the price of a candy drops from Rs.10 to Rs.5 and the demand increases from 10 candies to 15 candies. Here, the percentage of change in demand is equal to the percentage of change in price (50% divided by 50%, which is 1).

Perfectly Elastic (PED > 1)

If the percentage of change in demand is more than the percentage of change in price, then the demand is perfectly elastic. For instance, if a 10% increase in price causes a 20% drop in demand, then the coefficient of PED is 3, which means that the demand is perfectly elastic.



Factors that affect Price Elasticity

Now that you are familiar with the coefficient of the price elasticity of demand, let us understand the factors that affect the elasticity of demand.

1] Number of Substitutes Available

If there are several substitutes or brands available for a product, then the elasticity of demand for the product will be high because consumers can shift from one brand to another depending on the change in price. Chocolates, for instance, is a good example of substitutes. Consumers can choose between several brands of chocolates.

2] Price of Product in Relation to Income

Now when a household's income changes, the demand for goods and services also varies in response to the income. Hence, the demand for products and services becomes elastic.

3] Cost of Substitution

In some cases, the result of changing from one brand to another may be quite high. For instance, if a certain cable service has a lock-in period of deposit, then an existing consumer cannot change to another service, although inexpensive, without losing the deposit. Hence, the demand becomes inelastic.

4] Brand Loyalty

Sometimes, consumers are loyal to a specific product. In such cases, the price change in that product will not affect its associated demand. Brand loyalty, therefore, makes the demand inelastic.

5] Necessary Goods

Necessary goods such as medicines and petrol usually have an inelastic demand. As consumers have to purchase these goods irrespective of the change in price, the demand remains unresponsive.

Income Elasticity of Demand

Often, when you get a bonus at work or a raise, your first instinct is to celebrate by buying something expensive or pampering yourself and your family. But have you ever wondered how your sudden shopping spree might affect the demand for some products? In this article, we will discuss the concept of income elasticity of demand.

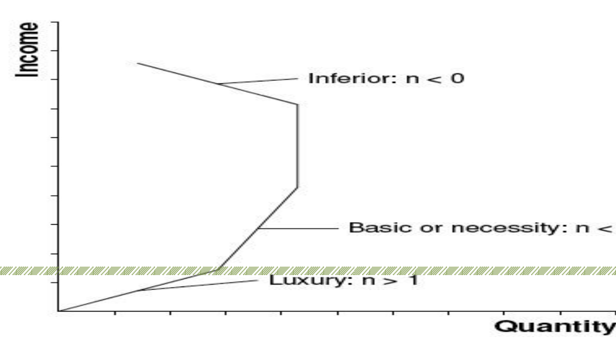
Income Elasticity of Demand

The elasticity of demand measures how factors such as price and income affect the demand for a product. The income elasticity of demand measures how the change in a consumer's income affects the demand for a specific product. You can express the income elasticity of demand mathematically as follows:

Income Elasticity of Demand (YED) = % change in quantity demanded / % change in income

The higher the income elasticity of demand for a specific product, the more responsive it becomes the change in consumers' income.

Now, we can measure the income elasticity of demand for different products by categorizing them as inferior goods and normal goods. The income elasticity of demand for a particular product can be



negative or positive, or even unresponsive.

Normal Goods and Luxuries

The income elasticity of demand for a product can elastic or inelastic based on its category— whether it is an inferior good or a normal good. Now, the coefficient for measuring income elasticity is YED.

When YED is more than zero, the product is income-elastic. Normal goods have positive YED. That is, when the consumers' income increases, the demand for these goods also increases.

However, normal goods can further be broken down into normal necessities and normal luxuries. Normal necessities have a positive but low income-elasticity compared to luxurious goods.

The income elasticity coefficient or YED for normal necessities is between 0 and 1. Normal necessities include basic needs such as milk, fuel, or medicines.

Factors such as a change in price or change in consumers' income do not affect the demand for necessary goods. The percentage of change in the demand for these products is less in proportion to the percentage of change in consumers' income.

Luxuries, on the other hand, are highly income-elastic. Examples of luxury goods include high-end electronics or jewellery. For instance, if a consumer's income increases, he/she may invest or purchase a high-end mobile or an HD television.

The percentage of change in demand is more in proportion to the change in income. However, it is important to note that the concept of luxury is contextual and it depends on the circumstances of consumers.

Inferior Goods

Inferior goods have a negative income elasticity; that is YED is less than 0. If the consumers' income increases, they demand less of these goods. Inferior goods are called inferior because they usually have superior alternatives.

For instance, if a consumer's income increases, then he/she might start taking a cab instead of opting for public transport. Public transport, in this case, is an inferior good.

Usually, when the economic growth is good and there is an increase in consumers' income, the demand for inferior goods reduces and there is an inward swing of the demand curve.

Consequently, when the incomes reduce and price of goods increases because of recession, then the demand for inferior goods increases, thereby causing an outward swing of the demand curve.

Cross Elasticity of Demand

It is the ratio of proportionate change in the quantity demanded of Y to a given proportionate change in the price of the related commodity X.

It is a measure of relative change in the quantity demanded of a commodity due to a change in the price of its substitute/complement. It can be expressed as:

$$Ce \frac{\text{Proportionate change in the quantity demanded of Y}}{\text{Proportionate change in the price of X}}$$

Cross elasticity may be infinite or zero if the slightest change in the price of X causes a substantial change in the quantity demanded of Y. It is always the case with goods which have perfect substitutes for one another. Cross elasticity is zero, if a change in the price of one commodity will not affect the quantity demanded of the other. In the case of goods which are not related to each other, cross elasticity of demand is zero.

Definition:

"The cross elasticity of demand is the proportional change in the quantity of X good demanded resulting from a given relative change in the price of a related good Y" Ferguson

"The cross elasticity of demand is a measure of the responsiveness of purchases of Y to change in the price of X" Leibafsky

Types of Cross Elasticity of Demand:

1. Positive:

When goods are substitute of each other then cross elasticity of demand is positive. In other words, when an increase in the price of Y leads to an increase in the demand of X. For instance, with the increase in price of tea, demand of coffee will increase.

In fig. 21 quantity has been measured on OX-axis and price on OY-axis. At price OP of Y- commodity, demand of X-commodity is OM. Now as price of Y commodity increases to OP1 demand of X-

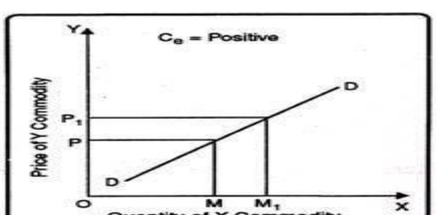
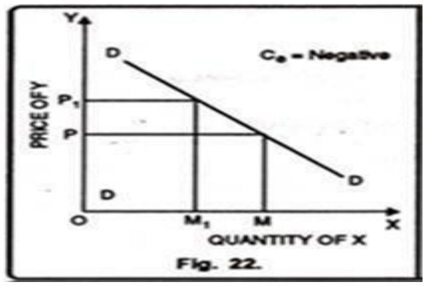


Fig. 21.

commodity increases to OM1 Thus, cross elasticity of demand is positive.

2. Negative:

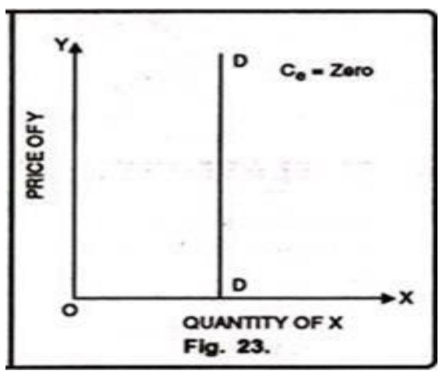
In case of complementary goods, cross elasticity of demand is negative. A proportionate increase in price of one commodity leads to a proportionate fall in the demand of another commodity because both are demanded jointly. In fig. 22 quantity has been measured on OX-axis while price has been measured on



OY-axis. When the price of commodity increases from OP to OP1 quantity demanded falls from OM to OM1. Thus, cross elasticity of demand is negative.

3. Zero:

Cross elasticity of demand is zero when two goods are not related to each other. For instance, increase in price of car does not effect the demand of cloth. Thus, cross elasticity of demand is zero. It has been



shown in fig. 23.

Therefore, it depends upon substitutability of goods. If substitutability is perfect, cross elasticity is infinite; if on the other hand, substitutability does not exist, cross elasticity is zero. In the case of complementary goods like jointly demanded goods cross elasticity is negative. A rise in the price of one commodity X will mean not only decrease in the quantity of X but also decrease in the quantity demanded of Y because both are demanded together.

Measurement of Cross Elasticity of Demand:

$$EC = \frac{\text{Percentage change in quantity demanded of Good} - X}{\text{Percentage change in the price of Good} - Y}$$

$$= \frac{\dfrac{\text{Change in quantity demanded of X}}{\text{Original Quantity of X}}}{\dfrac{\text{Change in Price of Y}}{\text{Original Price of Y}} \times 100} \times 100$$

$$= \frac{\dfrac{\Delta Q_x}{Q_x}}{\dfrac{\Delta Q_y}{Q_y}} = \frac{\Delta Q_x}{Q_x} \times \frac{P_y}{\Delta P_y}$$

$$EC = \frac{P_y}{\Delta_n} \times \frac{\Delta Q_n}{\Delta P_y}$$

Where Py = Original price of good-Y

ΔPy = Change in price of good-Y

Q_x = Original quantity demanded of X

ΔQ_x = Change in the quantity demanded of X

Cross elasticity of demand can be measured by the following formula:

Advertising Elasticity of Demand

This is a measure of effectiveness of increase in expenditure of advertising in increasing demand of a

$$AED = \frac{\%\ change\ in\ demand}{\%\ change\ in\ advertising\ expenditure} = \frac{\frac{\Delta d}{d}}{\frac{\Delta e}{e}}$$

product.

This ratio can lie between 0 and ∞ (infinity).

AED is always positive, meaning that the demand always increases with increase in advertising expenditure. Whereas values of this ratio below 1 mean that the increase in demand is less than the increase in advertising expenditure, while values greater than 1 indicate that the rise in demand is more than the rise in expenditure. While this is a good way to estimate expected rise in advertising costs for growth in demand or the expected growth with rise in expense toward advertising, this is not the most accurate way. This ratio assumes that several other factors that may affect demand are constant, which cannot be the case in real life.

The demand of a certain good/service depends on, apart from expense on advertising, the following to name a few factors:

1. Income of the people of the region (state of economy): Expensive advertising may not yield very good results if the region has been recently hit by an economic crisis where people have been laid off on a large scale and are struggling to make ends meet; or in a region with generally low income.
2. Price of the product: No matter how much money is put in the advertising, if a similar product is in the market for a lower price that may take away from the success of the advertising.
3. Quality/Appeal of the ad: High expense doesn't always mean high quality in terms of audio- visual or content. Or the ad just may lack appeal for the demographic the product is for.

All these factors can alter the demand of the product, and hence the AED. Thus, AED may not be the most accurate measure of effectiveness of increase in advertising expenditure.

Examples of advertising elasticity of demand: Wine=0.08, Cigarettes=0.04.

Demand Forecasting

Demand forecasting is a combination of two words; the first one is Demand and another forecasting. Demand means outside requirements of a product or service. In general, forecasting means making an estimation in the present for a future occurring event. Here we are going to discuss demand forecasting and its usefulness.

It is a technique for estimation of probable demand for a product or services in the future. It is based on the analysis of past demand for that product or service in the present market condition. Demand forecasting should be done on a scientific basis and facts and events related to forecasting should be considered.

Therefore, in simple words, we can say that after gathering information about various aspect of the market and demand based on the past, an attempt may be made to estimate future demand. This concept is called forecasting of demand.

For example, suppose we sold 200, 250, 300 units of product X in the month of January, February, and March respectively. Now we can say that there will be a demand for 250 units approx. of product X in the month of April, if the market condition remains the same.

Usefulness of Demand Forecasting

Demand plays a vital role in the decision making of a business. In competitive market conditions, there is a need to take correct decision and make planning for future events related to business like a sale, production, etc. The effectiveness of a decision taken by business managers depends upon the accuracy of the decision taken by them.

Demand is the most important aspect for business for achieving its objectives. Many decisions of business depend on demand like production, sales, staff requirement, etc. Forecasting is the necessity of business at an international level as well as domestic level.

Demand forecasting reduces risk related to business activities and helps it to take efficient decisions. For firms having production at the mass level, the importance of forecasting had increased more. A good forecasting helps a firm in better planning related to business goals.

There is a huge role of forecasting in functional areas of accounting. Good forecast helps in appropriate production planning, process selection, capacity planning, facility layout planning, and inventory management, etc.

Demand forecasting provides reasonable data for the organization’s capital investment and expansion decision. It also provides a way for the formulation of suitable pricing and advertisement strategies.

Following is the significance of Demand Forecasting:

- Fulfilling objectives of the business
- Preparing the budget

- Taking management decision
- Evaluating performance etc.

Moreover, forecasting is not completely full of proof and correct. It thus helps in evaluating various factors which affect demand and enables management staff to know about various forces relevant to the study of demand behavior.

The Scope of Demand Forecasting

The scope of demand forecasting depends upon the operated area of the firm, present as well as what is proposed in the future. Forecasting can be at an international level if the area of operation is international. If the firm supplies its products and services in the local market then forecasting will be at local level.

The scope should be decided considering the time and cost involved in relation to the benefit of the information acquired through the study of demand. Cost of forecasting and benefit flows from such forecasting should be in a balanced manner.

Types of Forecasting

There are two types of forecasting:

- Based on Economy
- Based on the time period

1. Based on Economy

There are three types of forecasting based on the economy:

i. Macro-level forecasting: It deals with the general economic environment relating to the economy as measured by the Index of Industrial Production(IIP), national income and general level of employment, etc.

ii. Industry level forecasting: Industry level forecasting deals with the demand for the industry's products as a whole. For example demand for cement in India, demand for clothes in India, etc.

iii. Firm-level forecasting: It means forecasting the demand for a particular firm's product. For example, demand for Birla cement, demand for Raymond clothes, etc.

2. Based on the Time Period

Forecasting based on time may be short-term forecasting and long-term forecasting

i. Short-term forecasting: It covers a short period of time, depending upon the nature of the industry. It is done generally for six months or less than one year. Short-term forecasting is generally useful in tactical decisions.

ii. Long-term forecasting casting: Long-term forecasts are for a longer period of time say, two to five years or more. It gives information for major strategic decisions of the firm. For example, expansion of plant capacity, opening a new unit of business, etc.

Methods of Demand Forecasting

Demand forecasting is the art as well as the science of predicting the likely demand for a product or service in the future. This prediction is based on past behavior patterns and the continuing trends in the present. Hence, it is not simply guessing the future demand but is estimating the demand scientifically and objectively. Thus, there are various methods of demand forecasting which we will discuss here.

There is no easy or simple formula to forecast the demand. Proper judgment along with the scientific formula is needed to correctly predict the future demand for a product or service. Some methods of demand forecasting are discussed below:

1] Survey of Buyer's Choice

When the demand needs to be forecasted in the short run, say a year, then the most feasible method is to ask the customers directly that what are they intending to buy in the forthcoming time period. Thus, under this method, potential customers are directly interviewed. This survey can be done in any of the following ways:

a. Complete Enumeration Method: Under this method, nearly all the potential buyers are asked about their future purchase plans.

b. Sample Survey Method: Under this method, a sample of potential buyers are chosen scientifically and only those chosen are interviewed.

c. End-use Method: It is especially used for forecasting the demand of the inputs. Under this method, the final users i.e. the consuming industries and other sectors are identified. The desirable norms of consumption of the product are fixed, the targeted output levels are estimated and these norms are applied to forecast the future demand of the inputs.

Hence, it can be said that under this method the burden of demand forecasting is on the buyer. However, the judgments of the buyers are not completely reliable and so the seller should take decisions in the light of his judgment also.

The customer may misjudge their demands and may also change their decisions in the future which in turn may mislead the survey. This method is suitable when goods are supplied in bulk to industries but not in the case of household customers.

2] Collective Opinion Method

Under this method, the salesperson of a firm predicts the estimated future sales in their region. The individual estimates are aggregated to calculate the total estimated future sales. These estimates are reviewed in the light of factors like future changes in the selling price, product designs, changes in competition, advertisement campaigns, the purchasing power of the consumers, employment opportunities, population, etc.

The principle underlying this method is that as the salesmen are closest to the consumers they are more likely to understand the changes in their needs and demands. They can also easily find out the reasons behind the change in their tastes.

Therefore, a firm having good sales personnel can utilize their experience to predict the demands. Hence, this method is also known as Salesforce opinion or Grassroots approach method. However, this method depends on the personal opinions of the sales personnel and is not purely scientific.

3] Barometric Method

This method is based on the past demands of the product and tries to project the past into the future. The economic indicators are used to predict the future trends of the business. Based on future trends, the demand for the product is forecasted. An index of economic indicators is formed. There are three types of economic indicators, viz. leading indicators, lagging indicators, and coincidental indicators.

The leading indicators are those that move up or down ahead of some other series. The lagging indicators are those that follow a change after some time lag. The coincidental indicators are those that move up and down simultaneously with the level of economic activities.

4] Market Experiment Method

Another one of the methods of demand forecasting is the market experiment method. Under this method, the demand is forecasted by conducting market studies and experiments on consumer behavior under actual but controlled, market conditions.

Certain determinants of demand that can be varied are changed and the experiments are done keeping other factors constant. However, this method is very expensive and time-consuming.

5] Expert Opinion Method

Usually, market experts have explicit knowledge about the factors affecting demand. Their opinion can help in demand forecasting. The Delphi technique, developed by Olaf Helmer is one such method.

Under this method, experts are given a series of carefully designed questionnaires and are asked to forecast the demand. They are also required to give the suitable reasons. The opinions are shared with the experts to arrive at a conclusion. This is a fast and cheap technique.

6] Statistical Methods

The statistical method is one of the important methods of demand forecasting. Statistical methods are scientific, reliable and free from biases. The major statistical methods used for demand forecasting are:

a. Trend Projection Method: This method is useful where the organization has a sufficient amount of accumulated past data of the sales. This date is arranged chronologically to obtain a time series. Thus, the time series depicts the past trend and on the basis of it, the future market trend can be predicted. It is assumed that the past trend will continue in the future. Thus, on the basis of the predicted future trend, the demand for a product or service is forecasted.

b. Regression Analysis: This method establishes a relationship between the dependent variable and the independent variables. In our case, the quantity demanded is the dependent variable and income, the price of goods, the price of related goods, the price of substitute goods, etc. are independent variables. The regression equation is derived assuming the relationship to be linear. Regression Equation: $Y = a + bX$. Where Y is the forecasted demand for a product or service.

UNIT - 3

THEORY OF PRODUCTION

Meaning & concept of Production

Since the primary purpose of economic activity is to produce utility for individuals, we count as production during a time period all activity which either creates utility during the period or which increases ability of the society to create utility in the future.

Business firms are important components (units) of the economic system.

They are artificial entities created by individuals for the purpose of organising and facilitating production. The essential characteristics of the business firm is that it purchases factors of production such as land, labour, capital, intermediate goods, and raw material from households and other business firms and transforms those resources into different goods or services which it sells to its customers, other business firms and various units of the government as also to foreign countries.

Definition of Production:

According to Bates and Parkinson:

“Production is the organised activity of transforming resources into finished products in the form of goods and services; the objective of production is to satisfy the demand for such transformed resources”.

According to J. R. Hicks:

“Production is any activity directed to the satisfaction of other peoples’ wants through exchange”. This definition makes it clear that, in economics, we do not treat the mere making of things as production. What is made must be designed to satisfy wants.

What is not Production?

The making or doing of things which are not wanted or are made just for the fun of it does not qualify as production. On the other hand, all jobs which do aim at satisfying wants are part of production.

Those who provide services Such as hair-dressers, solicitors, bus drivers, postmen, and clerks are as much a part of the process of satisfying wants as are farmers, miners, factory workers and bakers. The test of whether or not any activity is productive is whether or not anyone will buy its end-product. If we will buy something we must want it; if we are not willing to buy it then, in economic terms, we do not want it.

Importance of Exchange:

So from our above definition it is clear that many valuable activities such as the work done by people in their own houses and gardens (the so-called do it yourself exercise) and all voluntary work (such as free coaching, free-nursing, collection of subscription for a social cause such as flood-relief or earthquake-relief) immensely add to the quality of life but there is no practical way of measuring their economic worth (value).

This being so, and because in economics an important task is to measure changes in the volume of production, it is necessary to add the qualifying clause ‘through exchange’, i.e., in return for money, to the definition of production.

Three Types of Production:

For general purposes, it is necessary to classify production into three main groups:

1. Primary Production:

Primary production is carried out by ‘extractive’ industries like agriculture, forestry, fishing, mining and oil extraction. These industries are engaged in such activities as extracting the gifts of Nature from the earth’s surface, from beneath the earth’s surface and from the oceans.

2. Secondary Production:

This includes production in manufacturing industry, viz., turning out semi-finished and finished goods from raw materials and intermediate goods— conversion of flour into bread or iron ore into finished steel.

They are generally described as manufacturing and construction industries, such as the manufacture of cars, furnishing, clothing and chemicals, as also engineering and building.

3. Tertiary Production:

Industries in the tertiary sector produce all those services which enable the finished goods to be put in the hands of consumers. In fact, these services are supplied to the firms in all types of industry and directly to consumers. Examples cover distributive traders, banking, insurance, transport and communications. Government services, such as law, administration, education, health and defence, are also included.

Output:

Any activity connected with money earning and money-spending is called an economic activity. Production is an important economic activity. It results in the output (creation) of an enormous variety of economic goods and services.

Factors Affecting Production:

There are several factors which go to determine the volume of production in a country.

They are:

(i) Natural Factors:

The amount and nature of production in a country depends on its climate, nature of the soil, rainfall, etc. Production is diminished by natural calamities like earthquakes, floods, droughts and hailstorms.

(ii) Political Factors:

The form and character of government have a great deal to do with the volume of production in a country. In Russia, the Soviet Government has brought about tremendous increase in production through planning. For a long time, the foreign Government in India encouraged the production of raw materials and discouraged manufacturing industry in the country.

(iii) Technical Progress:

Production largely depends on the state of scientific knowledge and technical progress in the country. Discovery of new materials, new processes and new machines is bound to increase the volume of production. This is what is happening in India at present.

(iv) Development of Credit and Banking and Means of Transport and Communication:

In the absence of a sound banking system and efficient and cheap means of transport and communication, production is bound to Sutter. These are the prime needs of a country if production is to be increased. They are known as infrastructure.

(v) Character of the People:

Inhabiting a country also exerts a powerful influence on the nature and volume of its production. Hard working, educated and disciplined people can always produce relatively more and better goods than those who do not possess such qualities.

Factors of Production

Factors of production means resources used in the process of production of commodities. There are of four types viz., land, labour, capital and organization or enterprise. Here, land represents natural resources (such as soil, mineral deposits, seas, rivers, natural forests, fisheries etc). Labour represents human resources. Together, these two factors are called the 'primary factors of production'.

These two factors produce some units of goods for the purpose of consumption. And as consumption of these goods takes place, there is the possibility of some of these goods getting left over. Thus, saving is production minus consumption. This saved amount is called as capital, which serves as investment in the production process. Also, organisation or enterprise is a special form of labour. The third and the fourth factors are called 'secondary factors of production'.

These four factors depend on each other. They have a coordinated impact on production of goods and services.

1. Land

In ordinary sense 'land' refers to the soil or the surface of the earth or ground. But, in Economics, land means all gifts of Nature owned and controlled by human beings which yield an income. Land is the original source of all material wealth. The economic prosperity of a country depends on the richness of her natural resources. The quality and quantity of agricultural wealth are determined by the nature of soil, climate and rainfall.

The agricultural products are the basis of trade and industry. Industry survives on the availability of coal-mines or waterfall for electricity production. Hence, all aspects of economic life like agriculture, trade and industry are generally influenced by natural resources which are called as "Land" in economics.

Characteristics of Land

1) Land is a primary factor of production.

2) Land is a passive factor of production.

3) Land is the free gift of Nature.

4) Land has no cost of production.

5) Land is fixed in supply. It is inelastic in supply.

6) Land is permanent.

7) Land is immovable.

8) Land is heterogeneous as it differs in fertility.

9) Land has alternative uses.

10) Land is subject to Law of Diminishing Returns.

2. Labour

Labour is the active factor of production. In common parlance, labour means manual labour or unskilled work. But in Economics the term ‘labour’ has a wider meaning.

It refers to any work undertaken for securing an income or reward. Such work may be manual or intellectual. For example, the work done by an agricultural worker or a cook or rickshaw puller or a mason is manual. The work of a doctor or teacher or an engineer is intellectual. In short, labour in economics refers to any type of work performed by a labourer for earning an income.

According to Marshall, labour represents services provided by the factor labour, which helps in yielding an income to the owner of the labour-power.

Characteristics of Labour

1) Labour is the animate factor of production.

2) Labour is an active factor of production.

3) Labour implies several types: it may be manual (farmer) or intellectual (teacher, lawyer etc).

4) Labour is perishable.

5) Labour is inseparable from the Labourer.

6) Labour is less mobile between places and occupations.

7) Labour is a means as well as an end. It is both the cause of production and consumer of the product.

8) Labour units are heterogeneous.

9) Labour differs in ability.

10) Labour-supply determines its reward (wage).

11) Labour has weak bargaining power.

3. Capital

Marshall says "capital consists of all kinds of wealth other than free gifts of nature, which yield income". Bohm-Bawerk defines it as 'a produced means of production'. As said earlier, capital is a secondary means of production. It refers to that part of production which represents 'saving used as investment' in the further production process. For example, the entire mango is not eaten; a part of that (its nut) is used to produce more mangoes.

It is a stock concept. All capital is wealth but all wealth is not capital. For example, tractor is a capital asset which can be used in cultivation (production) of farm, but due to some reason the same is kept unused (idle) for some period. It cannot be termed as capital for that period. It is only wealth.

Characteristics of Capital

1) Capital is a man-made factor.

2) Capital is mobile between places and persons.

3) Capital is a passive factor of production.

4) Capital’s supply is elastic.

5) Capital’s demand is a derived demand.

6) Capital is durable.

Capital may be tangible or intangible. For example, buildings, plants and machinery, factories, inventories of inputs, warehouses, roads, highways etc are tangible capital. The examples for intangible capital are investment on advertisement, expenses on training programme etc.

Financial Capital means the assets needed by a firm to provide goods and services measured in term of money value . It is normally raised through debt and equity issues .The prime aim of it is to a mass wealth in terms of profit.

4. Entrepreneur

The man behind organizing the business is called as ‘Organizer’ or ‘Entrepreneur’. An organiser is the most important factor of production. He represents a special type of labour. Joseph Schumpeter says that an entrepreneur innovates, coordinates other factors of production, plans and runs a business. He not only runs the business, but bears the risk of business. His reward is residual. This residual is either positive (profit) or negative (loss) or zero.

Functions of an Entrepreneur

1) Initiation: An organizer is the initiator of the business, by considering the situation and availability of resources and planning the entire process of business or production.

2) Innovation: A successful entrepreneur is always an innovator. He introduces new methods in the production process.

3) Coordination: An organizer applies a particular combination of the factors of production to start and run the business or production.

4) Control, Direction and Supervision: An organiser controls so that nothing prevents the organisation from achieving its goal. He directs the factors to get better results and supervises for the efficient functioning of all the factors involved in the process of production.

5) Risk-taking and Uncertainty-bearing: There are risk-taking and uncertainty-bearing obstacles. Risks may be insured but uncertainties cannot be insured. They reduce the profit.

Production Function

Production is the result of co-operation of four factors of production viz., land, labour, capital and organization.

This is evident from the fact that no single commodity can be produced without the help of any one of these four factors of production.

Therefore, the producer combines all the four factors of production in a technical proportion. The aim of the producer is to maximize his profit. For this sake, he decides to maximize the production at minimum cost by means of the best combination of factors of production.

The producer secures the best combination by applying the principles of equi-marginal returns and substitution. According to the principle of equi-marginal returns, any producer can have maximum production only when the marginal returns of all the factors of production are equal to one another. For instance, when the marginal product of the land is equal to that of labour, capital and organisation, the production becomes maximum.

Meaning of Production Function:

In simple words, production function refers to the functional relationship between the quantity of a good produced (output) and factors of production (inputs).

"The production function is purely a technical relation which connects factor inputs and output." Prof. Koutsoyiannis

Defined production function as "the relation between a firm's physical production (output) and the material factors of production (inputs)." Prof. Watson

In this way, production function reflects how much output we can expect if we have so much of labour and so much of capital as well as of labour etc. In other words, we can say that production function is an indicator of the physical relationship between the inputs and output of a firm.

The reason behind physical relationship is that money prices do not appear in it. However, here one thing that becomes most important to quote is that like demand function a production function is for a definite period.

It shows the flow of inputs resulting into a flow of output during some time. The production function of a firm depends on the state of technology. With every development in technology the production function of the firm undergoes a change.

The new production function brought about by developing technology displays same inputs and more output or the same output with lesser inputs. Sometimes a new production function of the firm may be adverse as it takes more inputs to produce the same output.

Mathematically, such a basic relationship between inputs and outputs may be expressed as: $Q = f(L, C, N)$

Where Q = Quantity of output L = Labour

C = Capital N = Land.

Hence, the level of output (Q), depends on the quantities of different inputs (L, C, N) available to the firm. In the simplest case, where there are only two inputs, labour (L) and capital (C) and one output (Q), the production function becomes.

Q =f (L, C)

Definitions:

"The production function is a technical or engineering relation between input and output. As long as the natural laws of technology remain unchanged, the production function remains unchanged." Prof. L.R. Klein

"Production function is the relationship between inputs of productive services per unit of time and outputs of product per unit of time." Prof. George J. Stigler

"The relationship between inputs and outputs is summarized in what is called the production function. This is a technological relation showing for a given state of technological knowledge how much can be produced with given amounts of inputs." Prof. Richard J. Lipsey

Thus, from the above definitions, we can conclude that production function shows for a given state of technological knowledge, the relation between physical quantities of inputs and outputs achieved per period of time.

Features of Production Function:

Following are the main features of production function:

1. Substitutability:

The factors of production or inputs are substitutes of one another which make it possible to vary the total output by changing the quantity of one or a few inputs, while the quantities of all other inputs are held constant. It is the substitutability of the factors of production that gives rise to the laws of variable proportions.

2. Complementarity:

The factors of production are also complementary to one another, that is, the two or more inputs are to be used together as nothing will be produced if the quantity of either of the inputs used in the production process is zero.

The principles of returns to scale is another manifestation of complementarity of inputs as it reveals that the quantity of all inputs are to be increased simultaneously in order to attain a higher scale of total output.

3. Specificity:

It reveals that the inputs are specific to the production of a particular product. Machines and equipment's, specialized workers and raw materials are a few examples of the specificity of factors of production. The specificity may not be complete as factors may be used for production of other commodities too. This reveals that in the production process none of the factors can be ignored and in some cases ignorance to even slightest extent is not possible if the factors are perfectly specific. Production involves time; hence, the way the inputs are combined is determined to a large extent by the time period under consideration. The greater the time period, the greater the freedom the producer has to vary the quantities of various inputs used in the production process.

In the production function, variation in total output by varying the quantities of all inputs is possible only in the long run whereas the variation in total output by varying the quantity of single input may be possible even in the short run.

Fixed factors and Variable factors

Production is the result of combined efforts of the factors of production. These factors may be fixed or variable. A fixed factor is one, whose quantity cannot readily be changed in response to desired changes in output or market conditions.

Fixed Factors

Fixed factors are those which remain unchanged as out output of the firm changes in the shout- run. In other words as a firm increases or decreases its output in the short-run, fixed factors remain constant. They are independent of output in the short-run. Machines, factory buildings, plants, permanent employees etc. are the examples of fixed factors. To construct a new plant or expand the existing one for changing the output of the firm will take time. It is not possible in the short-run. What is possible is to employ more of labour and raw materials in the existing plant for expanding the output of the firm.

Variable Factors

Variable factors are those factor inputs which change with the change with the change of output in the short run. Raw materials, labour, fuel, power etc. are the examples of variable factors. If a firm wants to expand output in the short-run, then it can employ more labourers, purchase more raw materials and can use more power. Similarly if it wants to contract output, then it can retrench workers, purchase less of raw materials and fuel etc. This shows that as production increases, variable factors also increase and as production falls the quantities of variable factors also fall.

It is important to note that the distinction between the fixed factors and variable factors. So important in the short-run vanishes in the long-run. In the long-run. In the long-run all factors are variable.

Distinction Between Fixed and Variable Factors

FIXED FACTORS	VARIABLE FACTORS
1. Fixed factors exist only in the short-run.	1. Variable factors exist both in the short-run and long-run.
2. It is independent of output in the short-run.	2. It changes with the change of output in the short-run
3. Plant, machines etc. are the examples of fixed	3. Labour, raw materials etc. are the examples of variable factors.
4. It exists even in the zero level of output.	4. When output is zero, quantities of variable factors are reduced to zero.

The distinction between fixed and variable factors is related to two periods the short-run and the long-run. The period of short-run is too short to cause variation in fixed factors. Thus, in the short-run, some factors are fixed, while the others are variable. The production can be increased only by increasing the quantity of the variable factors or by having additional shifts or by increasing the hours of work. But, in the long-run (also called as planning period of the firm), all the factors are variable, i.e., the quantity of all the factors required can be varied to produce an output ranging from zero to an indefinite quantity.

All investment options are open including installation of new plant and machinery. In the long run, it is possible for a firm to branch out into new products or new areas or to modernise or reorganise its method of production through invention of new techniques.

The distinction between fixed and variable factors helps us to study the law of variable proportions and the law of returns of scale. These laws of production show the relationship between the factors of production and output in the short-run and long-run respectively.

Law of Variable Proportion (short run production analysis)

Law of Variable Proportions occupies an important place in economic theory. This law is also known as Law of Proportionality.

Keeping other factors fixed, the law explains the production function with one factor variable. In the short run when output of a commodity is sought to be increased, the law of variable proportions comes into operation. Therefore, when the number of one factor is increased or decreased, while other factors are constant, the proportion between the factors is altered. For instance, there are two factors of production viz., land and labour.

Land is a fixed factor whereas labour is a variable factor. Now, suppose we have a land measuring 5 hectares. We grow wheat on it with the help of variable factor i.e., labour. Accordingly, the proportion between land and labour will be 1: 5. If the number of laborers is increased to 2, the new proportion between labour and land will be 2: 5. Due to change in the proportion of factors there will also emerge a change in total output at different rates. This tendency in the theory of production called the Law of Variable Proportion.

Definitions:

"As the proportion of the factor in a combination of factors is increased after a point, first the marginal and then the average product of that factor will diminish." Benham

"An increase in some inputs relative to other fixed inputs will in a given state of technology cause output to increase, but after a point the extra output resulting from the same additions of extra inputs will become less and less." Samuelson

"The law of variable proportion states that if the inputs of one resource is increased by equal increment per unit of time while the inputs of other resources are held constant, total output will increase, but beyond some point the resulting output increases will become smaller and smaller." Leftwitch

Assumptions:

Law of variable proportions is based on following assumptions:

(i) Constant Technology:

The state of technology is assumed to be given and constant. If there is an improvement in technology the production function will move upward.

(ii) Factor Proportions are Variable:

The law assumes that factor proportions are variable. If factors of production are to be combined in a fixed proportion, the law has no validity.

(iii) Homogeneous Factor Units:

The units of variable factor are homogeneous. Each unit is identical in quality and amount with every other unit.

(iv) Short-Run:

The law operates in the short-run when it is not possible to vary all factor inputs.

Explanation of the Law:

In order to understand the law of variable proportions we take the example of agriculture. Suppose land and labour are the only two factors of production.

By keeping land as a fixed factor, the production of variable factor i.e., labour can be shown with the help of the following table:

Table 1.

Units of Land	Units of Labour	Total Production	Average Production	Marginal Production	
10 Acres	0	–	–	–	
”	1	20	20	20	1st stage MP > AP
”	2	50	25	30	
”	3	90	30	40	
”	4	120	30	30	AP = MP
”	5	140	28	20	2nd stage MP=0 and TP Maximum
”	6	150	25	10	
”	7	150	21.3	0	
”	8	140	17.5	–10	3rd stage MP < 0

From the table 1 it is clear that there are three stages of the law of variable proportion. In the first stage average production increases as there are more and more doses of labour and capital employed with fixed factors (land). We see that total product, average product, and marginal product increases but average product and marginal product increases up to 40 units. Later on, both start decreasing because proportion of workers to land was sufficient and land is not properly used. This is the end of the first stage.

The second stage starts from where the first stage ends or where AP=MP. In this stage, average product and marginal product start falling. We should note that marginal product falls at a faster rate than the average product. Here, total product increases at a diminishing rate. It is also maximum at 70 units of labour where marginal product becomes zero while average product is never zero or negative.

The third stage begins where second stage ends. This starts from 8th unit. Here, marginal product is negative and total product falls but average product is still positive. At this stage, any additional dose leads to positive nuisance because additional dose leads to negative marginal product.

Graphic Presentation:

In fig. 1, on OX axis, we have measured number of labourers while quantity of product is shown on OY axis. TP is total product curve. Up to point 'E', total product is increasing at increasing rate. Between points E and G it is increasing at the decreasing rate. Here marginal product has started falling. At point 'G' i.e., when 7 units of labourers are employed, total product is maximum while, marginal product is zero. Thereafter, it begins to diminish corresponding to negative marginal product. In the lower part of the figure MP is marginal product curve.

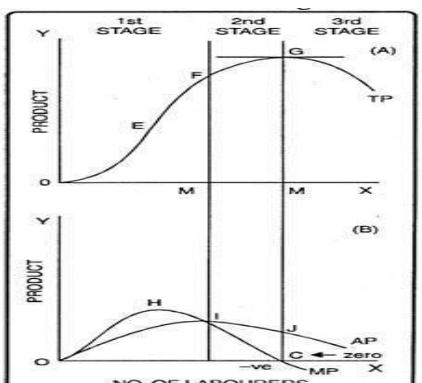


Fig 1.

Up to point 'H' marginal product increases. At point 'H', i.e., when 3 units of labourers are employed, it is maximum. After that, marginal product begins to decrease. Before point 'I' marginal product becomes zero at point C and it turns negative. AP curve represents average product. Before point 'I', average product is less than marginal product. At point 'I' average product is maximum. Up to point T, average product increases but after that it starts to diminish.

Three Stages of the Law:

1. First Stage:

First stage starts from point ‘O’ and ends up to point F. At point F average product is maximum and is equal to marginal product. In this stage, total product increases initially at increasing rate up to point E. between ‘E’ and ‘F’ it increases at diminishing rate. Similarly marginal product also increases initially and reaches its maximum at point ‘H’. Later on, it begins to diminish and becomes equal to average product at point T. In this stage, marginal product exceeds average product (MP > AP).

2. Second Stage:

It begins from the point F. In this stage, total product increases at diminishing rate and is at its maximum at point ‘G’ correspondingly marginal product diminishes rapidly and becomes ‘zero’ at point ‘C’. Average product is maximum at point ‘I’ and thereafter it begins to decrease. In this stage, marginal product is less than average product (MP < AP).

3. Third Stage:

This stage begins beyond point ‘G’. Here total product starts diminishing. Average product also declines. Marginal product turns negative. Law of diminishing returns firmly manifests itself. In this stage, no firm will produce anything. This happens because marginal product of the labour becomes negative. The employer will suffer losses by employing more units of labourers. However, of the three stages, a firm will like to produce up to any given point in the second stage only.

Total Product	Marginal Product	Average Product
Stage I		
First increases at increasing rate then at diminishing rate.	Increases in the beginning then reaches a maximum and begins to decrease.	First increases, continues to increase and becomes maximum.
Stage II		
Continues to increase at diminishing rate and becomes maximum.	Continues to diminish and becomes equal to zero.	Becomes equal to MP and then begins to diminish.
Stage III		
Diminishes	Becomes negative.	Continues to diminish but will always be greater than zero.

In Which Stage Rational Decision is Possible:

To make the things simple, let us suppose that, a is variable factor and b is the fixed factor. And a1, a2 , a3….are units of a and b1 b2b3…… are unit of b.

Stage I is characterized by increasing AP, so that the total product must also be increasing. This means that the efficiency of the variable factor of production is increasing i.e., output per unit of a is increasing. The efficiency of b, the fixed factor, is also increasing, since the total product with b1 is increasing.

The stage II is characterized by decreasing AP and a decreasing MP, but with MP not negative. Thus, the efficiency of the variable factor is falling, while the efficiency of b, the fixed factor, is increasing, since the TP with b1 continues to increase.

Finally, stage III is characterized by falling AP and MP, and further by negative MP. Thus, the efficiency of both the fixed and variable factor is decreasing.

Rational Decision:

Stage II becomes the relevant and important stage of production. Production will not take place in either of the other two stages. It means production will not take place in stage III and stage I. Thus, a rational producer will operate in stage II.

Suppose b were a free resource; i.e., it commanded no price. An entrepreneur would want to achieve the greatest efficiency possible from the factor for which he is paying, i.e., from factor a. Thus, he would want to produce where AP is maximum or at the boundary between stage I and II.

If on the other hand, a were the free resource, then he would want to employ b to its most efficient point; this is the boundary between stage II and III.

Obviously, if both resources commanded a price, he would produce somewhere in stage II. At what place in this stage production takes place would depend upon the relative prices of a and b.

Condition or Causes of Applicability:

There are many causes which are responsible for the application of the law of variable proportions.

They are as follows:

1. Under Utilization of Fixed Factor:

In initial stage of production, fixed factors of production like land or machine, is under-utilized. More units of variable factor, like labour, are needed for its proper utilization. As a result of employment of additional units of variable factors there is proper utilization of fixed factor. In short, increasing returns to a factor begins to manifest itself in the first stage.

2. Fixed Factors of Production.

The foremost cause of the operation of this law is that some of the factors of production are fixed during the short period. When the fixed factor is used with variable factor, then its ratio compared to variable factor falls. Production is the result of the co-operation of all factors. When an additional unit of a variable factor has to produce with the help of relatively fixed factor, then the marginal return of variable factor begins to decline.

3. Optimum Production:

After making the optimum use of a fixed factor, then the marginal return of such variable factor begins to diminish. The simple reason is that after the optimum use, the ratio of fixed and variable factors become defective. Let us suppose a machine is a fixed factor of production. It is put to optimum use when 4 labourers are employed on it. If 5 labourers are put on it, then total production increases very little and the marginal product diminishes.

4. Imperfect Substitutes:

Mrs. Joan Robinson has put the argument that imperfect substitution of factors is mainly responsible for the operation of the law of diminishing returns. One factor cannot be used in place of the other factor. After optimum use of fixed factors, variable factors are increased and the amount of fixed factor could be increased by its substitutes.

Such a substitution would increase the production in the same proportion as earlier. But in real practice factors are imperfect substitutes. However, after the optimum use of a fixed factor, it cannot be substituted by another factor.

Applicability of the Law of Variable Proportions:

The law of variable proportions is universal as it applies to all fields of production. This law applies to any field of production where some factors are fixed and others are variable. That is why it is called the law of universal application.

The main cause of application of this law is the fixity of any one factor. Land, mines, fisheries, and house building etc. are not the only examples of fixed factors. Machines, raw materials may also become fixed in the short period. Therefore, this law holds good in all activities of production etc. agriculture, mining, manufacturing industries.

1. Application to Agriculture:

With a view of raising agricultural production, labour and capital can be increased to any extent but not the land, being fixed factor. Thus when more and more units of variable factors like labour and capital are applied to a fixed factor then their marginal product starts to diminish and this law becomes operative.

2. Application to Industries:

In order to increase production of manufactured goods, factors of production has to be increased. It can be increased as desired for a long period, being variable factors. Thus, law of increasing returns operates in industries for a long period. But, this situation arises when additional units of labour, capital and enterprise are of inferior quality or are available at higher cost.

As a result, after a point, marginal product increases less proportionately than increase in the units of labour and capital. In this way, the law is equally valid in industries.

Postponement of the Law:

The postponement of the law of variable proportions is possible under following conditions:

(i) Improvement in Technique of Production:

The operation of the law can be postponed in case variable factors techniques of production are improved.

(ii) Perfect Substitute:

The law of variable proportion can also be postponed in case factors of production are made perfect substitutes i.e., when one factor can be substituted for the other.

Law of Return to Scale (long run production analysis)

The law of returns to scale explains the proportional change in output with respect to proportional change in inputs.

In other words, the law of returns to scale states when there are a proportionate change in the amounts of inputs, the behavior of output also changes.

In the long run all factors of production are variable. No factor is fixed. Accordingly, the scale of production can be changed by changing the quantity of all factors of production.

Definition:

“The term returns to scale refers to the changes in output as all factors change by the same proportion.” Koutsoyiannis

The degree of change in output varies with change in the amount of inputs. For example, an output may change by a large proportion, same proportion, or small proportion with respect to change in input.

On the basis of these possibilities, law of returns can be classified into three categories:

i. Increasing returns to scale

ii. Constant returns to scale

iii. Diminishing returns to scale

Explanation:

In the long run, output can be increased by increasing all factors in the same proportion. Generally, laws of returns to scale refer to an increase in output due to increase in all factors in the same proportion. Such an increase is called returns to scale.

Suppose, initially production function is as follows:

P = f (L, K)

Now, if both the factors of production i.e., labour and capital are increased in same proportion i.e., x, product function will be rewritten as.

$$P_1 = f(xL, xK)$$

1. If P_1 increases in the same proportion as the increase in factors of production i.e., $\frac{P_1}{P} = x$, it will be constant returns to scale.

2. If P_1 increases less than proportionate increase in the factors of production *i.e.*, $\frac{P_1}{P} < x$, it will be diminishing returns to scale.

3. If P_1 increases more than proportionate increase in the factors of production, *i.e.*, $\frac{P_1}{P} > x$, it will be increasing returns to scale. Returns to scale can be shown with the help of table 8.

Table 8. Showing different stages of return to scale

Units of Labour	Units of capital	%age increase in Labour & Capital	Total Product	%age increase in TP	Returns to scale
1	3	–	10	–	
2	9	100%	30	200%	Increasing
3	9	50%	60	100%	
4	12	33%	80	33%	Constant
5	15	25%	100	25%	
6	18	20%	120	10%	Decreasing
7	21	16.6%	130	8.3%	

The above stated table explains the following three stages of returns to scale:

1. Increasing Returns to Scale:

Increasing returns to scale or diminishing cost refers to a situation when all factors of production are increased, output increases at a higher rate. It means if all inputs are doubled, output will also increase at the faster rate than double. Hence, it is said to be increasing returns to scale. This increase is due to many reasons like division external economies of scale. Increasing returns to scale can be illustrated with the help of a diagram 8.

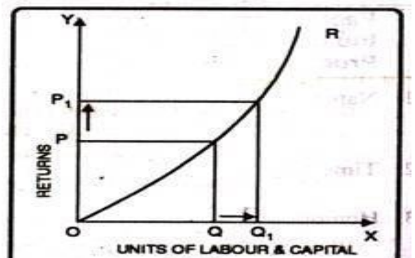


Fig 8.

In figure 8, OX axis represents increase in labour and capital while OY axis shows increase in output. When labour and capital increases from Q to Q1, output also increases from P to P1 which is higher than the factors of production i.e. labour and capital.

2. Constant Returns to Scale:

Constant returns to scale or constant cost refers to the production situation in which output increases exactly in the same proportion in which factors of production are increased. In simple terms, if factors of production are doubled output will also be doubled.

In this case internal and external economies are exactly equal to internal and external diseconomies. This situation arises when after reaching a certain level of production, economies of scale are balanced by diseconomies of scale. This is known as homogeneous production function. Cobb-Douglas linear homogenous production function is a good example of this kind. This is shown in diagram 10. In figure 10, we see that increase in factors of production i.e. labour and capital are equal to the proportion of output increase. Therefore, the result is constant returns to scale.

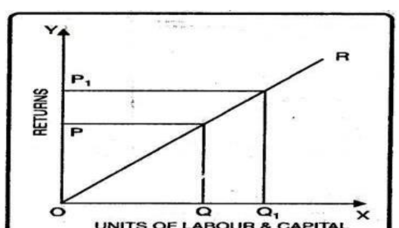


Fig. 10.

3. Diminishing Returns to Scale:

Diminishing returns or increasing costs refer to that production situation, where if all the factors of production are increased in a given proportion, output increases in a smaller proportion. It means, if inputs are doubled, output will be less than doubled. If 20 percent increase in labour and capital is followed by 10 percent increase in output, then it is an instance of diminishing returns to scale.

The main cause of the operation of diminishing returns to scale is that internal and external economies are less than internal and external diseconomies. It is clear from diagram 9.

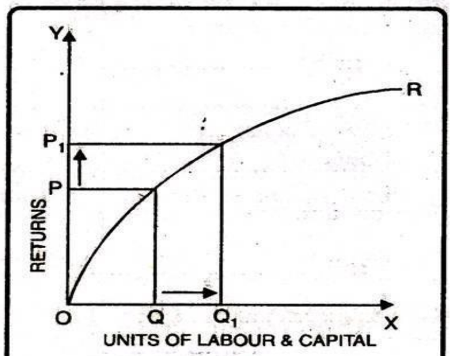


Fig. 9.

In this diagram 9, diminishing returns to scale has been shown. On OX axis, labour and capital are given while on OY axis, output. When factors of production increase from Q to Q1 (more quantity) but as a result increase in output, i.e. P to P1 is less. We see that increase in factors of production is more and increase in production is comparatively less, thus diminishing returns to scale apply.

Diminishing returns to scale is due to diseconomies of scale, which arises because of the managerial inefficiency. Generally, managerial inefficiency takes place in large-scale organizations. Another cause of diminishing returns to scale is limited natural resources. For example, a coal mining organization can increase the number of mining plants, but cannot increase output due to limited coal reserves.

Law of Returns to Scale through the use of ISOQUANTS

The laws of returns to scale can also be explained in terms of the isoquant approach. The laws of returns to scale refer to the effects of a change in the scale of factors (inputs) upon output in the long-run when the combinations of factors are changed in some proportion.

If by increasing two factors, say labour and capital, in the same proportion, output increases in exactly the same proportion, there are constant returns to scale.

If in order to secure equal increases in output, both factors are increased in larger proportionate units, there are decreasing returns to scale. If in order to get equal increases in output, both factors are increased in smaller proportionate units, there are increasing returns to scale.

The returns to scale can be shown diagrammatically on an expansion path "by the distance between successive 'multiple-level-of-output' isoquants, that is, isoquants that show levels of output which are multiples of some base level of output, e.g., 100, 200, 300, etc."

Increasing Returns to Scale:

Figure 11 shows the case of increasing returns to scale where to get equal increases in output, lesser proportionate increases in both factors, labour and capital, are required. It follows that in the figure

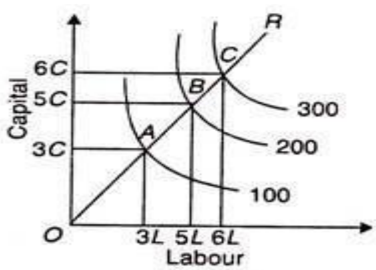


Fig. 11

100 units of output require 3C + 3L 200 units of output require 5C + 5L 300 units of output require 6C + 6L

so that along the expansion path OR, OA > AB > BC. In this case, the production function is homogeneous of degree greater than one.

The increasing returns to scale are attributed to the following factors:

1. There may be indivisibilities in machines, management, labour, finance, etc. Some items of equipment or some activities have a minimum size and cannot be divided into smaller units. When a business unit expands, the returns to scale increase because the indivisible factors are employed to their full capacity.

2. Increasing returns to scale also result from specialisation and division of labour. When the scale of the firm expands, there is wide scope for specialisation and division of labour. Work can be divided into small tasks and workers can be concentrated to narrower range of processes. For this, specialized equipment can be installed. Thus with specialization, efficiency increases and increasing returns to scale follow.

3. As the firm expands, it enjoys internal economies of production. It may be able to install better machines, sell its products more easily, borrow money cheaply, procure the services of more efficient manager and workers, etc. All these economies help in increasing the returns to scale more than proportionately.

4. A firm also enjoys increasing returns to scale due to external economies. When the industry itself expands to meet the increased long-run demand for its product, external economies appear

which are shared by all the firms in the industry.

When a large number of firms are concentrated at one place, skilled labour, credit and transport facilities are easily available. Subsidiary industries crop up to help the main industry. Trade journals, research and training centres appear which help in increasing the productive efficiency of the firms. Thus these external economies are also the cause of increasing returns to scale.

Decreasing Returns to Scale:

Figure 12 shows the case of decreasing returns where to get equal increases in output, larger proportionate increases in both labour and capital are required. It follows that

100 units of output require 2C + 2L

200 units of output require 5C + 5L 300 units of output require 9C + 9L so that along the expansion path OR, OG < GH < HK.

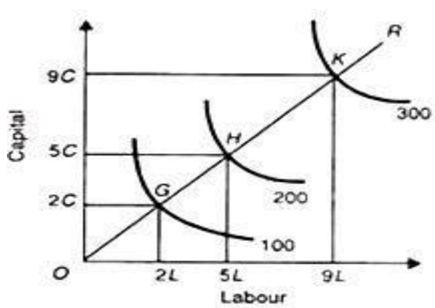


Fig. 12

In this case, the production function is homogeneous of degree less than one.

Returns to scale may start diminishing due to the following factors:

1. Indivisible factors may become inefficient and less productive.

2. The firm experiences internal diseconomies. Business may become unwieldy and produce problems of supervision and coordination. Large management creates difficulties of control and rigidities.

3. To these internal diseconomies are added external diseconomies of scale. These arise from higher factor prices or from diminishing productivities of the factors. As the industry continues to expand the demand for skilled labour, land, capital, etc. rises.

There being perfect competition, intensive bidding raises wages, rent and interest. Prices of raw materials also go up. Transport and marketing difficulties emerge. All these factors tend to raise costs and the expansion of the firms leads to diminishing returns to scale so that doubling the scale would not lead to doubling the output.

Constant Returns to Scale:

Figure 13 shows the case of constant returns to scale. Where the distance between the isoquants 100, 200 and 300 along the expansion path OR is the same, i.e., OD = DE = EE It means that if units of both

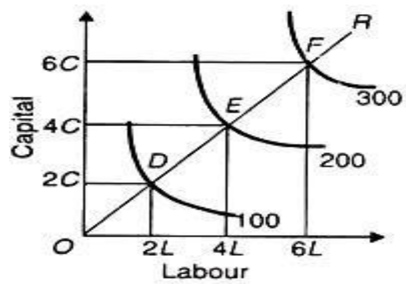


Fig. 13 ·

factors, labour and capital, are doubled, the output is doubled. To treble output, units of both factors are trebled. It follows that

100 units of output require 1 (2C + 2L) = 2C + 2L 200 units of output require 2(2C + 2L) = 4C + 4L 300 units of output require 3(2C + 2L) = 6C + 6L

Returns to scale are constant due to the following factors:

1. The returns to scale are constant when internal economies enjoyed by a firm are neutralized by internal diseconomies so that output increases in the same proportion.
2. Another reason is the balancing of external economies and external diseconomies.
3. Constant returns to scale also result when factors of production are perfectly divisible, substitutable, and homogeneous and their supplies are perfectly elastic at given prices.

That is why, in the case of constant returns to scale, the production function is homogeneous of degree one.

Alternate Method:

We have explained above the three laws of returns to scale separately on the assumption that there are three processes and each process shows the same returns over all ranges of output. "However, the technological conditions of production may be such that returns to scale may vary over different ranges of output. Over some range, we may have constant returns to scale, while over another range we may have increasing or decreasing returns to scale".

To explain it, we draw an expansion path OR from the origin. This is divided into segments by the successive isoquants representing equal increments in output, i.e., 100, 200, 300 and so on. As we move

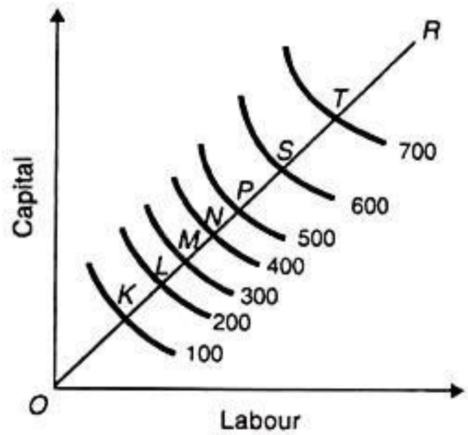


Fig. 13A

along the expansion path, the distance between the successive isoquants diminishes; it is a case of increasing returns to scale. This stage is shown in Figure 13(A) from K to M.

The distance between KL and LM becomes smaller LM < KL. The firm, therefore, requires smaller increases in the quantities of labour and capital to produce equal increments of output. If the segments between two isoquants are of equal length, there are constant returns to scale.

If labour and capital are doubled, the output would also be doubled. Thus when output increases from 300 to 400 and to 500 units, the isoquants representing these output levels mark off equal distances along the scale line, up to point P, i.e., MN = NP.

If there are decreasing returns to scale, the distance between a pair of isoquants would become longer on the expansion path. ST is longer than PS. It shows that to increase output larger increases in quantities of labour and capital are required. Thus, on the same expansion path from K to M, there are increasing returns to scale, from M to P, there are constant returns to scale and from P to T, and there are diminishing returns to scale.

UNIT – 4

COST ANALYSIS & PRICE OUTPUT DECISION

Concept of Cost

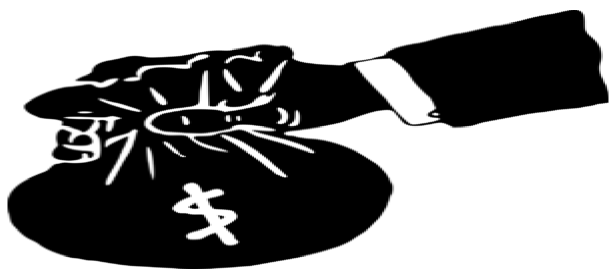
It is a commonly accepted fact that physical inputs or resources are important for enhancing production. We, however, tend to miss out on the financial aspect of this rule. Some of the most important decisions pertaining to business often relate to the cost of production, instead of physical resources themselves.

Hence, it is important for producers to understand cost analysis. Let’s understand the general concept of costs for that.

Concept of Costs

In order to understand the general concept of costs, it is important to know the following types of costs:

1. Accounting costs and Economic costs
2. Outlay costs and Opportunity costs
3. Direct/Traceable costs and Indirect/Untraceable costs
4. Incremental costs and Sunk costs
5. Private costs and Social costs
6. Fixed costs and Variable costs



Concept of Costs in terms of Treatment

1. Accounting costs

Accounting costs are those for which the entrepreneur pays direct cash for procuring resources for production. These include costs of the price paid for raw materials and machines, wages paid to workers, electricity charges, the cost incurred in hiring or purchasing a building or plot, etc. Accounting costs are treated as expenses. Chartered accountants record them in financial statements.

2. Economic costs

There are certain costs that accounting costs disregard. These include money which the entrepreneur forgoes but would have earned had he invested his time, efforts and investments in other ventures. For example, the entrepreneur would have earned an income had he sold his services to others instead of working on his own business Similarly, potential returns on the capital he employed in his business instead of giving it to others, the output generated by his resources which he could have used for others’ benefits, etc. are other examples of economic costs. Economic costs help the entrepreneur calculate supernormal profits, i.e. profits he would earn above the normal profits by investing in ventures other than his.

Concept of Costs in terms of the Nature of Expenses

1. Outlay costs

The actual expenses incurred by the entrepreneur in employing inputs are called outlay costs. These include costs on payment of wages, rent, electricity or fuel charges, raw materials, etc. We have to treat them are general expenses for the business.

2. Opportunity costs

Opportunity costs are incomes from the next best alternative that is foregone when the entrepreneur makes certain choices. For example, the entrepreneur could have earned a salary had he worked for others instead of spending time on his own business. These costs calculate the missed opportunity and calculate income that we can earn by following some other policy.

Concept of Costs in terms of Traceability

1. Direct costs

Direct costs are related to a specific process or product. They are also called traceable costs as we can directly trace them to a particular activity, product or process.

They can vary with changes in the activity or product. Examples of direct costs include manufacturing costs relating to production, customer acquisition costs pertaining to sales, etc.

2. Indirect costs

Indirect costs, or untraceable costs, are those which do not directly relate to a specific activity or component of the business. For example, an increase in charges of electricity or taxes payable on income. Although we cannot trace indirect costs, they are important because they affect overall profitability.

Concept of Costs in terms of the Purpose

1. Incremental costs

These costs are incurred when the business makes a policy decision. For example, change of product line, acquisition of new customers, upgrade of machinery to increase output are incremental costs.

2. Sunk costs

Suck costs are costs which the entrepreneur has already incurred and he cannot recover them again now. These include money spent on advertising, conducting research, and acquiring machinery.

Concept of Costs in terms of Payers

1. Private costs

These costs are incurred by the business in furtherance of its own objectives. Entrepreneurs spend them for their own private and business interests. For example, costs of manufacturing, production, sale, advertising, etc.

2. Social costs

As the name suggests, it is the society that bears social costs for private interests and expenses of the business. These include social resources for which the firm does not incur expenses, like atmosphere, water resources and environmental pollution.

Concept of Costs in terms of Variability

1. Fixed costs

Fixed costs are those which do not change with the volume of output. The business incurs them regardless of their level of production. Examples of these include payment of rent, taxes, interest on a loan, etc.

2. Variable costs

These costs will vary depending upon the output that the business generates. Less production will cost fewer expenses, and vice versa, the business will pay more when its production is greater. Expenses on the purchase of raw material and payment of wages are examples of variable costs.

Cost Function:

The relationship between output and costs is expressed in terms of cost function. By incorporating prices of inputs into the production function, one obtains the cost function since cost function is derived from production function. However, the nature of cost function depends on the time horizon. In microeconomic theory, we deal with short run and long run time.

A cost function may be written as:

Cq = f(Qf Pf)

Where Cq is the total production cost, Qf is the quantities of inputs employed by the firm, and Pf is the prices of relevant inputs. This cost equation says that cost of production depends on prices of inputs and quantities of inputs used by the firm.

Importance of Cost Function:

The study of business behaviour concentrates on the production process—the conversion of inputs into outputs—and the relationship between output and costs of production.

We have already studied a firm's production technology and how inputs are combined to produce output. The production function is just a starting point for the supply decisions of a firm. For any business decision, cost considerations play a great role.

Cost function is a derived function. It is derived from the production function which captures the technology of a firm. The theory of cost is a concern of managerial economics. Cost analysis helps allocation of resources among various alternatives. In fact, knowledge of cost theory is essential for making decisions relating to price and output.

Whether production of a new product is a wiser one on the part of a firm greatly depends on the evaluation of costs associated with it and the possibility of earning revenue from it. Decisions on capital investment (e.g., new machines) are made by comparing the rate of return from such investment with the opportunity cost of the funds used.

The relevance of cost analysis in decision-making is usually couched in terms of short and long periods of time by economists. In all market structures, short run costs are crucial in the determination of price and output. This is due to the fact that the basis for cost function is production and the prices of inputs that a firm pays.

On the other hand, long run cost analysis is used for planning the optimal scale of plant size. In other words, long run cost functions provide useful information for planning the growth as well as the investment policies of a firm. Growth of a firm largely depends on cost considerations.

The position of the U-shaped long run AC of a firm is suggestive of the direction of the growth of a firm. That is to say, a firm can take a decision whether to build up a new plant or to look for diversification in other markets by studying its existence on the long run AC curve. Further, it is the cost that decides the merger and takeover of a sick firm.

Non-profit sector or the government sector must also have a knowledge of cost function for decision-making. Whether the Narmada Dam is to be built or not, it should evaluate the costs and benefits 'flowing' from the dam.

Short Run Cost

Conceptually, in the short run, the quantity of at least one input is fixed and the quantities of the other inputs can be varied.

In the short-run period, factors, such as land and machinery, remain the same.

On the other hand, factors, such as labor and capital, vary with time. In the short run, the expansion is done by hiring more labor and increasing capital. The existing size of the plant or building cannot be increased in case of the short run.

Following are the cost concepts that are taken into consideration in the short run:

i. Total Fixed Costs (TFC):

Refer to the costs that remain fixed in the short period. These costs do not change with the change in the level of output. For example, rents, interest, and salaries. In the words of Ferguson, "Total fixed cost is the sum of the 'short run explicit fixed costs and implicit costs incurred by the entrepreneur." Fixed costs have implication even when the production of an organization is zero. These costs are also called supplementary costs, indirect costs, overhead costs, historical costs, and unavoidable costs.

TFC remains constant with respect to change in the level of output. Therefore, the slope of TFC curve is a horizontal straight line.

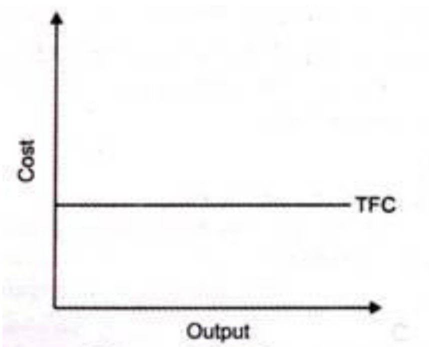


Figure-3: TFC Curve

Figure-3 depicts the TFC curve:

As shown in Figure-3, TFC curve is horizontal to x- axis. From Figure-3, it can be seen that TFC remains the same at all the levels with respect to change in the level of output.

ii. Total Variable Costs (TVC):

Refer to costs that change with the change in the level of production. For example, costs incurred on purchasing raw material, hiring labor, and using electricity. According to Ferguson, “total variable cost is the sum of amounts spent for each of the variable inputs used” If the output is zero, then the variable cost is also zero. These costs are also called prime costs, direct costs, and avoidable costs.

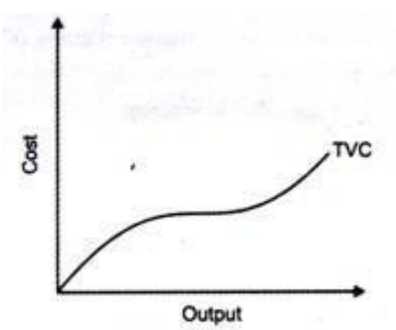


Figure-4: TVC Curve

Figure-4 shows the TVC curve:

In Figure-4, it can be seen that TVC curve changes with the change in the level of output.

iii. Total Cost (TC):

Involves the sum of TFC and TVC.

It can be calculated as follows:

Total Cost = TFC + TVC

TC also changes with the changes in the level of output as there is a change in TVC.

Figure-5 shows the total cost curve derived from sum of TVC and TFC:

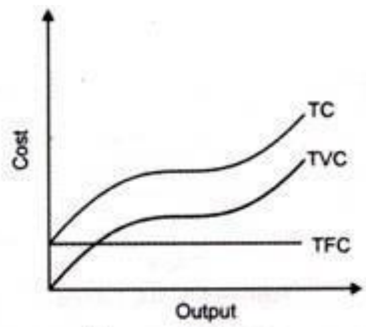


Figure-5: TC Curve

It should be noted that both TVC and TC increase initially at decreasing rate and then they increase at increasing rate Here, decreasing rate implies that the rate at which cost increases with respect to output is less, whereas increasing rate implies the rate at which cost increases with respect to output is more.

iv. Average Fixed Costs (AFC):

Refers to the per unit fixed costs of production. In other words, AFC implies fixed cost of production divided by the quantity of output produced.

It is calculated as:

AFC = TFC/Output

TFC is constant as production increases, thus AFC falls.

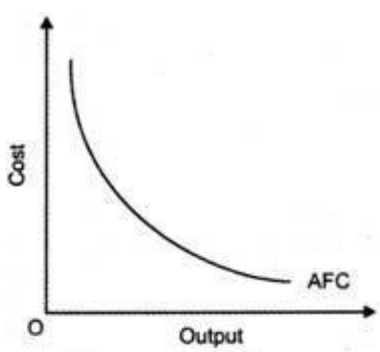


Figure-6: AFC Curve

Figure-6 shows the AFC curve:

In Figure-6 AFC curve is shown as a declining curve, which never touches the horizontal axis. This is because fixed cost can never be zero. The curve is also called rectangular hyperbola, which represents that total fixed costs remain same at all the levels.

v. Average Variable Costs (AVC):

Refer to the per unit variable cost of production. It implies organization's variable costs divided by the quantity of output produced.

It is calculated as:

AVC = TVC/ Output

Initially, AVC decreases as output increases. After a certain point of time, AVC increases with respect to increase in output.

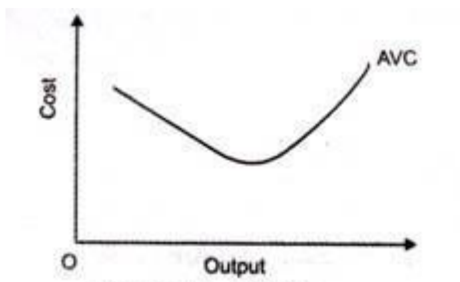


Figure -7: AVC Curve

Thus, it is a U- shaped curve, as shown in Figure-7:

vi. Average Cost (AC):

Refer to the total costs of production per unit of output.

AC is calculated as:

AC = TC/ Output

AC is also equal to the sum total of AFC and AVC. AC curve is also U-shaped curve as average cost initially decreases when output increases and then increases when output increases.

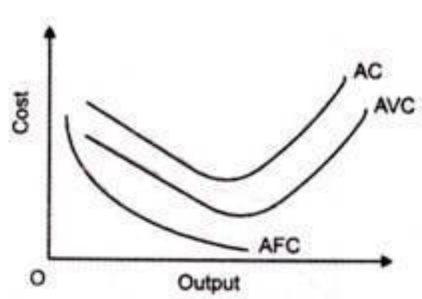


Figure -8: AC Curve

Figure-8 shows the AC curve:

vii. Marginal Cost:

Refer to the addition to the total cost for producing an additional unit of the product.

Marginal cost is calculated as:

MC = TCn = TCn-1

n= Number of units produced

It is also calculated as:

MC = ΔTC/ΔOutput

MC curve is also a U-shaped curve as marginal cost initially decreases as output increases and afterwards, rises as output increases. This is because TC increases at decreasing rate and then increases at increasing rate.

Figure- 9 shows the MC curve:

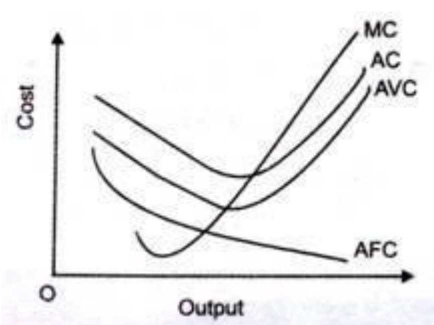


Figure-9: MC Curve

Let us learn the aforementioned cost concepts numerically with the help of Table-1:

Table-1: Calculation of Short-run Costs

Units of Output	TFC =30	TVC	TC= TFC + TVC	AFC= TFC/Output	AVC= TVC/Output	AC= AFC+ AVC	MC
0	30	0	30	-	-	-	-
1	30	10	40	30	10	40	10
2	30	18	48	15	9	24	8
3	30	24	54	10	8	18	6
4	30	32	62	7.5	8	15.5	8
5	30	50	80	6	10	16	18
6	30	72	102	5	12	17	22

Long Run Cost

In the long run, all the factors of production used by an organization vary. The existing size of the plant or building can be increased in case of long run.

There are no fixed inputs or costs in the long run. Long run is a period in which all the costs change as all the factors of production are variable.

There is no distinction between the Long run Total Costs (LTC) and long run variable cost as there are no fixed costs. It should be noted that the ability of an organization of changing inputs enables it to produce at lower cost in the long run.

1. Long Run Total Cost:

Long run Total Cost (LTC) refers to the minimum cost at which given level of output can be produced. According to Leibhafasky, "the long run total cost of production is the least possible cost of producing any given level of output when all inputs are variable." LTC represents the least cost of different quantities of output. LTC is always less than or equal to short run total cost, but it is never more than short run cost.

The LTC curve is shown in Figure-10:

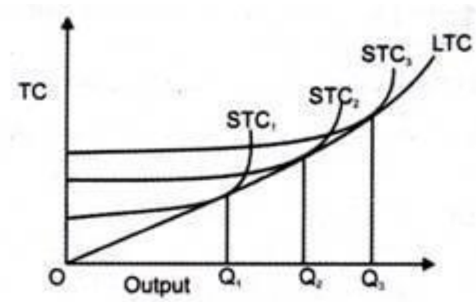


Figure-10: LTC Curve

As shown in Figure-10, short run total costs curves; STC1, STC2, and STC3 are shown depicting different plant sizes. The LTC curve is made by joining the minimum points of short run total cost curves. Therefore, LTC envelopes the STC curves.

2. Long Run Average Cost:

Long run Average Cost (LAC) is equal to long run total costs divided by the level of output. The derivation of long run average costs is done from the short run average cost curves. In the short run, plant is fixed and each short run curve corresponds to a particular plant. The long run average costs curve is also called planning curve or envelope curve as it helps in making organizational plans for expanding production and achieving minimum cost.

Figure-11 shows the derivation of LAC curve:

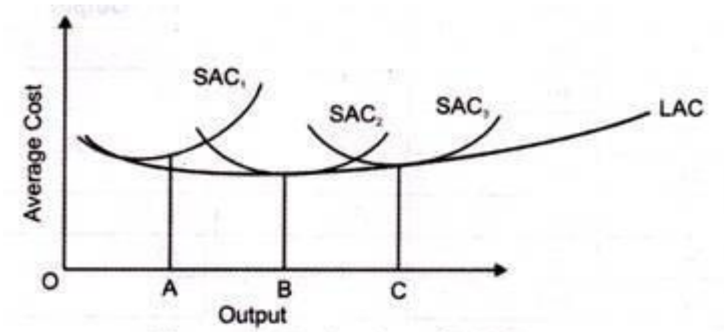


Figure-11: Derivation of LAC Curve

Suppose there are three sizes of the plant and no other size of the plant can be built. In short run, the plant sizes are fixed thus, organization increase or decrease the variable factors. However, in the long run, the organization can select among the plants which help in achieving minimum possible cost at a given level of output.

From Figure-11, it can be noted that till OB amount of production, it is beneficial for the organization to operate on the plant SAC2 as it entails lower costs than SAC1. If the plant SAC2 is used for producing OA, then cost incurred would be more. Thus, in the long run, it is clear that the producer would produce till OB on plant SAC2. On SAC2, the producer would produce till OC amount of output. If an organization wants to exceed output from OC, it will be beneficial to produce at SAC3 than SAC2.

Thus, in the long run, an organization has a choice to use the plant incurring minimum costs at a given output. LAC depicts the lowest possible average cost for producing different levels of output. The LAC curve is derived from joining the lowest minimum costs of the short run average cost curves.

It first falls and then rises, thus it is U- shaped curve. The returns to scale also affect the LTC and LAC. Returns to scale implies a change in output of an organization with a change in inputs. In the long run, the output changes with respect to change in all inputs of production.

In case of increasing returns to scale (IRS), organizations can double the output by using less than twice of inputs. LTC increases less than the increase in the output, thus, LAC falls. In case of constant returns to scale (CRS), organizations can double the output by using inputs twice.

LTC increases proportionately to the output; therefore, LAC becomes constant. On the other hand, in case of decreasing returns to scale (DRS), organizations can double the output by using inputs more than twice. Thus, LTC increases more than the increase in output. As a result, LAC increases.

Figure-12 shows the effect on LAC because of returns to scale:

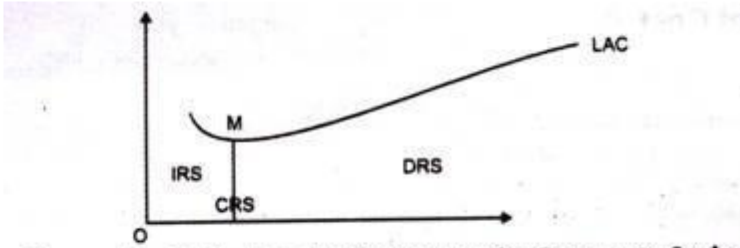


Figure-12: Derivation of LAC curve under Returns to Scale

As shown in Figure-12, up to M, LAC slopes downward. This is because at this stage IRS is applied. On the other hand, at M, LAC becomes constant. After M, LAC slopes upwards implying DRS.

3. Long Run Marginal Cost:

Long run Marginal Cost (LMC) is defined as added cost of producing an additional unit of a commodity when all inputs are variable. This cost is derived from short run marginal cost. On the graph, the LMC is derived from the points of tangency between LAC and SAC.

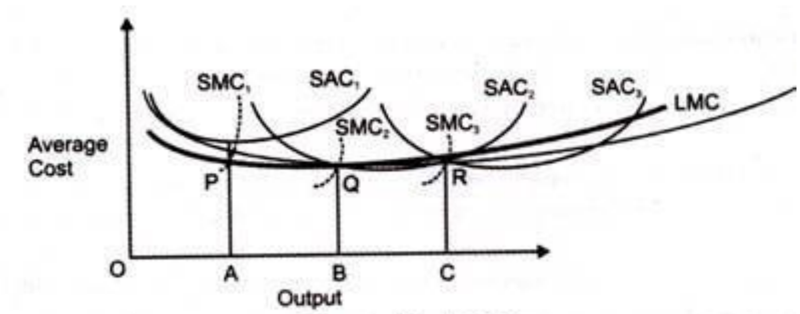


Figure-13: LMC Curve

LMC curve can be learned through Figure-13:

If perpendiculars are drawn from point A, B, and C, respectively; then they would intersect SMC curves at P, Q, and R respectively. By joining P, Q, and R, the LMC curve would be drawn. It should be noted that LMC equals to SMC, when LMC is tangent to the LAC.

In Figure-13, OB is the output at which:

SAC2 = SMC2 = LAC = LMC

We can also draw the relation between LMC and LAC as follows:

When LMC < LAC, LAC falls When LMC = LAC, LAC is constant When LMC > LAC, LAC rises

Economies and Diseconomies of Scale

Economies of scale are defined as the cost advantages that an organization can achieve by expanding its production in the long run.

In other words, these are the advantages of large scale production of the organization. The cost advantages are achieved in the form of lower average costs per unit.

It is a long term concept. Economies of scale are achieved when there is an increase in the sales of an organization. As a result, the savings of the organization increases, which further enables the organization to obtain raw materials in bulk. This helps the organization to enjoy discounts. These benefits are called as economies of scale.

The economies of scale are divided in to internal economies and external economies discussed as follows:

i. Internal Economies:

Refer to real economies which arise from the expansion of the plant size of the organization. These economies arise from the growth of the organization itself.

The examples of internal economies of scale are as follows:

a. Technical economies of scale:

Occur when organizations invest in the expensive and advanced technology. This helps in lowering and controlling the costs of production of organizations. These economies are enjoyed because of the technical efficiency gained by the organizations. The advanced technology enables an organization to produce a large number of goods in short time. Thus, production costs per unit falls leading to economies of scale.

b. Marketing economies of scale:

Occur when large organizations spread their marketing budget over the large output. The marketing economies of scale are achieved in case of bulk buying, branding, and advertising. For instance, large organizations enjoy benefits on advertising costs as they cover larger audience. On the other hand, small organizations pay equal advertising expenses as large organizations, but do not enjoy such benefits on advertising costs.

c. Financial economies of scale:

Take place when large organizations borrow money at lower rate of interest. These organizations have good credibility in the market. Generally, banks prefer to grant loans to those organizations that have strong foothold in the market and have good repaying capacity.

d. Managerial economies of scale:

Occur when large organizations employ specialized workers for performing different tasks. These workers are experts in their fields and use their knowledge and experience to maximize the profits of the organization. For instance, in an organization, accounts and research department are created and managed by experienced individuals, SO that all costs and profits of the organization can be estimated properly.

e. Commercial economies:

Refer to economies in which organizations enjoy benefits of buying raw materials and selling of finished goods at lower cost. Large organizations buy raw materials in bulk; therefore, enjoy benefits in transportation charges, easy credit from banks, and prompt delivery of products to customers.

ii. External economies:

Occur outside the organization. These economies occur within the industries which benefit organizations. When an industry expands, organizations may benefit from better transportation network, infrastructure, and other facilities. This helps in decreasing the cost of an organization. Some of the examples of external economies of scale are discussed as follows:

a. Economies of Concentration:

Refer to economies that arise from the availability of skilled labor, better credit, and transportation facilities.

b. Economies of Information:

Imply advantages that are derived from publication related to trade and business. The central research institutions are the source of information for organizations.

c. Economies of Disintegration:

Refer to the economies that arise when organizations split their processes into different processes.

Diseconomies of scale occur when the long run average costs of the organization increases. It may happen when an organization grows excessively large. In other words, the diseconomies of scale cause larger organizations to produce goods and services at increased costs.

There are two types of diseconomies of scale, namely, internal diseconomies and external diseconomies, discussed as follows:

i. Internal diseconomies of scale:

Refer to diseconomies that raise the cost of production of an organization. The main factors that influence the cost of production of an organization include the lack of decision, supervision, and technical difficulties.

ii. External diseconomies of scale:

Refer to diseconomies that limit the expansion of an organization or industry. The factors that act as restraint to expansion include increased cost of production, scarcity of raw materials, and low supply of skilled laborer.

There are a number of causes for diseconomies of scale.

Some of the causes which lead to diseconomies of scale are as follows:

i. Poor Communication:

Act as a major reason for diseconomies of scale. If production goals and objectives of an organization are not properly communicated to employees within the organization, it may lead to overproduction or production. This may lead to diseconomies of scale.

Apart from this, if the communication process of the organization is not strong then the employees would not get adequate feedback. As a result, there would be less face-to-face interaction among employees- thus the production process would be affected.

ii. Lack of Motivation:

Leads to fall in productivity levels. In case of a large organization, workers may feel isolated and are less appreciated for their work, thus their motivation diminishes. Due to poor communication network, it is harder for employers to interact with the employees and build a sense of belongingness. This leads to fall in the productivity levels of output owing to lack of motivation. This further leads to increase in costs of the organization.

iii. Loss of Control:

Acts as the main problem of large organizations. Monitoring and controlling the work of every employee in a large organization becomes impossible and costly. It is harder to make out that all the employees of an organization are working towards the same goal. It becomes difficult for managers to supervise the sub-ordinates in large organizations.

iv. Cannibalization:

Implies a situation when an organization faces competition from its own product. A small organization faces competition from products of other organizations, whereas sometimes large organizations find that their own products are competing with each other.

Explicit Cost & Implicit Cost

When most people think of cost, they think only of the explicit cost — the actual payment by firms to labour, capital, and other factors of production. Whether these costs are fixed or variable, they are straightforward: they are the amounts that firms must pay to owners of the resources in order to bid these resources away from alternative uses.

Producers also incur some costs referred to as implicit costs and in any complete analysis of costs one must take implicit costs into consideration. Managerial decision makers should do the same. To aid in analysing the nature of implicit costs, consider two firms that produce identical amounts of the good. The owner of one firm rents the building in which the good is produced.

The owner of the other firm inherited the building the firm uses and therefore pays no rent. Whose costs are higher? For decision making purposes, the costs for both are the same even though the second firm makes lower payments to outside factors of production. The reason for the same costs is that using the building to produce goods costs the second firm the amount of income that could have been earned had it been leased at the prevailing rent.

Since these two buildings are the same, presumably the market rentals would be the same. In other words, a part of the cost incurred by the second firm is the (implicit) payment from firm owners to themselves as the owners of a resource (the building). Thus, one implicit cost would include what the owner of a firm could make from selling or leasing the capital he own instead of using it in his own firm.

Another implicit cost would include the value of the firm owner’s time that is used to manage the business. Presumably, if the owner of a firm were not managing the business or working for the firm in another way, he or she could obtain a job with some other firm, possibly as a manager. The salary that could be earned in this alternative occupation is an implicit cost that should be considered as part of the total cost of production. These implicit costs are just as real as explicit costs.

The implicit costs incurred by firms in producing a specific commodity consist of the amounts that could be earned in the best alternative use of the owner-manager’s time and of any other of his resources currently used to produce the commodity in question. Implicit costs must be added to explicit costs in order to obtain total costs.

In short, explicit cost is called outlay cost and refers to any payment to an outsider and is reflected in a company’s book of account. By contrast, implicit cost is opportunity cost and is not taken into consideration by the accountant. But in economics it is added to explicit cost and iscalled normal profit. Due to inclusion of implicit cost economic profit is less than accounting profit.

Comparison Chart

BASIS FOR COMPARISON	EXPLICIT COST	IMPLICIT COST

Meaning	The costs which involve outflow of cash due to the use of factors of production is known as Explicit Cost.	The costs in which there is no cash outlay, is known as Implicit Cost.
Alternatively known as	Out-of-pocket Costs	Imputed Costs
Occurrence	Actual	Implied
Recording and Reporting	Yes	No
Estimation of Cost	Objective	Subjective
Which profit can be calculated with the help of cost?	Accounting Profit and Economic Profit	Economic Profit
Example	Salaries, rent, advertisement, wages, etc.	Interest on owner's capital, Salary to owner, rent of owner's building, etc. which do not occur in reality.

Private & Social Cost

Private Cost:

Private cost refers to the cost of production incurred and provided for by an individual firm engaged in the production of a commodity. It is found out to get private profits.

This cost has nothing to do with the society. It includes both explicit as well as implicit cost. A firm is interested in minimising private cost.

Social Cost:

Social cost refers to the cost of producing a commodity to the society as a whole. It takes into consideration all those costs, which are borne by the society directly or indirectly. Social cost is not borne by the firm. It is rather passed on to persons not involved in the activity in the direct way. Social cost is a much broader concept.

It is found out to get social profits rather than private profits. The production of a commodity by a firm generates advantages (benefits) as well as disadvantages (cost) to other members of society, called external benefits and external costs respectively.

These benefits are available free of cost. For instance, to facilitate easier movement of raw materials and finished products, a producer constructs a road, linking it with a highway. This road may be used by others, who will not pay for the benefits derived. On the similar lines, no producer compensates others for the costs incurred to them as a result of his production.

Water pollution caused by the disposal of wastes into a river (or sea) or air pollution and consequent health hazards by the smoke generation by factories or buses plying in big cities are some other examples.

Noise pollution and accident proneness are some other social costs due to rising traffic in big cities. While computing social costs, market prices of goods and factor of production are adjusted as social and shadow prices.

Social cost is the sum of private cost and external cost. Alternatively, external cost is the difference between social cost and private cost, which may be positive or negative. If social cost is more than private cost, there is an external cost (or. negative externality). On the other hand, if social cost is less than private cost, there is an external benefit (or positive externality).

Social cost is an important concept. Knowledge of social cost and social benefit is extremely important in the efficient utilisation of limited resources. The concept of social cost can be linked with opportunity cost to which we now turn.

Pricing under Perfect Competition

Meaning and Definition of Perfect Competition:

A Perfect Competition market is that type of market in which the number of buyers and sellers is very large, all are engaged in buying and selling a homogeneous product without any artificial restrictions and possessing perfect knowledge of the market at a time.

In other words it can be said—"A market is said to be perfect when all the potential buyers and sellers are promptly aware of the prices at which the transaction take place. Under such conditions the price of the commodity will tend to be equal everywhere."

In this connection Mrs. Joan Robinson has said—"Perfect Competition prevails when the demand for the output of each producer is perfectly elastic."

According to Boulding—"A Perfect Competition market may be defined as a large number of buyers and sellers all engaged in the purchase and sale of identically similar commodities, who are in close contact with one another and who buy and sell freely among themselves."

Characteristics of Perfect Competition:

The following characteristics are essential for the existence of Perfect Competition:

1. Large Number of Buyers and Sellers:

The first condition is that the number of buyers and sellers must be so large that none of them individually is in a position to influence the price and output of the industry as a whole. In the market the position of a purchaser or a seller is just like a drop of water in an ocean.

2. Homogeneity of the Product:

Each firm should produce and sell a homogeneous product so that no buyer has any preference for the product of any individual seller over others. If goods will be homogeneous then price will also be uniform everywhere.

3. Free Entry and Exit of Firms:

The firm should be free to enter or leave the firm. If there is hope of profit the firm will enter in business and if there is profitability of loss, the firm will leave the business.

4. Perfect Knowledge of the Market:

Buyers and sellers must possess complete knowledge about the prices at which goods are being bought and sold and of the prices at which others are prepared to buy and sell. This will help in having uniformity in prices.

5. Perfect Mobility of the Factors of Production and Goods:

There should be perfect mobility of goods and factors between industries. Goods should be free to move to those places where they can fetch the highest price.

6. Absence of Price Control:

There should be complete openness in buying and selling of goods. Here prices are liable to change freely in response to demand and supply conditions.

7. Perfect Competition among Buyers and Sellers:

In this purchasers and sellers have got complete freedom for bargaining, no restrictions in charging more or demanding less, competition feeling must be present there.

8. Absence of Transport Cost:

There must be absence of transport cost. In having less or negligible transport cost will help complete market in maintaining uniformity in price.

9. One Price of the Commodity:

There is always one price of the commodity available in the market.

10. Independent Relationship between Buyers and Sellers:

There should not be any attachment between sellers and purchasers in the market. Here, the seller should not show prick and choose method in accepting the price of the commodity. If we will see from the close we will find that in real life "Perfect Competition is a pure myth."

Factors Affecting Price Determination:

Let us discuss these two factors:

(i) Total Demand:

Demand does not mean the amount of a commodity, say, tea, which people need, or would like to have, but the effective demand, the amount which people are willing to buy at various prices. Every unit of a commodity has a demand price, the price at which that unit finds a buyer.

Other things being equal, a fall in price leads to an increase in quantity demanded and rise in price leads to reduction in quantity demanded. A fall in the price induces the existing buyers to buy more and attracts new customers. It may also have substitution and income effects. The quantity demanded thus varies with every change in price.

(ii) Total Supply:

Supply means the quantity offered for sale by producers. Thus the supply of tea does not mean the actual stock of tea; it means the amount of tea which the producers are willing to put on the market at various prices.

Every unit of a commodity has a supply price-the price at which it is offered for sale. Other things being equal, a rise in price leads to an increase in the quantity offered for sale, and a fall in price reduces the quantity offered for sale.

Equilibrium between Demand & Supply Prices:

Equilibrium between demand & supply price is obtained by the interaction of these two forces.

Table 1

Equilibrium between Demand and Supply

Price	Demand	Supply		Pressure on price
5	12	1		
10	10	2	D > S	Excess
15	8	4		Demand
20	6	6	D = S	Equilibrium
25	4	8		
30	2	10	D < S	Excess
35	1	12		Supply

It is being discussed with the help of a table 1 and a diagram 1 below:

In the table 1, at price Rs. 20 the quantity demanded and quantity supplied are equal to each other. Thus, the price Rs. 20 will be settled down in the market where both the buyers and sellers are satisfied. Once this equilibrium price prevails in the market, there will be no tendency for it to change.

Now, suppose the price is Rs. 25 in the market, the buyers would demand only four units while the sellers are ready to supply 8 units. For this the sellers will compete with each other to sell more units of the commodity.

This process will continue, till the equilibrium price is reached. On the other hand, if the price falls below the equilibrium level say Rs. 15 the buyers will demand 8 units and the sellers will be ready to supply 4 units.

Diagrammatic Representation:

In figure 1, DD is the demand curve, SS is the supply curve. Demand and supply are in equilibrium at point E where both curves intersect each other.

Here, equilibrium output is OX1 and price is OP. Now, suppose the price is greater than equilibrium price i.e. OP2. At this price quantity demanded is P2D2, while the quantity supplied is P2S2. Thus, D2S2 is the excess supply which the buyers will not take off the market at price OP2. In order to dispose of the said excess supply, the seller will compete with each other and will bring down the price.

Hence, the price will fall to the level of OP. At price OP1 the buyers will demand P1D1 while the sellers are prepared to supply P1 S1. In this way, unsatisfied buyers will compete with each other to have the limited supply. Therefore, there will be tendency for the price to rise to the level of OP.

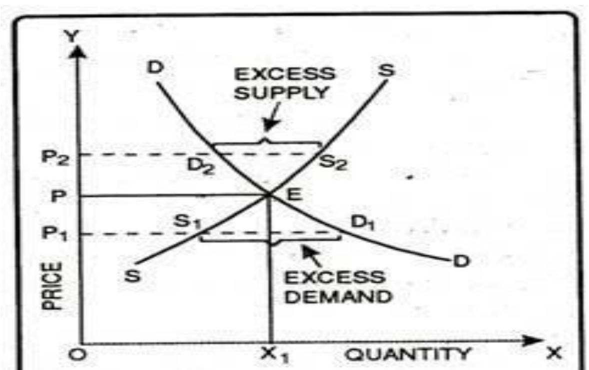


Fig. 1.

Equilibrium Price of Firm and Industry:

Prices under perfect competition are determined by industry and each firm will have to follow this price. It can be shown with the help of fig. 2.

In Fig. 2A demand curve DD cuts the industry supply curve SS at point E. Thus, E is the equilibrium point which determines OP as an equilibrium price. Fig. 2B reflects the firms demand curve. The firm will have to sell all its output at OP price.

The firm cannot increase or decrease their price as it is determined by industry. It is so because under perfect competition firm is a price taker and not a price maker.

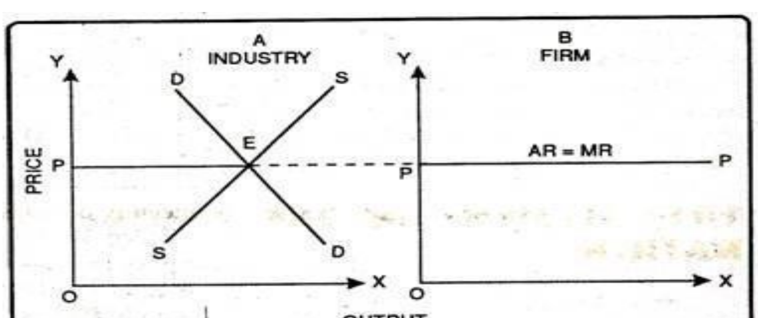


Fig. 2.

Changes in Equilibrium:

The equilibrium price and quantity will change as a result of shift either in demand curve or supply curve or both.

It can be shown as follows:

(i) Effect of Shift in Demand:

Demand changes whenever there is a change in the conditions of demand- income, tastes, prices of substitutes and complements etc. If demand increases due to a change in any one of these conditions the demand curve moves upward to the right. If, on the other hand, demand decreases, the demand curve moves downward to the left. Let us consider the effects of changes in demand on price increases (i.e., at the old price a greater quantity is demanded) price and quantity demanded both increase.

Let us show the effect of a change in demand on price and quantity. Given the supply curve, an upward shift in the demand curve causes the price to rise. A downward shift of the demand curve causes the price to fall.

This is shown in the figure given below:

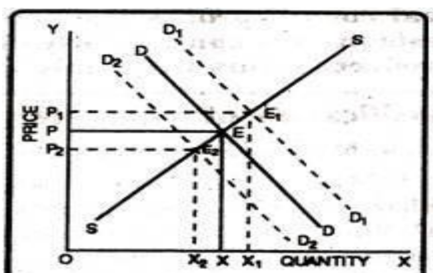


Fig. 3.

In figure 3 Quantity demanded and supplied is shown on X-axis whereas price on Y-axis. DD is the original demand curve. Initially, the demand and supply curves intersect each other at point

E. Thus, OP is the equilibrium price and OX is the equilibrium quantity demanded.

Now, suppose the demand increases and takes the form of D1 D1. The new demand curve intersects the supply curve at point E1, where the quantity demanded goes to OX1 and price at OP1. In the same way, if the demand curve shifts downward and assumes the form D2D2, the equilibrium will be at point E2 and, therefore, quantity demanded will be OX2 and price is OP2. Thus, we conclude that if the supply conditions are given then there is direct relation between demand and price.

(ii) Effect of Shift in Supply:

We shall consider the effects of changes in supply on price assuming that demand remains unchanged. An increase in supply is represented by a shift of the supply curve to the right (or downward) and a decrease in supply is represented by a shift to the left or upward. The general rule is: if the supply increases, price falls and if supply decreases price rises.

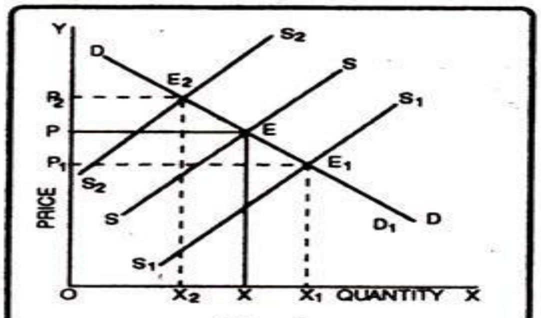


Fig. 4.

In figure 4 both the demand and supply curves intersect each other at point E. Thus, OP is the equilibrium price and OX is the equilibrium quantity. Now, supply curve shifts from SS to S1S1– At this level quantity demanded increases form OX to OX1 but the price falls from OP to OP1. Similarly, if the supply curve takes the form of S2S2 instead of SS, that way quantity demanded falls to OX2 while die price increases from OP to OP2– Here; we may observe that there is inverse correlation between quantity demanded and price.

Pricing under Monopoly

Monopoly: Meaning, Definitions & Features:
Meaning:

The word monopoly has been derived from the combination of two words i.e., 'Mono' and 'Poly'. Mono refers to a single and poly to control.
In this way, monopoly refers to a market situation in which there is only one seller of a commodity.
There are no close substitutes for the commodity it produces and there are barriers to entry. The single producer may be in the form of individual owner or a single partnership or a joint stock company. In other words, under monopoly there is no difference between firm and industry.
Monopolist has full control over the supply of commodity. Having control over the supply of the commodity he possesses the market power to set the price. Thus, as a single seller, monopolist may be a king without a crown. If there is to be monopoly, the cross elasticity of demand between the product of the monopolist and the product of any other seller must be very small.

Definitions:

“Pure monopoly is represented by a market situation in which there is a single seller of a product for which there are no substitutes; this single seller is unaffected by and does not affect the prices and outputs of other products sold in the economy.” Bilas

“Monopoly is a market situation in which there is a single seller. There are no close substitutes of the commodity it produces, there are barriers to entry”. -Koutsoyiannis

“Under pure monopoly there is a single seller in the market. The monopolist demand is market demand. The monopolist is a price-maker. Pure monopoly suggests no substitute situation”. -A.
J. Braff

“A pure monopoly exists when there is only one producer in the market. There are no dire competitions.” -Ferguson

“Pure or absolute monopoly exists when a single firm is the sole producer for a product for which there are no close substitutes.” -McConnel

Features:

We may state the features of monopoly as:

1. One Seller and Large Number of Buyers:

The monopolist’s firm is the only firm; it is an industry. But the number of buyers is assumed to be large.

2. No Close Substitutes:

There shall not be any close substitutes for the product sold by the monopolist. The cross elasticity of demand between the product of the monopolist and others must be negligible or zero.

3. Difficulty of Entry of New Firms:

There are either natural or artificial restrictions on the entry of firms into the industry, even when the firm is making abnormal profits.

4. Monopoly is also an Industry:

Under monopoly there is only one firm which constitutes the industry. Difference between firm and industry comes to an end.

5. Price Maker:

Under monopoly, monopolist has full control over the supply of the commodity. But due to large number of buyers, demand of any one buyer constitutes an infinitely small part of the total demand. Therefore, buyers have to pay the price fixed by the monopolist.

Under monopoly, for the equilibrium and price determination there are two different conditions which are:

1. Marginal revenue must be equal to marginal cost.
2. MC must cut MR from below.

However, there are two approaches to determine equilibrium price under monopoly viz.;

1. Total Revenue and Total Cost Approach.
2. Marginal Revenue and Marginal Cost Approach.

Total Revenue and Total Cost Approach:

Monopolist can earn maximum profits when difference between TR and TC is maximum. By fixing different prices, a monopolist tries to find out the level of output where the difference between TR and TC is maximum. The level of output where monopolist earns maximum profits is called the equilibrium situation. This can be explained with the help of fig. 2.

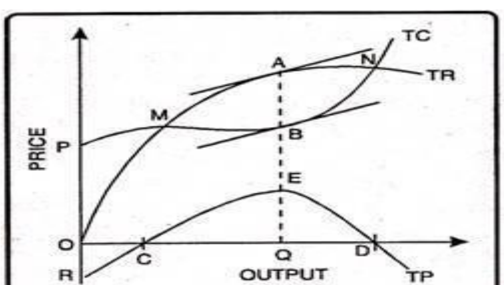


Fig. 2.

In Fig. 2, TC is the total cost curve. TR is the total revenue curve. TR curve starts from the origin. It indicates that at zero level of output, TR will also be zero. TC curve starts from P. It reflects that even if the firm discontinues its production, it will have to suffer the loss of fixed costs.

Total profits of the firm are represented by TP curve. It starts from point R showing that initially firm is faced with negative profits. Now as the firm increases its production, TR also increases. But in the initial stage, the rate of increase in TR is less than TC.

Therefore, RC part of TP curve reflects that firm is incurring losses. At point M, total revenue is equal to total cost. It shows that firm is working under no profit, no loss basis. Point M is called the breakeven point. When firm produces more than point M, TR will be more than TC. TP curve also slopes upward. It shows that firm is earning profit. Now as the TP curve reaches point E then the firm will be earning maximum profits. This amount of output will be termed as equilibrium output.

Marginal Revenue and Marginal Cost Approach:

According to marginal revenue and marginal cost approach, a monopolist will be in equilibrium when two conditions are fulfilled i.e., (i) MC=MR and (ii) MC must cut MR from below. The study of equilibrium price according to this analysis can be conducted in two time periods.

1. The Short Run
2. The Long Run

1. Short Run Equilibrium under Monopoly:

Short period refers to that period in which the monopolist has to work with a given existing plant. In other words, the monopolist cannot change the fixed factors like, plant, machinery etc. in the short period. Monopolist can increase his output by changing the variable factors. In this period, the monopolist can enjoy super-normal profits, normal profits and sustain losses.

These three possibilities are described as follows:

(a) Super Normal Profits:

If the price determined by the monopolist in more than AC, he will get super normal profits. The monopolist will produce up to the level where MC=MR. This limit will indicate equilibrium output. In Figure 3 output is measured on X-axis and price on Y-axis. SAC and SMC are the short run average cost and marginal cost curves while AR or MR are the average revenue or marginal revenue curves respectively.

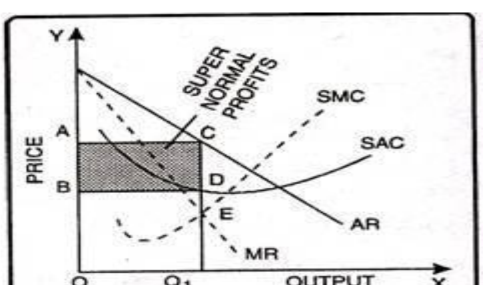


Fig. 3.

The monopolist is in equilibrium at point E because at point E both the conditions of equilibrium are fulfilled i.e., MR = MC and MC intersects the MR curve from below. At this level of equilibrium the monopolist will produce OQ1 level of output and sells it at CQ1 price which is more than average cost DQ1 by CD per unit. Therefore, in this case total profits of the monopolist will be equal to shaded area ABDC.

(b) Normal Profits:

A monopolist in the short run would enjoy normal profits when average revenue is just equal to average cost. We know that average cost of production is inclusive of normal profits. This situation can be illustrated with the help of fig 4.

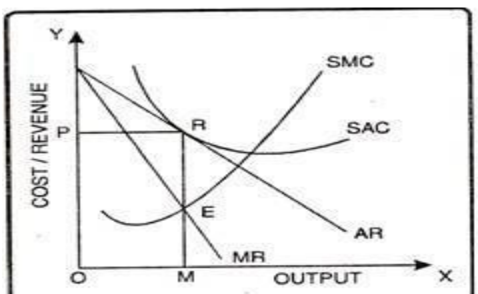


Fig. 4.

In Fig. 4 the firm is in equilibrium at point E. Here marginal cost is equal to marginal revenue. The firm is producing OM level of output. At OM level of output average cost curve touches the average revenue curve at point P. Therefore, at point ‘P’ price OR is equal to average cost of the total product. In this way, monopoly firm enjoys the normal profits.

© Minimum Losses:

In the short run, the monopolist may have to incur losses. This situation occurs if in the short run price falls below the variable cost. In other words, if price falls due to depression and fall in demand, the monopolist will continue to produce as long as price covers the average variable cost. Once the price falls

Below the average variable cost, monopolist will stop production. Thus, a monopolist in the short run equilibrium has to bear the minimum loss equal to fixed costs. Therefore, equilibrium price will be equal to average variable cost. This situation can also be explained with the help of Fig. 5.

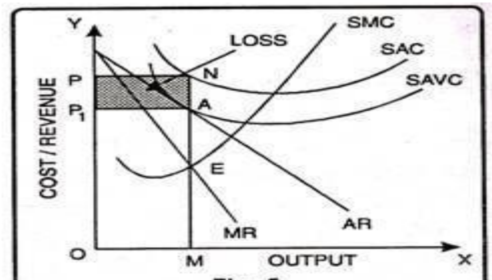


Fig. 5.

In Fig. 5 monopolist is in equilibrium at point E. At point E marginal cost is equal to marginal revenue and he produces OM level of output. At OM level of output, equilibrium price fixed by the monopolist is OP1. At OP1 price, AVC touches the AR curve at point A.

It signifies that the firm will cover only average variable cost from the prevailing price. At OP1 price, firm will bear loss of fixed cost i.e., A per unit. The firm will bear the total loss equal to the shaded area PP1 AN. Now if the price falls below OP1, the monopolist will stop production. It is so because if he continues production, he will have to bear the loss of variable costs along with fixed costs.

2. Long Run Equilibrium under Monopoly:

Long-run is the period in which output can be changed by changing the factors of production. In other words, all variable factors can be changed and monopolist would choose that plant size which is most appropriate for specific level of demand. Here, equilibrium would be attained at that level of output where the long-run marginal cost cuts marginal revenue curve from below. This can be shown with the help of Fig. 6.

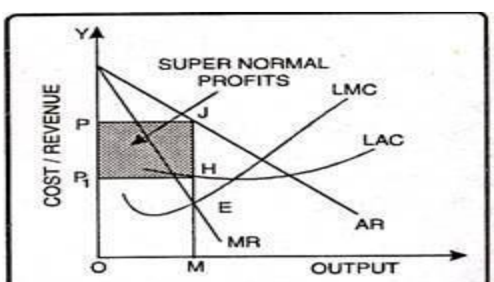


Fig. 6.

In Fig. 6 monopolist is in equilibrium at OM level of output. At OM level of output marginal revenue is equal to long run marginal cost and the monopolist fixes OP price. HM is the long run average cost? Price OP being more than LAC i.e., HM which fetch the monopolist super normal profits. Accordingly, the monopolist earns JM – HM = JH super normal profit per unit. His total super normal profits will be equal to shaded area PJHP1.

Price Discrimination

In monopoly, there is a single seller of a product called monopolist. The monopolist has control over pricing, demand, and supply decisions, thus, sets prices in a way, so that maximum profit can be earned.

The monopolist often charges different prices from different consumers for the same product. This practice of charging different prices for identical product is called price discrimination.

According to Robinson, “Price discrimination is charging different prices for the same product or same price for the differentiated product.”

According to Stigler, “Price discrimination is the sale of various products at prices which are not proportional to their marginal costs.”

In the words of Dooley, “Discriminatory monopoly means charging different rates from different customers for the same good or service.”

According to J.S. Bains, “Price discrimination refers strictly to the practice by a seller to charging different prices from different buyers for the same good.”

Let us learn different types of price discrimination.

Types of Price Discrimination:

Price discrimination is a common pricing strategy’ used by a monopolist having discretionary pricing power. This strategy is practiced by the monopolist to gain market advantage or to capture market position.

There are three types of price discrimination, which are shown in Figure-13:

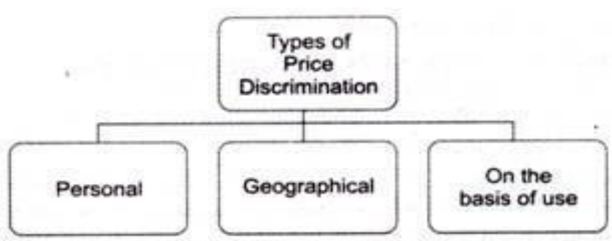


Figure-13: Price Discrimination

The different types of price discrimination (as shown in Figure-13) are explained as follows:

i. Personal:

Refers to price discrimination when different prices are charged from different individuals. The different prices are charged according to the level of income of consumers as well as their willingness to purchase a product. For example, a doctor charges different fees from poor and rich patients.

ii. Geographical:

Refers to price discrimination when the monopolist charges different prices at different places for the same product. This type of discrimination is also called dumping.

iii. On the basis of use:

Occurs when different prices are charged according to the use of a product. For instance, an electricity supply board charges lower rates for domestic consumption of electricity and higher rates for commercial consumption.

Degrees of Price Discrimination:

Price discrimination has become widespread in almost every market. In economic jargon, price discrimination is also called monopoly price discrimination or yield management. The degree of price discrimination vanes in different markets.

Figure-14 shows the degrees of price discrimination:

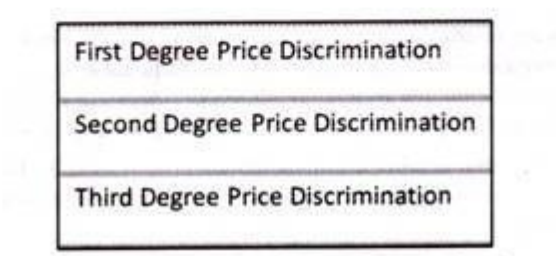


Figure-14: Degrees of Price Discrimination

These three degrees of price discrimination (as shown in Figure-14) are explained as follows:

i. First-degree Price Discrimination:

Refers to a price discrimination in which a monopolist charges the maximum price that each buyer is willing to pay. This is also known as perfect price discrimination as it involves maximum exploitation of consumers. In this, consumers fail to enjoy any consumer surplus. First degree is practiced by lawyers and doctors.

ii. Second-degree Price Discrimination:

Refers to a price discrimination in which buyers are divided into different groups and different prices are charged from these groups depending upon what they are willing to pay. Railways and airlines practice this type of price discrimination.

iii. Third-degree Price Discrimination:

Refers to a price discrimination in which the monopolist divides the entire market into submarkets and different prices are charged in each submarket. Therefore, third-degree price discrimination is also termed as market segmentation.

In this type of price discrimination, the monopolist is required to segment market in a manner, so that products sold in one market cannot be resold in another market. Moreover, he/she should identify the price elasticity of demand of different submarkets. The groups are divided according to age, sex, and location. For instance, railways charge lower fares from senior citizens. Students get discount in cinemas, museums, and historical monuments.

Necessary Conditions for Price Discrimination:

Price discrimination implies charging different prices for identical goods.

It is possible under the following conditions:

i. Existence of Monopoly:

Implies that a supplier can discriminate prices only when there is monopoly. The degree of the price discrimination depends upon the degree of monopoly in the market.

ii. Separate Market:

Implies that there must be two or more markets that can be easily separated for discriminating prices. The buyer of one market cannot move to another market and goods sold in one market cannot be resold in another market.

iii. No Contact between Buyers:

Refers to one of the most important conditions for price discrimination. A supplier can discriminate prices if there is no contact between buyers of different markets. If buyers in one market come to know that prices charged in another market are lower, they will prefer to buy it in other market and sell in own market. The monopolists should be able to separate markets and avoid reselling in these markets.

iv. Different Elasticity of Demand:

Implies that the elasticity of demand in the markets should differ from each other. In markets with high elasticity of demand, low price will be charged, whereas in markets with low elasticity of demand, high prices will be charged. Price discrimination fails in case of markets having same elasticity- of demand.

Advantages and Disadvantages of Price Discrimination:

A monopolist practices price discrimination to gain profits. However, it acts as a loss for the consumers.

Following are some of the advantages of price discrimination:

i. Helps organizations to earn revenue and stabilize the business

ii. Facilitates the expansion plans of organizations as more revenue is generated

iii. Benefits customers, such as senior citizens and students, by providing them discounts In spite of advantages, there are certain disadvantages of price discrimination.

Some of the disadvantages of price discrimination as follows:

i. Leads to losses as some consumers end up paying higher prices

ii. Involves administration costs for separating markets.

Pricing under Monopolistic Competition

Meaning and Definition of Monopolistic Competition:

Before 1933, the traditional Marshallian theory of value was prevalent.

But in 1933 a revolution in the approach to price theory was initiated by the publication of two works of modern economists, Chamberlin and Mrs. Joan Robinson.

E.H. Chamberlin's work was entitled "The Theory of Monopolistic Competition" and Mrs. Robinson's "The Economics of Imperfect Competition".

Both economists challenged the concept of perfect competition and monopoly as unrealistic and attempted to present a new theory which is more realistic of the two new approaches, the view of Chamberlin's theory of monopolistic competition received wide acclamation. Critics also regarded Chamberlin's contribution as novel and superior to that of Mrs. Robinson's. In fact the real credit goes to Chamberlin for setting a new and realistic trend in the economics value.

Concept of Monopolistic Competition:

Monopolistic Competition refers to the market situation in which there is a keen competition, but neither perfect nor pure, among a group of a large number of small producers or suppliers having some degree of monopoly because of the differentiation of their products. Thus, we can say that monopolistic competition (or imperfect competition) is a mixture of competition and a certain degree of monopoly, on the basis of a correct appraisal of the market situation.

Chamberlin has asserted that monopoly and competition are not mutually exclusive rather both are frequently blend together. In short, we can say that a market with a blending of monopoly and competition is called monopolistic competition or imperfect competition.

Definition:

1. Monopolistic Competition refers to competition among a large number of sellers producing close but not perfect substitutes for each other.

2. According to Prof. Lerner – “The condition of imperfect competition arises when a seller has to face the falling demand curve.”

3. According to Prof. J. K. Mehta – “It has been more fully realised that every case of exchange is a case of what may be called partial monopoly and partial monopoly is looked at from the other said a case of imperfect competition. There is a blending of both competition element and monopoly element in each situation.”

4. According to Prof. Leftwich – “Monopolistic Competition (or imperfect competition) is that condition of industrial market in which a particular commodity of one seller creates an idea of difference from that of the other sellers in the minds of the consumers.”

Characteristics or Main Features of Monopolistic Competition:

Important characteristics of monopolistic competition are as follows:

1. Less Number of Buyers and Sellers:

In this market neither buyers nor sellers are too many as under perfect competition nor there is only one seller as under monopoly. Mostly, it is a situation in between. Every producer for his produced commodity has some special buyers. Every consumer and seller can influence demand and supply in the market.

2. Difference in the Quality and Shape of the Goods:

Although the commodities produced by different producers can serve as perfect substitutes to those produced by others, yet they are different in colour, form, packing, design, name etc. So there is product differentiation in the market.

3. Lack of Knowledge on the Part of Consumers:

Neither consumers nor sellers have full knowledge of market conditions, so there is international difference in the price of goods from those of others.

4. High Transportation Cost:

In this high transportation cost play an important role in order to create discrimination among commodities. Similar goods because of different transport costs are bought and sold at different prices.

5. Advertisement:

Here, advertisement plays an important role because buyers are influenced to prefer by advertisement, which plays upon their mind and makes them the product of one firm to those of another. Through advertisement, they are brought to his notice through radio, television and other audio-visual aids in a more pleasing and more forceful manner. Thus, rival firms compete against each other in quantity, in facilities as well as in price.

6. Ignorance of the Buyers:

There are some people who think that high priced goods will be better and of higher quality. So, they avoid buying low priced goods.

7. Differences in the Establishment of Industry:

In the imperfect competitive market, there is neither freedom of entry or exit as is under perfect competition nor there is perfect control as in monopoly but there are some restrictions on the entry of industry only.

Price-Output Equilibrium under Monopolistic Competition: Equilibrium in Short-Run and Long Run!

Under monopolistic competition, organizations need to make optimum adjustments in the prices and output sold to attain equilibrium.

Apart from this, under monopolistic competition, organizations also need to pay attention toward the design of the product and the way the product is promoted in the market.

Moreover, an organization under monopolistic competition is not only required to study its individual equilibrium, but group equilibrium of all organizations existing in the market. Let us first understand individual equilibrium of an organization under monopolistic competition.

As we know every seller, irrespective of the market structure, is willing to maximize his/her profits. In monopolistic competition, profits are maximized at a point where marginal revenue is equal to marginal cost. The price determined at this point is known as equilibrium price and the output produced at this point is called equilibrium output.

If the marginal revenue of a seller is greater than marginal cost, he/she may plan to expand his/her output. On the other hand, if marginal revenue is lesser than marginal cost, it would be profitable for the seller to reduce his/her output to the level where marginal revenue is equal to marginal cost.

Equilibrium in Short Run:

The short-run equilibrium of a monopolistic competitive organization is the same as that of an organization under monopoly. In the short run, an organization under monopolistic competition attains its equilibrium where marginal revenue equals marginal cost and sets its price according to its demand curve. This implies that in the short run, profits are maximized when MR=MC.

Figure-2 shows the equilibrium in the short run:

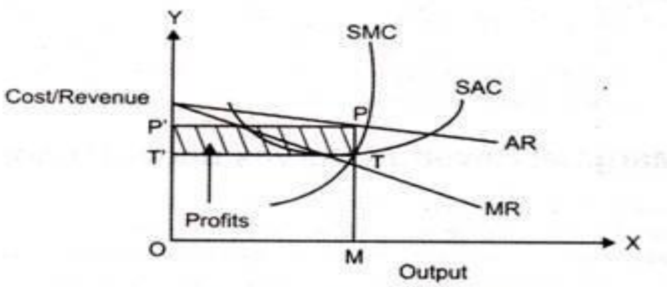


Figure-2: Equilibrium in the Short Run

In Figure-2, AR is the average revenue curve, MR represents the marginal revenue curve, SAC curve denotes the short run average cost curve, while SMC signifies the short run marginal cost. In Figure-2, it can be seen that MR intersects SMC at output OM where price is OP' (which is equal to MP). This is because P is the, point on AR curve, which is price.

From Figure-2, it can be interpreted from that the organization is earning supernormal profit. Supernormal profit per unit of output is the difference between the average revenue and average cost. In Figure-2, average revenue at equilibrium point is MP and average cost is MT.

Therefore, PT is the supernormal profit per unit of output. In the present case, supernormal profit would be measured by the area of rectangle P'PTT' (which is output multiplied by supernormal profit per unit of output).

On the other hand, when marginal cost is greater than marginal revenue, organizations would incur losses, as shown in Figure-3:

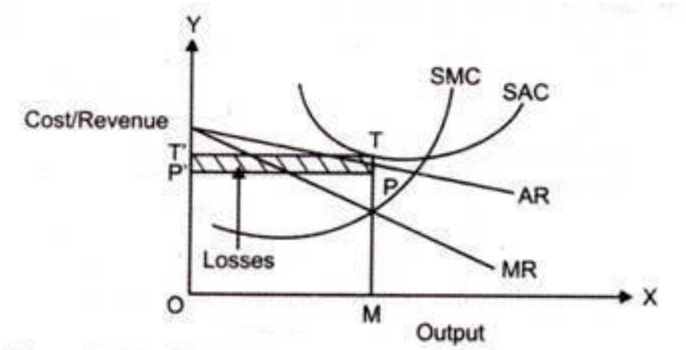


Figure-3: Equilibrium in the Short Run in Case of Losses

Figure-3 shows the condition of losses in the short run under monopolistic competition. Here, OP’ is smaller than MT, which implies that average revenue is smaller than average cost. TP is representing the loss that has incurred per unit of output. Therefore total loss is depicted from rectangle T’TPP’.

Equilibrium in Long Run:

In the preceding sections, we have discussed that in the short run, organizations can earn supernormal profits. However, in the long run, there is a gradual decrease in the profits of organizations. This is because in the long run, several new organizations enter the market due to freedom of entry and exit under monopolistic competition.

When these new organizations start production the supply would increase and the prices would fall. This would automatically increase the level of competition in the market. Consequently, AR curve shifts from right to left and supernormal profits are replaced with normal profits.

In the long run, the AR curve is more elastic than that of in the short run. This is because of an increase in the number of substitute products in the long- run. The long-run equilibrium of monopolistically competitive organizations is achieved when average revenue is equal to average cost. In such a case, organizations receive normal profits.

Figure-4 shows the long-run equilibrium position under monopolistic competition:

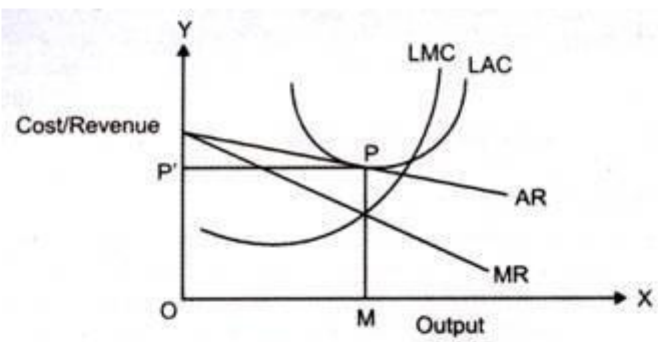


Figure-4: Equilibrium in the Long Run

In Figure-4, P is the point at which AR curve touches the average cost curve (LAC) as a tangent. P is regarded as the equilibrium point at which the price level is MP (which is also equal to OF) and output is OM.

In the present case average cost is equal to average revenue that is MP. Therefore, in long run, the profit is normal. In the short run, equilibrium is attained when marginal revenue is equal to marginal cost. However, in the long run, both the conditions (MR=MC and AR=AC) must hold to attain equilibrium.

Pricing under Pligopoly

Oligopoly is a market situation in which there are a few firms selling homogeneous or differenti- ated products. It is difficult to pinpoint the number of firms in the oligopolist market. There may be three, four or five firms.

It is also known as competition among the few. With only a few firms in the market, the action of one firm is likely to affect the others. An oligopoly industry produces either a homogeneous product or heterogeneous products.

The former is called pure or perfect oligopoly and the latter is called imperfect or differentiated oligopoly. Pure oligopoly is found primarily among producers of such industrial products as aluminium, cement, copper, steel, zinc, etc. Imperfect oligopoly is found among producers of such consumer goods as automobiles, cigarettes,' soaps and detergents, TVs, rubber tyres, refrigerators, typewriters, etc.

Characteristics of Oligopoly

Now that the Oligopoly definition is clear, it's time to look at the characteristics of Oligopoly:

Few firms

Under Oligopoly, there are a few large firms although the exact number of firms is undefined. Also, there is severe competition since each firm produces a significant portion of the total output.

Barriers to Entry

Under Oligopoly, a firm can earn super-normal profits in the long run as there are barriers to entry like patents, licenses, control over crucial raw materials, etc. These barriers prevent the entry of new firms into the industry.

Non-Price Competition

Firms try to avoid price competition due to the fear of price wars and hence depend on non-price methods like advertising, after sales services, warranties, etc. This ensures that firms can influence demand and build brand recognition.

Interdependence

Under Oligopoly, since a few firms hold a significant share in the total output of the industry, each firm is affected by the price and output decisions of rival firms. Therefore, there is a lot of interdependence among firms in an oligopoly. Hence, a firm takes into account the action and reaction of its competing firms while determining its price and output levels.

Nature of the Product

Under oligopoly, the products of the firms are either homogeneous or differentiated.

Selling Costs

Since firms try to avoid price competition and there is a huge interdependence among firms, selling costs are highly important for competing against rival firms for a larger market share.

No unique pattern of pricing behavior

Under Oligopoly, firms want to act independently and earn maximum profits on one hand and cooperate with rivals to remove uncertainty on the other hand.

Depending on their motives, situations in real-life can vary making predicting the pattern of pricing behavior among firms impossible. The firms can compete or collude with other firms which can lead to different pricing situations.

Indeterminateness of the Demand Curve

Unlike other market structures, under Oligopoly, it is not possible to determine the demand curve of a firm. This is because on one hand, there is a huge interdependence among rivals. And on the other hand there is uncertainty regarding the reaction of the rivals. The rivals can react in different ways when a firm changes its price and that makes the demand curve indeterminate.

2. Price Determination under Oligopoly:

We shall confine our study to the non-collusive oligopoly model of Sweezy, and to the collusive oligopoly models relating to cartels and price leadership.

1. The Sweezy Model of Kinked Demand Curve (Rigid Prices):

In his article published in 1939, Prof. Sweezy presented the kinked demand curve analysis to explain price rigidities often observed in oligopolistic markets. Sweezy assumes that if the oligopolistic firm lowers its price, its rivals will react by matching that price cut m order to avoid losing their customers. Thus the firm lowering the price will not be able to increase its demand much. This portion of its demand curve is relatively inelastic.

On the other hand, if the oligopolistic firm increases its price, its rivals will not follow it and change their prices. Thus the quantity demanded of this firm will fall considerably. This portion of the demand curve is relatively elastic. In these two situations, the demand curve of the oligopolistic firm has a kink at the prevailing market price which explains price rigidity.

Assumptions:

The kinked demand curve hypothesis of price rigidity is based on the following assumptions:

(1) There are few firms in the oligopolistic industry.

(2) The product produced by one firm is a close substitute for the other firms.

(3) The product is of the same quality. There is no product differentiation.

(4) There are no advertising expenditures.

(5) There is an established or prevailing market price for the product at which all the sellers are satisfied.

(6) Each seller's attitude depends on the attitude of his rivals.

(7) Any attempt on the part of a seller to push up his sales by reducing the price of his product will be counteracted by the other sellers who will follow his move.

(8) If he raises the price, others will not follow him. Rather they will stick to the prevailing price and cater to the customers, leaving the price-raising seller.

(9) The marginal cost curve passes through the dotted portion of the marginal revenue curve so that changes in marginal cost do not affect output and price.

The Model:

Given these assumptions, the price-output relationship in the oligopolist market is explained in Figure 1 where KPD is the kinked demand curve and OP0 the prevailing price in the oligopoly market for the OR product of one seller. Starting from point P, corresponding to the current price OP1, any increase in price above it will considerably reduce his sales, for his rivals are not expected to follow his price increase. This is so because the KP portion of the kinked demand curve is elastic, and the corresponding portion KA of the MR curve is positive. Therefore, any price-increase will not only reduce his total sale but also his total revenue and profit.

On the other hand, if the seller reduces the price of the product below OPQ (or P), his rivals will also reduce their prices. Though he will increase his sales, his profit would be less than before. The reason is that the PD portion of the kinked demand curve below P is less elastic and the corresponding part of marginal revenue curve below R is negative. Thus in both the price-raising and price-reducing situations, the seller will be a loser. He would stick to the prevailing market price OP0 which remains rigid.

In order to study the working of the kinked demand curve, let us analyse the effect of changes in cost and demand conditions on price stability in the oligopolistic market.

Changes in Costs:

In oligopoly under the kinked demand curve analysis changes in costs within a certain range do not affect the prevailing price. Suppose the cost of production falls so that the new MC curve is MC1, to the right, as in Figure 2. It cuts the MR curve in the gap AB so that the profit maximising output is OR which can be sold at OP0 price.

It should be noted that with any cost reduction the new MC curve will always cut the MR curve in the gap

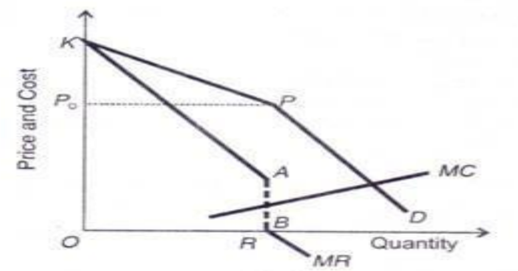


Fig. 1

because as costs fall the gap AB continues to widen due to two reasons:

(1) As costs fall, the upper portion KP of the demand curve becomes more elastic because of the greater certainty that a price rise by one seller will not be followed by rivals and his sales would be considerably reduced.

(2) With the reduction in costs the lower portion PD of the kinked curve becomes more inelastic, because of the greater certainty that a price reduction by one seller will be followed by the other rivals.

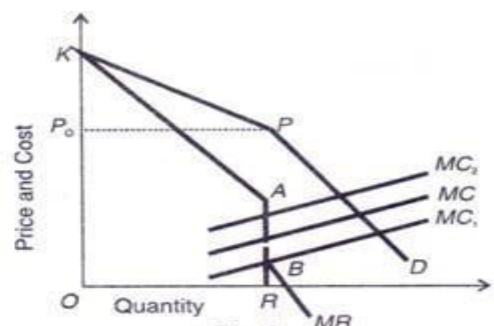


Fig. 2

Thus the angle KPD tends to be a right angle at P and the gap AB widens so that any MC curve below point A will cut the marginal revenue curve inside the gap. The net result is the same output OR at the same price OP0 and larger profits for the oligopolistic sellers.

In case the cost of production rises the marginal cost curve will shift to the left of the old curve MC as MC2. So long as the higher MC curve intersects the MR curve within the gap upto point A, the price situation will be rigid. However, with the rise in costs the price is not likely to remain stable indefinitely and if the MC curve rises above point A, it will intersect the MC curve in the portion KA so that a lesser quantity is sold at a higher price. We may conclude that there may be price stability under oligopoly even when costs change so long as the MC curve cuts the MR curve in its discontinuous portion. However, chances of the existence of price rigidity are greater where there is a reduction in costs than there is a rise in costs.

Changes in Demand:

We now explain price rigidity where there is a change in demand with the help of Figure 3, D2 is the original demand curve, MR2 is its corresponding marginal revenue curve and MC is the marginal cost curve. Suppose there is decrease in demand shown by D1 curve and MR1 is its marginal revenue curve. When demand decreases, a price-reduction move by one seller will be followed by other rivals.

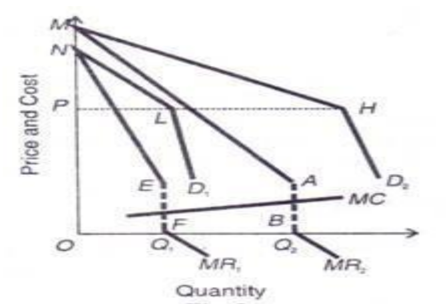


Fig. 3

This will make LD1, the lower portion of the new demand curve, more inelastic than the lower portion HD2 of the old demand curve. This will tend to make the angle at L approach a right angle. As a result, the gap EF in MR1 curve is likely to be wider than the gap AB of the MR2 curve.

The marginal cost curve MC will, therefore, intersect the lower marginal revenue curve MRX inside the gap EF, thus indicating a stable price for the oligopolistic industry. Since the level of the kinks H and L of the two demand curves remains the same, the same price OP is maintained after the decrease in demand. But the output level falls from OQ2 to OQ1.

This case can be reversed to show increase in demand by taking D2 and MR2 as the original demand and marginal revenue curves and D 2and MR2 as the higher demand and marginal revenue curves respectively. The price OP is maintained but the output rises from OQ1 to OQ. So long as the MC curve continues to intersect the MR curve in the discontinuous portion, there will be price rigidity.

The whole analysis of the kinked demand curve points out that price rigidity in oligopolistic markets is likely to prevail if there is a price reduction move on the part of all sellers. Changes in costs and demand also lead to price stability under normal conditions so long as the MC curve intersects the MR curve in its discontinuous portion. But price increase rather than price rigidity may be found in response to rising cost or increased demand.

Reasons for Price Stability:

There are a number of reasons for price rigidity in certain oligopoly markets:

(1) Individual sellers in an oligopolistic industry might have learnt through experience the futility of

price wars and thus prefer price stability.

(2) They may be content with the current prices, outputs and profits and avoid any involvement in unnecessary insecurity and uncertainty.

(3) They may also prefer to stick to the present price level to prevent new firms from entering the industry.

(4) The sellers may intensify their sales promotion efforts at the current price instead of reducing it. They may view non-price competition better than price rivalry.

(5) After spending a lot of money on advertising his product, a seller may not like to raise its price to deprive himself of the fruits of his hard labour. Naturally, he would stick to the going price of the product.

(6) If a stable price has been set through agreement or collusion, no seller would like to disturb it, for fear of unleashing a price war and thus engulfing himself into an era of uncertainty and insecurity.

(7) It is the kinked demand curve analysis which is responsible for price rigidity in oligopolistic markets.

It's Shortcomings:

But the theory of kinked demand curve in oligopoly pricing is not without shortcomings.

(1) Even if we accept all its assumptions it is not likely that the gap in the marginal revenue curve will be wide enough for the marginal cost curve to pass through it. It may be shortened even under conditions of fall in demand or costs thereby making price unstable.

(2) One of its major shortcomings, according to Prof. Stigler, is that "the theory does not explain why prices that have once changed should settle down, again acquire stability, and gradually produce a new kink." For instance in Figure 2, the kink occurs at P because OP0 is the prevailing price. But the theory does not explain the forces that established the initial price OP0.

(3) Price stability may be illusory because it is not based on the actual market behaviour. Sales do not always occur at list prices. There are often deviations from posted prices because of trade- ins, allowance and secret price concessions. The oligopolistic seller may outwardly keep the price stable but he may reduce the quality or quantity of the product. Thus price stability becomes illusory.

(4) Moreover, it is not possible to statistically compile actual sales prices in the case of many products that may reflect stable prices for them. It is, therefore, doubtful that price stability actually exists in oligopoly.

(5) Critics point out that the kinked demand curve analysis holds during the short run, when the knowledge about the reactions of rivals is low. But it is difficult to guess correctly the rivals' reactions in the long run. Thus the theory is not applicable in the long-run.

(6) According to some economists, the kinked demand curve analysis applies to an oligopolistic industry in its initial stages or to that industry in which new and previously unknown rivals enter the market.

(7) The kinked demand curve analysis is based on two assumptions: first, other firms will follow a price cut and, second, they will not follow a price rise. Stigler has shown on empirical evidence that in an inflationary period the rise in output prices is not confined only to one firm but is industry-wide. So all firms having similar costs will follow one another in raising price. In Stigler's words: "There is little historical basis for a firm to believe that price increases will not be matched by rivals and that price decreases will be matched."

(8) Economists have concluded from this that the kinked demand curve analysis is applicable only under depression. For in an inflationary period when demand increases, the oligopolistic firm will raise price and other firms will also follow it. In such a situation, the demand curve of the oligopolist will have inverted kink. This reverse kink is based on his expectation that all his competitors will follow him when he raises the price of his product, but none will follow a price cut because of inflationary condition.

(9) Stigler further points out those cases in oligopoly industries where the number of sellers is either very small or somewhat large, the kinked demand curve is not likely to be there.

"However", as pointed out by Professor Baumol, "the analysis does show how the oligopolistic firm's view of competitive reaction patterns can affect the changeability of whatever price it happens to be charging."

2. Collusive Oligopoly:

Collusive oligopoly is a situation in which firms in a particular industry decide to join together as a single unit for the purpose of maximising their joint profits and to negotiate among themselves so as to share the market. The former is known as the joint profit maximisation cartel and the latter as the market-sharing cartel.

There is another type of collusion, known as leadership, which is based on tacit agreements. Under it, one firm acts as the price leader and fixes the price for the product while other firms follow it. Price leadership is of three types: low-cost firm, dominant firm, and barometric.

(A) Cartels:

A cartel is an association of independent firms within the same industry. The cartel follows common policies relating to prices, outputs, sales and profit maximisation and distribution of products. Cartels may be voluntary or compulsory and open or secret depending upon the policy of the government with regard to their formation. Thus cartels have many forms and use many devices in order to follow varied common policies depending upon the type of the cartel. We discuss below the two most common types of cartels: (1) Joint profit maximisation or perfect cartel; and (2) market-sharing cartel.

1. Joint Profit Maximisation Cartel:

The uncertainty to be found in an oligopolistic market provides an incentive to rival firms to form a perfect cartel. Perfect cartel is an extreme form of perfect collusion. In this, firms producing a homogeneous product form a centralised cartel board in the industry. The individual firms surrender their price-output decisions to this central board.

The board determines output quotas for its members, the price to be charged and the distribution of industry profits. Since the central board manipulates prices, outputs, sales and distribution of profits, it acts like a single monopoly whose main aim is to maximise the joint profits of the oligopolistic industry.

Assumptions:

The analysis of joint profit maximisation cartel is based on the following assumptions:

1. Only two firms A and B are assumed in the oligopolistic industry that forms the cartel.
2. Each firm produces and sells a homogeneous product that is a perfect substitute for each other.
3. The number of buyers is large.
4. The market demand curve for the product is given and is known to the cartel.
5. The cost curves of the firms are different but are known to the cartel.
6. The price of the product determines the policy of the cartel.
7. The cartel aims at joint profit maximisation.

Joint Profit Maximisation Solution:

Given these assumptions, and given the market demand curve and its corresponding MR curve, joint profits will be maximised when the industry MR equals the industry MC. Figure 4 illustrates this situation where D is the market (or cartel) demand curve and MR is its corresponding marginal revenue curve. The aggregate marginal cost curve of the industry ΣMC is drawn by the lateral summation of the MC curves of firms A and B, so that ΣMC = MCa + MCb. The cartel solution that maximises joint profit is determined at point E where the ШС curve intersects the industry MR curve.

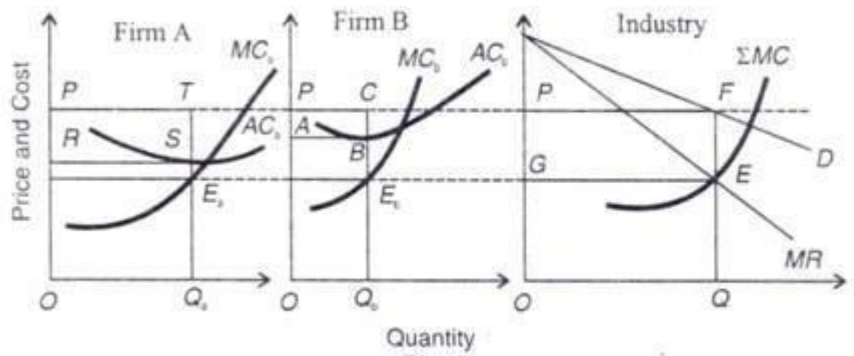


Fig. 4

Consequently, the total output is OQ which will be sold at OP = (QF) price. As under monopoly, the cartel board will allocate the industry output by equating the industry MR to the marginal cost of each firm. The share of each firm in the industry output is obtained by drawing a straight line from E to the vertical axis which passes through the curves MCb and MCa of firms B and A at points Eb and Ea respectively.

Thus the share of firm A is OQa and that of firm B is OOb which equal the total output OQ (= OQa + OQb). The price OP and the output OQ distributed between A and B (inns in the ratio of OQa: OQb is the monopoly solution. Firm A with the lower costs sells larger output OQa than firm B with higher costs so that OQa > OQb. But this does not mean that A will be getting more profit than B.

The joint maximum profit is the sum of RSTP and ABCP earned by A and B respectively. It will be pooled into a fund and distributed by the cartel board according to the agreement arrived at by the two firms at the time of the formation of the cartel. A pooling agreement of this type will make it possible for both firms to maximise their joint profit provided the total profits earned by them independently do not exceed the former.

Advantages:

Thus perfect collusion by oligopolistic firms in the form of a cartel has certain advantages. It avoids price wars among rivals. The firms forming a cartel gain at the expense of customers who are charged a high price for the product. The cartel operates like a monopoly organisation which maximises the joint profit of firms. Joint profits are generally more than the total profits earned by them if they were to act independently.

Difficulties of a Cartel:

The above analysis is based on perfect collusion in which all firms relinquish their individual price-output decisions to a central board of the cartel which acts like a multi-plant monopolist. But this is only a theoretical possibility in the short run because in practice the joint profit maximisation objective cannot be achieved by a cartel. In the long run, there are a number of difficulties faced by a cartel which tend to break it down.

They are as under:

1. It is difficult to make an accurate estimate of the market demand curve. Each firm thinks that its own demand curve is more elastic than the market demand curve because its product is a perfect substitute for the product of its rivals. Thus if the market demand curve is underestimated so will be its corresponding MR curve which will make the estimation of the market price inaccurate by the cartel.

2. Similarly, the estimation of the market MC curve may be inaccurate because of the supply of wrong data about their MCs by individual firms to the cartel. There is every possibility that the individual firms may supply low-cost data to the central cartel board in order to have larger share of output and profits. This may ultimately lead to the breakdown of the cartel.

3. The formation of a cartel is a slow process which takes a long time for the agreement to arrive at by firms especially if their number is very large. In the meantime, there may be changes in the cost structure and market demand for the product. This renders the cartel agreement useless and it breaks down soon.

4. If a firm's product is preferred more by consumers than that of the other members of the cartel, the market demand for it may be higher than the quota fixed by the cartel. It may, therefore, secretly sell more than its quota and if followed by other firms, the cartel will break down.

5. The larger the number of firms in a cartel, the less is its chances of survival for long because of the distrust, threatening and bargaining resorted to by them. The cartel will, therefore, break down.

6. In theory, the cartel-members agree on joint profit maximisation. But in practice, they seldom agree on profit distribution. Large firms want a lower price, a higher output quota and larger profits. So when such problems arise in joint profit distribution in contravention of the cartel agreement, they lead to the breakdown of the cartel.

7. I he price of the product fixed by the cartel cannot be changed even if the market conditions require it to be changed. This is because it takes long time for the members to arrive at an agreed price. This stickiness of the price often leads to the breakdown of the cartel when some members defect from it.

8. Price stickiness gives rise to "chislers" who secretly cut the price or violate the quota agreement. Such secret dealings by firms to raise their own profits tend to break down the cartel.

9. Unless all member firms in the cartel are strongly committed to cooperation, outside disturbances, such as a sharp fall in demand, may lead to the breakdown of the cartel.

10. When a cartel raises the price of the product and increases the profits of its members, it creates an incentive for new firms to enter the industry. Even if the entry of new firms is blocked, it is only a short-run phenomenon because the success of the cartel will lead to the entry of firms in the long run. This will force the cartel to break down. If the new firms are allowed to enter the cartel, it will become unmanageable, increase the

defectors and bring its end.

11. Some high-cost uneconomic firms may refuse to shut down or leave the cartel despite the cartel board's request. This is likely to distort the profit maximisation level of the cartel and thus break it.

12. The cartel's policy of fixing high price and restricting the quantity of the product may lead to the emergence of substitutes in the long run. The other firms may invent and produce cheaper substitutes which may be accepted by consumers. This will tend to reduce the demand for the cartel's product, make it more elastic, reduce its joint profits and thus break the cartel.

13. The cartel may not be able to maximise joint profits by not charging a very high price for fear of government interference and regulation.

14. Similarly, the cartel may not charge a very high price and maximise its joint profits in order to have a good public image or reputation.

Thus the chances are greater for individual firms to leave the cartel on account of personal bickering and antagonism of member firms over allotment of quotas and division of profits which are likely to affect adversely joint profit maximisation and end the cartel agreement.

Besides these problems in the working of a cartel, it is more difficult to form and run a cartel for long in the case of a differentiated product than in the case of a homogeneous product. For, it is not possible to rationalise and sort out the differences in the qualities of the product.

2. Market-Sharing Cartel:

Another type of perfect collusion in an oligopolistic market is found in practice which relates to market-sharing by the member firms of a cartel. The firms enter into a market-sharing agreement to form a cartel "but keep a considerable degree of freedom concerning the style of their output, their selling activities and other decisions."

There are two main methods of market-sharing:

(a) non-price competition; and

(b) Quota system.

They are discussed as under:

(a) Non-Price Competition Cartel:

The non-price competition agreement among oligopolistic firms is a loose form of cartel. Under this type of cartel, the low-cost firms press for a low price and the high-cost firms for a high price. But ultimately, they agree upon a common price below which they will not sell.

Such a price must allow them some profits. The firms can compete with one another on a non- price basis by varying the colour, design, shape, packing, etc. of their product and having their own different advertising and other selling activities. Thus each firm shares the market on a non- price basis while selling the product at the agreed common price.

This type of cartel is inherently unstable because if one low-cost firm cheats the other firms by charging a lower price than the common price, it will attract the customers of other member firms and earn larger profits. When other firms come to know of this, they will leave the cartel. A price war will start and ultimately the lowest-cost firm will remain in the industry.

In case the cost curves of the firms forming a cartel differ, the low-cost firms may not stick to the common price. They may try to increase their share of the market by means of secret price concessions. They may also resort to better sales promotion methods. Such policies tend to change their demand-cost conditions further. Consequently, price variations among firms become more common. Ultimately, the cartel agreement becomes a farce and a price war starts. This leads to the breaking up of the cartel agreement.

(b) Market Sharing by Quota Agreement:

The second method of market sharing is the quota agreement among firms. All firms in an oligopolistic industry enter into a collusion for charging an agreed uniform price. But the main agreement relates to the sharing of the market equally among member firms so that each firm gets profits on its sales.

Assumptions:

This analysis is based on the following assumptions:

1. There are only two firms that enter into market-sharing agreement on the basis of the quota system.
2. Each firm produces and sells a homogeneous product which is a perfect substitute for each other.
3. The number of buyers is large.
4. The market demand curve for the product is given and known to the cartel.
5. Each firm has its own demand curve having the same elasticity as that of the market demand curve.
6. The cost curves of the two firms are identical.
7. Both firms share the market equally.
8. Each sells the product at the agreed uniform price.
9. There is no threat of entry' by new firms.

Market-Sharing Solution:

Given these assumptions, the equal market sharing between the two firms is explained in terms of Figure 5 where D is the market demand curve and d/MR is its corresponding MR curve. EMC is the aggregate MC curve of the industry. The ZMC curve intersects the d/MR curve at point E which determines QA (= OP) price and total output OQ for the industry. This is the monopoly solution in the market-sharing cartel.

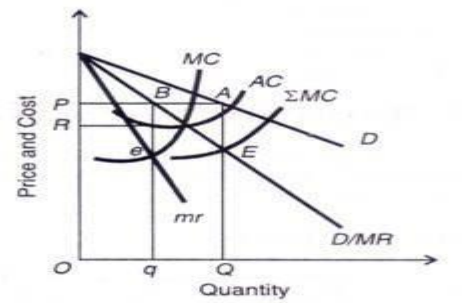


Fig. 5

How will the industry output be shared equally between the two firms? Now assume that the d/ MR is the demand curve of each firm and mr is its corresponding MR curve. AC and MC are their identical cost curves. The MC curve intersects the mr curve at point e so that the profit maximisation output of each firm is Oq.

Since the total output of the industry is OQ which is equal to 2 x Oq = (OQ = 2Oq), it is equally shared by the two firms as per the quota agreement between them. Thus each firm sells Oq output at the same price qB (=OP) and earns RP per unit profit. The total profit earned by each firm is RP x Oq and by both is RP x 2Oq or RP x OQ.

However, in actuality, there are more than two firms in an oligopolistic industry which do not share the market equally. Moreover, their cost curves are also not identical. In case their cost curves differ, their market shares will also differ. Each firm will charge an independent price in accordance with its own MC and MR curves.

They may not sell the same quantity at the agreed common price. They may be charging a price slightly above or below the profit maximisation price depending upon its cost conditions. But each will try to be nearest the profit maximisation price. This will ultimately lead to the breaking up of the market sharing agreement.

With Threat of Entry:

So far our analysis has been confined to collusive oligopoly without any threat of entry of new firms in the industry. Suppose there is a constant threat of entry into the oligopolistic industry. In that case if the firms agree on the price OP, new firms will enter the industry, reduce their sales and profits. This may ultimately lead to excess capacity and uneconomic firms in the industry. The existence of excess capacity and uneconomic firms will raise the average costs AC to the level of B (not shown in Figure 5) and the firms will be earning only normal profits. Each firm will sell less than Oq.

If the existing oligopolists are wiser, they may forestall entry by charging a price lower than the profit maximisation price OP. In this way the collusive oligopolists by charging a lower price in the present will be earning larger profits in the long-run, and continue their exclusive control over the market by keeping the new entrants out for ever.

We may conclude that perfect collusive oligopoly pricing has not any set pattern of price behav- iour. The resultant price and output will depend upon the reaction of the collusive oligopolists towards the profit maximisation price and their attitude towards the existing and potential rivals.

(B) Price Leadership:

Price leadership is imperfect collusion among the oligopolistic firms in an industry when all firms follow the lead of one big firm.

There is a tacit agreement among the firms to sell the product at a price set by the leader of the industry (i.e. the big firm). Sometimes, there is a formal meeting and a definite agreement with the leader firm. If the products are homogeneous, a uniform price is established. In case of a differentiated product also prices can be uniform. Whatever price changes take place, the leader announces from time to time, and the other firms follow him.

In America, examples of price leadership industries are:

Biscuits, cement, cigarettes, flour, fertilizers, petroleum, milk, rayon, steel, etc. They relate both to pure and differentiated oligopoly.

Price leadership is of various types. But there are three most common price leadership models which we discuss now:

1. The Low-Cost Price Leadership Model:

In the low-cost price leadership model, an oligopolistic firm having lower costs than the other firms sets a lower price which the other firms have to follow. Thus the low-cost firm becomes the price leader.

It’s Assumptions:

The low-cost firm model is based on the following assumptions:

1. There are two firms A and B.
2. Their costs differ. A is the low-cost firm and B is the high-cost firm.
3. They have identical demand and MR curves. The demand curve faced by them is 1/2 of the market demand curve.
4. The number of buyers is large.
5. The market industry demand curve for the product is known to both the firms.

The Model:

Given these assumptions, both firms enter into a tacit agreement whereby the high-cost firm B will follow the price set by the price leader firm A and to share the market equally. The price policy to be followed by both is illustrated in Figure 6. D is the industry demand curve and d/MR is its corresponding marginal revenue curve which is the demand curve for both the firms and mr is their marginal revenue curve. The cost curves of the low-cost firm A are AC and MC and of the high-cost firm B are ACa and MCb.

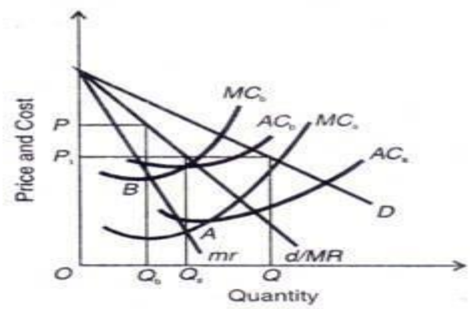


Fig. 6

If the two firms were to act independently, the high-cost firm B would charge OP price per unit and sell OQb quantity, as determined by point B where its MCb curve cuts the mr curve. Similarly, the low-cost firm A would charge OP1price per unit and sell OQa quantity, as determined by point A where its MCa curve cuts the mr curve.

As there is a tacit agreement between the two firms, the high-cost firm B has no choice but to follow the price leader firm A. It will, therefore, sell OQa quantity at a lower price OP1even though it will not be earning maximum profits. On the other hand, the price leader A will earn much higher profits at OP1price by selling OQ quantity. Since both A and B sell the same quantity OQa, the total market demand OQ is equally divided between the two, OQ = 20Qa. But if firm B sticks to OP price, its sales will be zero because the product being homogeneous, all its customers will shift to firm A.

The price-leader firm A can, however, drive firm B out of the market by setting a lower price than OFF, lower than the average cost ACb of firm B. Firm A would become a monopoly firm. But in such a situation it will have to face legal problems. Therefore, it will be in its interest to fix OP1 price and tolerate firm B in order to share the market equally and maximise its profits.

Price Leadership Model with Unequal Market Share:

In the case of the price leadership model with unequal market share, the two firms will have different demand curves along with their different cost curves. The low-cost firm's demand curve will be more elastic than that of the high-cost firm. The high-cost firm would maximise its profits by selling less at a higher price while the low-cost firm would sell more at a lower price and maximise its profits.

If they enter into a common price agreement, it would be in the interest of the high-cost firm to sell more quantity at a lower price set by the price leader by earning a little less than the maximum profits. But this is only possible so long as the price set by the leader covers the AC of the high-cost firm.

The price leadership model with unequal market shares is illustrated in Figure 7, where the market demand curve

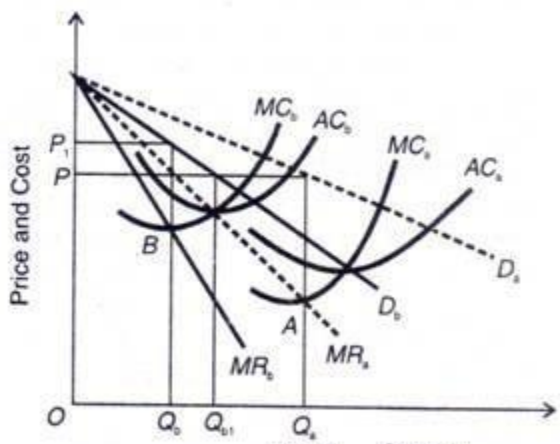


Fig. 7

is not shown to simplify the analysis. In the figure, Da is the demand curve of the low-cost firm A and MRa is its marginal revenue curve. The demand curve and MR curve of the high-cost firm B are Db and MRb. The low-cost firm A sets the price OP and the quantity OQa when its MCa curve cuts its MRa curve at point A.

The price OP1 and quantity OQb of the high-cost firm B are determined when its MCb curve cuts its MRb curve at point B. Following the price leader firm A, when firm B accepts the price OP, it sells more quantity OQb1 and earns less than maximum profits. It will pay the follower firm to sell this quantity at OP price so long as this price covers its average cost.

If it does not follow the leader firm and tries to sell OQb quantity at its profit maximisation price OP1 it will have to close down because its customers will switch over to the leader firm which charges low price OP. However, if there is no agreement for sharing the market between the leader and the follower firms, the follower can adopt the price of the leader (OP) but produce a lower quantity (less than OQb1) than required to maintain the price in the market, and thus push the leader to a non-profit maximisation position by producing less output.

2. The Dominant Firm Price Leadership Model:

This is a typical case of price leadership where there is one large dominant firm and a number of small firms in the industry. The dominant firm fixes the price for the entire industry and the small firms sell as much product as they like and the remaining market is filled by the dominant firm itself. It will, therefore, select that price which brings more profits to itself.

Assumptions:

This model is based on the following assumptions:

(1) The oligopolistic industry consists of a large dominant firm and a number of small firms.

(2) The dominant firm sets the market price.

(3) All other firms act like pure competitors, which act as price takers. Their demand curves are perfectly elastic for they sell the product at the dominant firm's price.

(4) The dominant firm alone is capable of estimating the market demand curve for the product.

(5) The dominant firm is in a position to predict the supplies of other firms at each price set by it.

The Model:

Given these assumptions, when each firm sells its product at the price set by the dominant firm, its demand curve is perfectly elastic at that price. Thus its marginal revenue curve coincides with the horizontal demand curve. The firm will produce that output at which its marginal cost equals marginal revenue. The MC curves of all the small firms combined laterally establish their aggregate supply curve. All these firms behave competitively while the dominant firm behaves passively. It fixes the price and allows the small firms to sell all they wish at that price.

The case of price leadership by the dominant firm is explained in terms of Figure 8 where DD1 is the market demand curve. ΣMCs is the aggregate supply curve of all the small firms. By subtracting ΣMC from DD1 at each price, we get the demand curve faced by the dominant firm, PNMBD1 which can be drawn as follows. Suppose

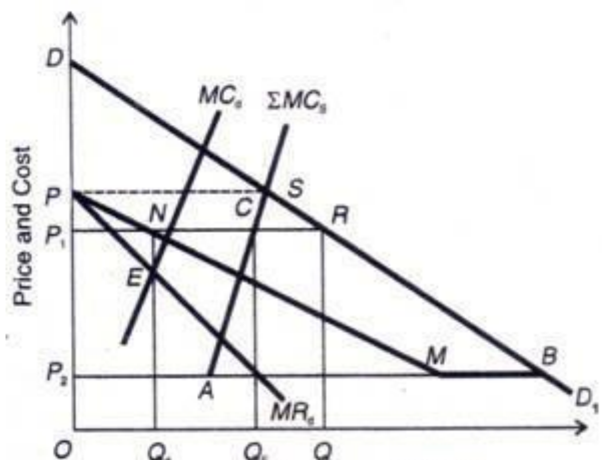


Fig. 8

the dominant firm sets the price OP.

At this price, it allows the small firms to meet the entire market demand by supplying PS quantity. But the dominant firm would supply nothing at the price OP. Point P is, therefore, the starting point of its demand curve. Now take a price OP1 less than OP. The small firms would supply P1C (=OQs) output at this price OP1 when their ΣMCs curve cuts their horizontal demand curve P1 R at point C. Since the total quantity demanded at

OP1price is P1R (=OQ) and the small firms supply P1C quantity, CR (=QSQ) quantity would be supplied by the dominant firm.

By taking P1N= CR on the horizontal line P1R, the dominant firm's supply becomes P1 N (=OQd). Thus we derive point N on the dominant firm's demand curve by subtracting the horizontal distance from point P1to N from the demand curve DDV Since the small firms supply nothing at prices below OP1 because their ΣMCs curve exceeds this price, the dominant firm's demand curve coincides with the horizontal line P1B over the range MB and then with the market demand curve over the segment BD1.Thus the dominant firm's demand curve is PNMBD1.

The dominant firm will maximise its profits at that output where its marginal cost curve MCd cuts its MRd, the marginal revenue curve. It establishes the equilibrium point E at which the dominant firm sells OQa output at OP1 price. The small firms will sell OQs output at this price for, the marginal cost curve of the small firms equals the horizontal price line P1 R at C.

The total output of the industry will be OQ = OQd + OQs. If OP2 price is set by the dominant firm, the small firms would sell P2A and the dominant firm AB. In case a price below OP2 is set the dominant firm would meet the entire industry demand- and the sales of the small firms would be zero. The above analysis shows that the price- quantity solution is stable because the small firms behave passively as price-takers.

However, the real test of dominant firm's price leadership is the extent to which the other firms follow its lead. The moment the firms cease to follow the price leader, the model breaks down. Besides, if the other firms have different cost curves, the same price may not maximise short-run profits for all the firms.

The dominant-firm model of price leadership can have a number of variations. There may be two or more large firms among a number of small firms which may enter into collusion for sharing the market at various prices. There may be product differentiation. Nevertheless, the conclusions arrived at help to explain price output policies in all such situations.

3. The Barometric Price Leadership Model:

The barometric price leadership is that in which there is no leader firm as such but one firm among the oligopolistic firms with the wisest management which announces a price change first which is followed by other firms in the industry. The barometric price leader may not be the dominant firm with the lowest cost or even the largest firm in the industry.

It is a firm which acts like a barometer in forecasting changes in cost and demand conditions in the industry and economic conditions in the economy as a whole. On the basis of a formal or informal tacit agreement, the other firms in the industry accept such a firm as the leader and follow it in making price changes for the product.

The barometric price leadership develops due to the following reasons:

1. As a reaction to the earlier experience of violent price changes and cut-throat competition among oligopolistic firms, they accept one firm as the price leader.
2. Most firms do not possess the expertise to calculate cost and demand conditions of the industry. So they leave their estimation to one leader firm which has the ability to do so.
3. Oligopolistic firms accept one among them as the barometric leader firm which possesses better knowledge and predictive power about changes in direct costs or style and quality changes and changes in the economic conditions as a whole.

It is not essential that a firm selected as the barometric leader must belong to the industry. Even a firm not belonging to the industry may be chosen as the barometric leader.

3. Non-Price Competition in Oligopoly:

There is not much of active price competition in oligopolistic markets. There are occasionally price wars among firms which are due to the failure of communication channels among firms. Usually, prices are stable in an oligopolistic market. Competition among firms is, therefore, for increased market share of the product. The oligopolistic firms know that if they try to increase their market share through price cut, competition among them will lead to an unabated fall in the price and all of them would be losers in the process. Thus, instead of competing through price, they resort to non-price competition.

Non-price Competition refers to the efforts on the part of one oligopolistic firm to increase its sales by some means other than a price reduction. Some other means are advertising, product differentiation and customer service. These, in turn, include publicity, sales promotion and personal selling; product quality, stylistic and aesthetic quality, brand name and packaging; and service agreement, warranty, guarantee, selling on credit, installment selling, etc.

Thus non-price competition involves efforts by an oligopolist to differentiate his product from that of his rivals by establishing real or imaginary differences in the minds of consumers through the quality of the product, its technological level, and through service, marketing and promotional means.

Economists tend to bundle these different dimensions of non-price competition under product differentiation. An oligopolistic firm tries to differentiate its product from that of its rivals in order to raise the demand for its product and to make its demand curve less elastic.

To achieve these objectives, a firm may seek to have successful product differentiation in a number of ways. It may spend more on advertising and promotion, rather than on the attributes of its product. Or, it may change the features and packaging of its product in such a way that it appeals more to buyers.

Product Attributes:

In the case of product variation the firm may choose a brand name or a brand mark for its product which may create an element of distinctiveness and make it easier to identify the product by buyers. The firm may choose those product attributes which buyers value highly, which it can provide cost-effectively and which its rivals are not in a position to provide. Through its technological efforts, the firm may seek both product and process development for enhancing the quality and features of its product. Similarly, it may direct its technological effort towards the specific requirements of its targeted buyers.

Advertising and Promotion:

The main purpose of advertising and promotion is to shift the demand curve for the product upward to the right. So the oligopolistic firm is able to sell more at each and every price. Advertising differentiates one product from another and makes the product better known than others.

Thus advertising pushes the sales of the product of a particular firm as against that of its rivals. Appealing posters, short films on TV showing a famous film star or a model uttering words in praise of a particular product brand and commercial broadcasts all aim at pushing the sales of one firm at the cost of others.

Economists measure advertising and promotion efforts on the part of a firm in terms of Adver- tising (Promotional) Elasticity of Demand which measures the responsiveness of sales to changes in advertising and promotional expenses.

Thus advertising elasticity:

$Ea = \Delta Q/ \Delta A .A/Q$

Where Q refers to sales or demand and A to advertising and promotional expenses.

Ea is positive because advertising expenses are supposed to increase sales. The higher the advertising elasticity, the greater the incentive for the firm to advertise its product. In fact, E is a measure of the effectiveness of advertising. As advertising expenses increase, their effectiveness increases.

But in the case of an oligopolistic firm, the larger the firm’s market shares in the industry, the lower the advertising elasticity of demand is likely to be. If the rival firms react to increases in the firm’s advertising expenses by increasing their own advertising expenses, then these expenses will tend to cancel each other, thereby reducing the advertising elasticity of demand.

Marketing Channels:

In the traditional oligopoly theory, no reference is found to the marketing channels which play an important role in the promotion of a product. This is based on the implicit assumption that there is direct marketing of the product to buyers.

COMPUTER APPLICATION (109)

UNIT I

Computer

A computer is an electronic device, operating under the control of instructions stored in its own memory that can accept data (input), process the data according to specified rules, produce information (output), and store the information for future use.

Functionalities of a computer

- Takes data as input
- Stores the data/instructions in its memory and use them when required
- Processes the data and converts it into useful information
- Generates the output
- Controls all the above four steps

Computer Components

Any kind of computers consists of **HARDWARE** AND **SOFTWARE**.

Hardware:

Computer hardware is the collection of physical elements that constitutes a computer system. Computer hardware refers to the physical parts or components of a computer such as the monitor, mouse, keyboard, computer data storage, hard drive disk (HDD), system unit (graphic cards, sound cards, memory, motherboard and chips), etc. all of which are physical objects that can be touched.

Input Devices:

Input device is any peripheral (piece of computer hardware equipment to provide data and control signals to an information processing system such as a computer or other information appliance.

Input device Translate data from **form** that humans understand to one that the computer can work with. Most common are keyboard and mouse.

Example of Input Devices:- 1. Keyboard	2. Mouse (pointing device)	3. Microphone
4. Touch screen	5. Scanner	6. Webcam
7. Touchpads	8. MIDI keyboard	
9. Graphics Tablets	10. Cameras	11. Pen Input
12. Video Capture Hardware	13. Microphone	14. Trackballs
15. Barcode reader	16. Digital camera	17. Joystick
18. Gamepad	19. Electronic Whiteboard	

Central Processing Unit (CPU)

A CPU is brain of a computer. It is responsible for all functions and processes. Regarding computing power, the CPU is the most important element of a computer system.

The CPU is comprised of three main parts:

*** Arithmetic Logic Unit (ALU)**: Executes all arithmetic and logical operations. Arithmetic calculations like as addition, subtraction, multiplication and division. Logical operation like compare numbers, letters, or special characters

*** Control Unit (CU):** controls and co-ordinates computer components.

1. Read the code for the next instruction to be executed.

2. Increment the program counter so it points to the next instruction.

3. Read whatever data the instruction requires from cells in memory.

4. Provide the necessary data to an ALU or register.

5. If the instruction requires an ALU or specialized hardware to complete, instruct the hardware to perform the requested operation.

* ***Registers*** : Stores the data that is to be executed next, "very fast storage area".

Primary Memory:-

1. **RAM**: Random Access Memory (RAM) is a memory scheme within the computer system responsible for storing data on a temporary basis, so that it can be promptly accessed by the processor as and when needed. It is volatile in nature, which means that data will be erased once supply to the storage device is turned off. RAM stores data randomly and the processor accesses these data randomly from the RAM storage. RAM is considered "random access" because you can access any memory cell directly if you know the row and column that intersect at that cell.

2. **ROM** (Read Only Memory): ROM is a permanent form of storage. ROM stays active regardless of whether power supply to it is turned on or off. ROM devices do not allow data stored on them to be modified.

Secondary Memory:-

Stores data and programs permanently: its retained after the power is turned off

1. **Hard drive (HD):** A hard disk is part of a unit, often called a "disk drive," "hard drive," or "hard disk drive," that store and provides relatively quick access to large amounts of data on an electromagnetically charged surface or set of surfaces.

2. **Optical Disk:** An optical disc drive (ODD) is a disk drive that uses laser light as part of the process of reading or writing data to or from optical discs. Some drives can only read from discs, but recent drives are commonly both readers and recorders, also called burners or writers. Compact discs, DVDs, and Blu-ray discs are common types of optical media which can be read and recorded by such drives. Optical drive is the generic name; drives are usually described as "CD" "DVD", or "Bluray", followed by "drive", "writer", etc. There are three main types of

optical media: CD, DVD, and Blu-ray disc. CDs can store up to 700 megabytes (MB) of data and DVDs can store up to 8.4 GB of data. Blu-ray discs, which are the newest type of optical media, can store up to 50 GB of data. This storage capacity is a clear advantage over the floppy disk storage media (a magnetic media), which only has a capacity of 1.44 MB.

3. Flash Disk: A storage module made of flash memory chips. A Flash disks have no mechanical platters or access arms, but the term "disk" is used because the data are accessed as if they were on a hard drive. The disk storage structure is emulated.

<u>Output devices</u>

An output device is any piece of computer hardware equipment used to communicate the results of data processing carried out by an information processing system (such as a computer) which converts the electronically generated information into human-readable form.

Example on Output Devices: 1. Monitor	2. LCD Projection Panels
3. Printers (all types)	4. Computer Output Microfilm (COM)
5. Plotters	6. Speaker(s)
7. Projector	

<u>Software</u>

Software is a generic term for organized collections of computer data and instructions, often broken into two major categories: system software that provides the basic non-task-specific functions of the computer, and application software which is used by users to accomplish specific tasks.

Software Types

A. System software is responsible for controlling, integrating, and managing the individual hardware components of a computer system so that other software and the users of the system see it as a functional unit without having to be concerned with the low-level details such as transferring data from memory to disk, or rendering text onto a display. Generally, system software consists of an operating system and some fundamental utilities such as disk formatters, file managers, display managers, text editors, user authentication (login) and management tools, and networking and device control software.

B. **Application software** is used to accomplish specific tasks other than just running the computer system. Application software may consist of a single program, such as an image viewer; a small collection of programs (often called a software package) that work closely together to accomplish a task, such as a spreadsheet or text processing system; a larger collection (often called a software suite) of related but independent programs and packages that have a common user interface or shared data format, such as Microsoft Office, which consists of closely integrated word processor, spreadsheet, database, etc.; or a software system, such as a database management system, which is a collection of fundamental programs that may provide some service to a variety of other independent applications.

Relationship between Hardware and Software

- Hardware and software are mutually dependent on each other. Both of them must work together to make a computer produce a useful output.
- Software cannot be utilized without supporting hardware.
- Hardware without a set of programs to operate upon cannot be utilized and is useless.
- To get a particular job done on the computer, relevant software should be loaded into the hardware.
- Hardware is a one-time expense.
- Software development is very expensive and is a continuing expense.
- Different software applications can be loaded on a hardware to run different jobs.
- A software acts as an interface between the user and the hardware.
- If the hardware is the 'heart' of a computer system, then the software is its 'soul'. Both are complementary to each other.

Classification of Computers

Computers can be generally classified by size and power as follows, though there is Considerable overlap:

• Personal computer: A small, single-user computer based on a microprocessor. In addition to the microprocessor, a personal computer has a keyboard for entering data, a monitor for displaying information, and a storage device for saving data.

• workstation: A powerful, single-user computer. A workstation is like a personal computer, but it has a more powerful microprocessor and a higher-quality monitor.

• minicomputer: A multi-user computer capable of supporting from 10 to hundreds of users simultaneously.

• mainframe : A powerful multi-user computer capable of supporting many hundreds or thousands of users simultaneously.

• supercomputer : An extremely fast computer that can perform hundreds of millions of instructions per second.

Characteristics of Computer

Speed, accuracy, diligence, storage capability and versatility are some of the key characteristics of a computer. A brief overview of these characteristics are :

• Speed: The computer can process data very fast, at the rate of millions of instructions per second. Some calculations that would have taken hours and days to complete otherwise, can be completed in a few seconds using the computer. For example, calculation and generation of salary slips of thousands of employees of an organization, weather forecasting that requires analysis of a large amount of data related to temperature, pressure and humidity of various places, etc.

• Accuracy: Computer provides a high degree of accuracy. For example, the computer can accurately give the result of division of any two numbers up to 10 decimal places.

• Diligence: When used for a longer period of time, the computer does not get tired or fatigued. It can perform long and complex calculations with the same speed and accuracy from the start till the end.

• Storage Capability: Large volumes of data and information can be stored in the computer and also retrieved whenever required. A limited amount of data can be stored, temporarily, in the primary memory. Secondary storage devices like floppy disk and compact disk can store a large amount of data permanently.

• Versatility: Computer is versatile in nature. It can perform different types of tasks with the same ease. At one moment you can use the computer to prepare a letter document and in the next moment you may play music or print a document. Computers have several limitations too. Computer can only perform tasks that it has been programmed to do.

COMPUTER MEMORY

A memory is just like a human brain. It is used to store data and instructions. Computer memory is the storage space in the computer, where data is to be processed and instructions required for processing are stored. The memory is divided into large number of small parts called cells. Each location or cell has a unique address, which varies from zero to memory size minus one. For example, if the computer has 64k words, then this memory unit has 64 * 1024 = 65536 memory locations. The address of these locations varies from 0 to 65535.

Memory is primarily of three types −

- Cache Memory
- Primary Memory/Main Memory
- Secondary Memory

Cache Memory

Cache memory is a very high speed semiconductor memory which can speed up the CPU. It acts as a buffer between the CPU and the main memory. It is used to hold those parts of data and program which are most frequently used by the CPU. The parts of data and programs are transferred from the disk to cache memory by the operating system, from where the CPU can access them.

Advantages

The advantages of cache memory are as follows −

- Cache memory is faster than main memory.
- It consumes less access time as compared to main memory.
- It stores the program that can be executed within a short period of time.
- It stores data for temporary use.

Disadvantages

The disadvantages of cache memory are as follows −

- Cache memory has limited capacity.
- It is very expensive.

Primary Memory (Main Memory)

Primary memory holds only those data and instructions on which the computer is currently working. It has a limited capacity and data is lost when power is switched off. It is generally made up of semiconductor device. These memories are not as fast as registers. The data and instruction required to be processed resides in the main memory. It is divided into two subcategories RAM and ROM.

Characteristics of Main Memory

- These are semiconductor memories.
- It is known as the main memory.
- Usually volatile memory.
- Data is lost in case power is switched off.
- It is the working memory of the computer.
- Faster than secondary memories.
- A computer cannot run without the primary memory.

Secondary Memory

This type of memory is also known as external memory or non-volatile. It is slower than the main memory. These are used for storing data/information permanently. CPU directly does not access these memories; instead they are accessed via input-output routines. The contents of secondary memories are first transferred to the main memory, and then the CPU can access it. For example, disk, CD-ROM, DVD, etc.

Characteristics of Secondary Memory

- These are magnetic and optical memories.
- It is known as the backup memory.
- It is a non-volatile memory.
- Data is permanently stored even if power is switched off.
- It is used for storage of data in a computer.
- Computer may run without the secondary memory.
- Slower than primary memories.

RAM (Random Access Memory) is the internal memory of the CPU for storing data, program, and program result. It is a read/write memory which stores data until the machine is working. As soon as the machine is switched off, data is erased.

Access time in RAM is independent of the address, that is, each storage location inside the memory is as easy to reach as other locations and takes the same amount of time. Data in the RAM can be accessed randomly but it is very expensive.

RAM is volatile, i.e. data stored in it is lost when we switch off the computer or if there is a power failure. Hence, a backup Uninterruptible Power System (UPS) is often used with computers. RAM is small, both in terms of its physical size and in the amount of data it can hold.

RAM is of two types −

- Static RAM (SRAM)
- Dynamic RAM (DRAM)

1) **Static RAM (SRAM)**

The word **static** indicates that the memory retains its contents as long as power is being supplied. However, data is lost when the power gets down due to volatile nature. SRAM chips use a matrix of 6-transistors and no capacitors. Transistors do not require power to prevent leakage, so SRAM need not be refreshed on a regular basis.

There is extra space in the matrix, hence SRAM uses more chips than DRAM for the same amount of storage space, making the manufacturing costs higher. SRAM is thus used as cache memory and has very fast access.

Characteristic of Static RAM

- Long life
- No need to refresh
- Faster
- Used as cache memory
- Large size
- Expensive
- High power consumption

Dynamic RAM (DRAM)

DRAM, unlike SRAM, must be continually **refreshed** in order to maintain the data. This is done by placing the memory on a refresh circuit that rewrites the data several hundred times per second. DRAM is used for most system memory as it is cheap and small. All DRAMs are made up of memory cells, which are composed of one capacitor and one transistor.

Characteristics of Dynamic RAM

- Short data lifetime
- Needs to be refreshed continuously
- Slower as compared to SRAM
- Used as RAM
- Smaller in size
- Less expensive
- Less power consumption

Difference between SRAM and DRAM

SRAM	DRAM
1. SRAM has lower access time, so it is faster compared to DRAM.	1. DRAM has higher access time, so it is slower than SRAM.
2. SRAM is costlier than DRAM.	2. DRAM costs less compared to SRAM.
3. SRAM requires constant power supply, which means this type of memory consumes more power.	3. DRAM offers reduced power consumption, due to the fact that the information is stored in the capacitor.
4. Due to complex internal circuitry, less storage capacity is available compared to the same physical size of DRAM memory chip.	4. Due to the small internal circuitry in the one-bit memory cell of DRAM, the large storage capacity is available.
5. SRAM has low packaging density.	5. DRAM has high packaging density.

ROM stands for **Read Only Memory**. The memory from which we can only read but cannot write on it. This type of memory is non-volatile. The information is stored permanently in such memories during manufacture. A ROM stores such instructions that are required to start a computer. This operation is referred to as **bootstrap**. ROM chips are not only used in the computer but also in other electronic items like washing machine and microwave oven.

The various types of ROMs and their characteristics.

1) **MROM (Masked ROM)**

The very first ROMs were hard-wired devices that contained a pre-programmed set of data or instructions. These kind of ROMs are known as masked ROMs, which are inexpensive.

2) **PROM (Programmable Read Only Memory)**

PROM is read-only memory that can be modified only once by a user. The user buys a blank PROM and enters the desired contents using a PROM program. Inside the PROM chip, there are small fuses which are burnt open during programming. It can be programmed only once and is not erasable.

3) **EPROM (Erasable and Programmable Read Only Memory)**

EPROM can be erased by exposing it to ultra-violet light for a duration of up to 40 minutes. Usually, an EPROM eraser achieves this function. During programming, an electrical charge is trapped in an insulated gate region. The charge is retained for more than 10 years because the charge has no leakage path. For erasing this charge, ultra-violet light is passed through a quartz crystal window (lid). This exposure to ultra-violet light dissipates the charge. During normal use, the quartz lid is sealed with a sticker.

4) **EEPROM (Electrically Erasable and Programmable Read Only Memory)**

EEPROM is programmed and erased electrically. It can be erased and reprogrammed about ten thousand times. Both erasing and programming take about 4 to 10 ms (millisecond). In EEPROM, any location can be selectively erased and programmed. EEPROMs can be erased one byte at a time, rather than erasing the entire chip. Hence, the process of reprogramming is flexible but slow.

Advantages of ROM

The advantages of ROM are as follows −

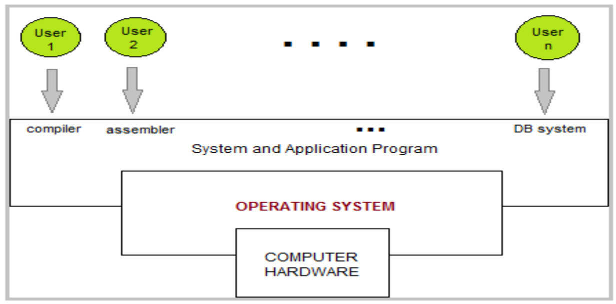
- Non-volatile in nature
- Cannot be accidentally changed
- Cheaper than RAMs
- Easy to test
- More reliable than RAMs
- Static and do not require refreshing
- Contents are always known and can be verified

UNIT- II

Introduction to Operating Systems

A computer system has many resources (hardware and software), which may be require to complete a task. The commonly required resources are input/output devices, memory, file storage space, CPU etc. The operating system acts as a manager of the above resources and allocates them to specific programs and users, whenever necessary to perform a particular task. Therefore operating system is the resource manager i.e. it can manage the resource of a computer system internally. The resources are processor, memory, files, and I/O devices. **In simple terms, an operating system is the interface between the user and the machine.**

Four Components of a Computer System



Two Views of Operating System

1. User's View
2. System View

Operating System: User View

The user view of the computer refers to the interface being used. Such systems are designed for one user to monopolize its resources, to maximize the work that the user is performing. In these cases, the operating system is designed mostly for ease of use, with some attention paid to performance, and none paid to resource utilization.

Operating System: System View

Operating system can be viewed as a resource allocator also. A computer system consists of many resources like - hardware and software - that must be managed efficiently. The operating system acts as the manager of the resources, decides between conflicting requests, controls execution of programs etc.

Operating System Management Tasks

1. **Processor management** which involves putting the tasks into order and pairing them into manageable size before they go to the CPU.
2. **Memory management** which coordinates data to and from RAM (random-access memory) and determines the necessity for virtual memory.
3. **Device management** which provides interface between connected devices.

4. **Storage management** which directs permanent data storage.
5. **Application** which allows standard communication between software and your computer.
6. **User interface** which allows you to communicate with your computer.

Functions of Operating System

1. It boots the computer
2. It performs basic computer tasks e.g. managing the various peripheral devices e.g. mouse, keyboard
3. It provides a user interface, e.g. command line, graphical user interface (GUI)
4. It handles system resources such as computer's memory and sharing of the central processing unit(CPU) time by various applications or peripheral devices.
5. It provides file management which refers to the way that the operating system manipulates, stores, retrieves and saves data.
6. Error Handling is done by the operating system. It takes preventive measures whenever required to avoid errors.

Types of Operating Systems

Following are some of the most widely used types of Operating system.

1. Simple Batch System
2. Multiprogramming Batch System
3. Multiprocessor System
4. Desktop System
5. Distributed Operating System
6. Clustered System
7. Real time Operating System
8. Handheld System

---

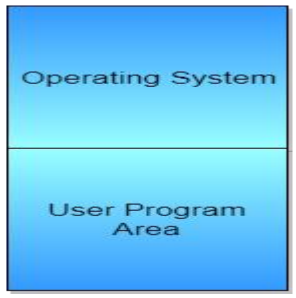
1) Simple Batch Systems

- In this type of system, there is **no direct interaction between user and the computer**.
- The user has to submit a job (written on cards or tape) to a computer operator.
- Then computer operator places a batch of several jobs on an input device.
- Jobs are batched together by type of languages and requirement.

- Then a special program, the monitor, manages the execution of each program in the batch.
- The monitor is always in the main memory and available for execution.

Advantages of Simple Batch Systems

1. No interaction between user and computer.
2. No mechanism to prioritise the processes.

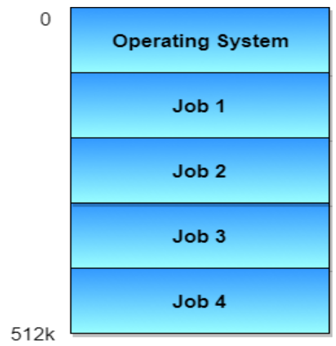


2) Multiprogramming Batch Systems

- In this the operating system picks up and begins to execute one of the jobs from memory.
- Once this job needs an I/O operation operating system switches to another job (CPU and OS always busy).
- Jobs in the memory are always less than the number of jobs on disk(Job Pool).
- If several jobs are ready to run at the same time, then the system chooses which one to run through the process of **CPU Scheduling**.
- In Non-multiprogrammed system, there are moments when CPU sits idle and does not do any work.
- In Multiprogramming system, CPU will never be idle and keeps on processing.

Time Sharing Systems are very similar to Multiprogramming batch systems. In fact time sharing systems are an extension of multiprogramming systems.

In Time sharing systems the prime focus is on **minimizing the response time**, while in multiprogramming the prime focus is to maximize the CPU usage.



3) Multiprocessor Systems

A Multiprocessor system consists of several processors that share a common physical memory. Multiprocessor system provides higher computing power and speed. In multiprocessor system all processors operate under single operating system. Multiplicity of the processors and how they do act together are transparent to the others.

Advantages of Multiprocessor Systems

1. Enhanced performance
2. Execution of several tasks by different processors concurrently, increases the system's throughput without speeding up the execution of a single task.
3. If possible, system divides task into many subtasks and then these subtasks can be executed in parallel in different processors. Thereby speeding up the execution of single tasks.

4) Desktop Systems

Earlier, CPUs and PCs lacked the features needed to protect an operating system from user programs. PC operating systems therefore were neither **multiuser** nor **multitasking**. However, the goals of these operating systems have changed with time; instead of maximizing CPU and peripheral utilization, the systems opt for maximizing user convenience and responsiveness. These systems are called **Desktop Systems** and include PCs running Microsoft Windows and the Apple Macintosh. Operating systems for these computers have benefited in several ways from the development of operating systems for **mainframes**.

Microcomputers were immediately able to adopt some of the technology developed for larger operating systems. On the other hand, the hardware costs for microcomputers are sufficiently **low** that individuals have sole use of the computer, and CPU utilization is no longer a prime concern. Thus, some of the design decisions made in operating systems for mainframes may not be appropriate for smaller systems.

---

5) Distributed Operating System

The motivation behind developing distributed operating systems is the availability of powerful and inexpensive microprocessors and advances in communication technology.

These advancements in technology have made it possible to design and develop distributed systems comprising of many computers that are inter connected by communication networks. The main benefit of distributed systems is its low price/performance ratio.

Advantages Distributed Operating System

1. As there are multiple systems involved, user at one site can utilize the resources of systems at other sites for resource-intensive tasks.
2. Fast processing.
3. Less load on the Host Machine.

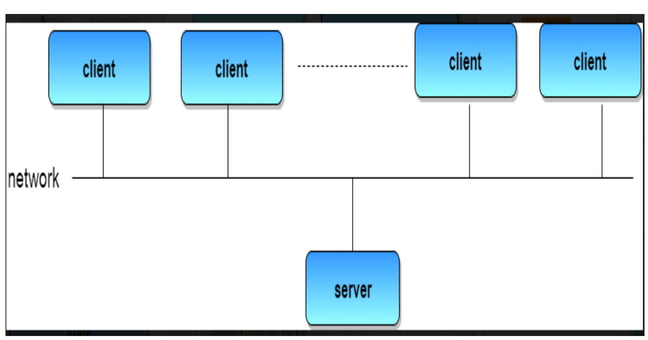
Types of Distributed Operating Systems

Following are the two types of distributed operating systems used:

1. Client-Server Systems
2. Peer-to-Peer Systems

Client-Server Systems

Centralized systems today act as **server systems** to satisfy requests generated by **client systems**. The general structure of a client-server system is depicted in the figure below:



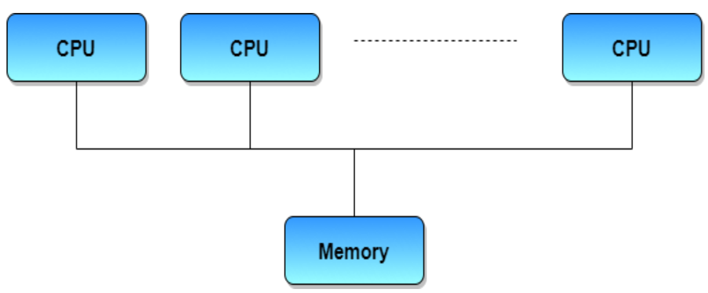
Server Systems can be broadly categorized as: **Compute Servers** and **File Servers**.

- **Compute Server systems**, provide an interface to which clients can send requests to perform an action, in response to which they execute the action and send back results to the client.
- **File Server systems**, provide a file-system interface where clients can create, update, read, and delete files.

Peer-to-Peer Systems

The growth of computer networks - especially the Internet and World Wide Web (WWW) – has had a profound influence on the recent development of operating systems. When PCs were introduced in the 1970s, they were designed for **personal** use and were generally considered standalone computers. With the beginning of widespread public use of the Internet in the 1990s for electronic mail and FTP, many PCs became connected to computer networks.

In contrast to the **Tightly Coupled** systems, the computer networks used in these applications consist of a collection of processors that do not share memory or a clock. Instead, each processor has its own local memory. The processors communicate with one another through various communication lines, such as high-speed buses or telephone lines. These systems are usually referred to as loosely coupled systems (or distributed systems). The general structure of a client-server system is depicted in the figure below:



6) Clustered Systems

- Like parallel systems, clustered systems gather together multiple CPUs to accomplish computational work.
- Clustered systems differ from parallel systems, however, in that they are composed of two or more individual systems coupled together.
- The definition of the term clustered is **not concrete;** the general accepted definition is that clustered computers share storage and are closely linked via LAN networking.
- Clustering is usually performed to provide **high availability**.

- A layer of cluster software runs on the cluster nodes. Each node can monitor one or more of the others. If the monitored machine fails, the monitoring machine can take ownership of its storage, and restart the application(s) that were running on the failed machine. The failed machine can remain down, but the users and clients of the application would only see a brief interruption of service.
- **Asymmetric Clustering -** In this, one machine is in hot standby mode while the other is running the applications. The hot standby host (machine) does nothing but monitor the active server. If that server fails, the hot standby host becomes the active server.
- **Symmetric Clustering -** In this, two or more hosts are running applications, and they are monitoring each other. This mode is obviously more efficient, as it uses all of the available hardware.
- **Parallel Clustering -** Parallel clusters allow multiple hosts to access the same data on the shared storage. Because most operating systems lack support for this simultaneous data access by multiple hosts, parallel clusters are usually accomplished by special versions of software and special releases of applications.

Clustered technology is rapidly changing. Clustered system's usage and it's features should expand greatly as **Storage Area Networks(SANs)**. SANs allow easy attachment of multiple hosts to multiple storage units. Current clusters are usually limited to two or four hosts due to the complexity of connecting the hosts to shared storage.

---

7) Real Time Operating System

It is defined as an operating system known to give maximum time for each of the critical operations that it performs, like OS calls and interrupt handling.

The Real-Time Operating system which guarantees the maximum time for critical operations and complete them on time are referred to as **Hard Real-Time Operating Systems.**

While the real-time operating systems that can only guarantee a maximum of the time, i.e. the critical task will get priority over other tasks, but no assurity of completeing it in a defined time. These systems are referred to as **Soft Real-Time Operating Systems**.

---

8) Handheld Systems

Handheld systems include **Personal Digital Assistants(PDAs)**, such as Palm-Pilots or Cellular Telephones with connectivity to a network such as the Internet. They are usually of limited size due to which most handheld devices have a small amount of memory, include slow processors, and feature small display screens.

- Many handheld devices have between **512 KB** and **8 MB** of memory. As a result, the operating system and applications must manage memory efficiently. This includes returning all allocated memory back to the memory manager once the memory is no longer being used.
- Currently, many handheld devices do **not use virtual memory** techniques, thus forcing program developers to work within the confines of limited physical memory.
- Processors for most handheld devices often run at a fraction of the speed of a processor in a PC. Faster processors require **more power**. To include a faster processor in a handheld device would require a **larger battery** that would have to be replaced more frequently.
- The last issue confronting program designers for handheld devices is the small display screens typically available. One approach for displaying the content in web pages is **web clipping**, where only a small subset of a web page is delivered and displayed on the handheld device.

Some handheld devices may use wireless technology such as **BlueTooth**, allowing remote access to e-mail and web browsing. **Cellular telephones** with connectivity to the Internet fall into this category. Their use continues to expand as network connections become more available and other options such as cameras and MP3 players, expand their utility.

BOOTING PROCESS

When we start our Computer then there is an operation which is performed automatically by the Computer which is also called as Booting. In the Booting, System will check all the hardware's and Software's those are installed or Attached with the System and this will also load all the Files those are needed for running a system.

In the Booting Process all the Files those are Stored into the ROM Chip will also be Loaded for Running the System. In the Booting Process the System will read all the information from the Files those are Stored into the ROM Chip and the ROM chip will read all the instructions those are Stored into these Files. After the Booting of the System this will automatically display all the information on the System. The Instructions those are necessary to Start the System will be read at the Time of Booting.

There are two Types of Booting

1) **Warm Booting**: when the System Starts from the Starting or from initial State Means when we Starts our System this is called as warm Booting. In the Warm Booting the System will be started from its beginning State means first of all, the user will press the Power Button, then this will read all the instructions from the ROM and the Operating System will b automatically gets loaded into the System.

2) **Cold Booting**: The Cold Booting is that in which System Automatically Starts when we are Running the System, For Example due to Light Fluctuation the system will Automatically Restarts So that in this Chances Damaging of system are More. And the System will no be start from its initial State So May Some Files will b Damaged because they are not Properly Stored into the System.

TRANSLATORS AND LANGUAGES

A program translator is a computer program that performs the translation of a program written in a given programming language into a functionally equivalent program in a different computer language, without losing the functional or logical structure of the original code (the "essence" of each program).

These include translations between high-level and human-readable computer languages such as C++, Java and COBOL, intermediate-level languages such as Java bytecode, low-level languages such as the assembly language and machine code, and between similar levels of language on different computing platforms, as well as from any of these to any other of these.

They also include translators between software implementations and hardware/ASIC microchip implementations of the same program, and from software descriptions of a microchip to the logic gates needed to build it.

1. COMPILERS

A compiler is a computer program (or set of programs) that transforms source code written in a programming language (the source language) into another computer language (the target language, often having a binary form known as object code).

The most common reason for converting a source code is to create an executable program.

The name "compiler" is primarily used for programs that translate source code from a high-level programming language to a lower level language (e.g., assembly language or machine code).

If the compiled program can run on a computer whose CPU or operating system is different from the one on which the compiler runs, the compiler is known as a cross-compiler. More generally, compilers are a specific type of translators.

A program that translates from a low level language to a higher level one is a decompiler.

A program that translates between high-level languages is usually called a source-to-source compiler or transpiler.

A language rewriter is usually a program that translates the form of expressions without a change of language.

The term compiler-compiler is sometimes used to refer to a parser generator, a tool often used to help create the lexer and parser.

A compiler is likely to perform many or all of the following operations:

1. Lexical analysis,
2. Preprocessing,
3. Parsing,
4. Semantic analysis (syntax-directed translation),
5. Code generation, and code optimization.

Program faults caused by incorrect compiler behavior can be very difficult to track down and work around; therefore, compiler implementors invest significant effort to ensure compiler correctness.

Compilers enabled the development of programs that are machine-independent.

Before the development of FORTRAN, the first higher-level language, in the 1950s, machine-dependent assembly language was widely used.

While assembly language produces more abstraction than machine code on the same architecture, just as with machine code, it has to be modified or rewritten if the program is to be executed on different computer hardware architecture.

With the advent of high-level programming languages that followed FORTRAN, such as COBOL, C, and BASIC, programmers could write machine-independent source programs. A compiler translates the high-level source programs into target programs in machine languages for the specific hardware. Once the target program is generated, the user can execute the program.

ADVANTAGES OF COMPILER

1. Source code is not included, therefore compiled code is more secure than interpreted code.
2. Tends to produce faster code than interpreting source code.
3. Produces an executable file, and therefore the program can be run without need of the source code.

DISADVANTAGES OF COMPILER

1. Object code needs to be produced before a final executable file, this can be a slow process.
2. The source code must be 100% correct for the executable file to be produced.

2. INTERPRETERS

In computer science, an ***interpreter*** is a computer program that directly executes, i.e. *performs*, instructions written in a programming or scripting language, without previously compiling them into a machine language program.

An *interpreter* is a program that reads in as input a source program, along with data for the program, and translates the source program instruction by instruction.

EXAMPLE

- The Java interpreter java translate a .class file into code that can be executed natively on the underlying machine.
- The program VirtualPC interprets programs written for the Intel Pentium architecture (IBM-PC clone) for the PowerPC architecture (Macintosh). This enable Macintosh users to run Windows programs on their computer.

An interpreter generally uses one of the following strategies for program execution:

1. parse the source code and perform its behavior directly.

2. translate source code into some efficient intermediate representation and immediately execute this.

3. explicitly execute stored precompiled code made by a compiler which is part of the interpreter system.

APPLICATIONS

1. Interpreters are frequently used to execute command languages, and glue languages since each operator executed in command language is usually an invocation of a complex routine such as an editor or compiler.

2. Self-modifying code can easily be implemented in an interpreted language. This relates to the origins of interpretation in Lisp and artificial intelligence research.

3. Virtualization. Machine code intended for one hardware architecture can be run on another using a virtual machine, which is essentially an interpreter.

4. Sandboxing: An interpreter or virtual machine is not compelled to actually execute all the instructions the source code it is processing. In particular, it can refuse to execute code that violates any security constraints it is operating under.

ADVANTAGES OF INTERPRETER

1. Easier to debug(check errors) than a compiler.

2. Easier to create multi-platform code, as each different platform would have an interpreter to run the same code.

3. Useful for prototyping software and testing basic program logic.

DISADVANTAGES OF INTERPRETER

1. Source code is required for the program to be executed, and this source code can be read making it insecure.

2. Interpreters are generally slower than compiled programs due to the per-line translation method.

3. ASSEMBLERS

An assembler translates assembly language into machine code.

An assembler is a program that creates object code by translating combinations of mnemonics and syntax for operations and addressing modes into their numerical equivalents.

Assembly language

- It consists of mnemonics for machine opcodes so assemblers perform a 1:1 translation from mnemonic to a direct instruction.
- An assembly language (or assembler language) is a low-level programming language for a computer, or other programmable device, in which there is a very strong (generally one-to-one) correspondence between the language and the architecture's machine code instructions.
- Each assembly language is specific to a particular computer architecture, in contrast to most high-level programming languages, which are generally portable across multiple architectures, but require interpreting or compiling.
- Assembly language is converted into executable machine code by a utility program referred to as an assembler; the conversion process is referred to as assembly, or assembling the code.

For example:

`LDA #4` converts to `0001001000100100`

Conversely, one instruction in a high level language will translate to one or more instructions at machine level.

TYPES OF ASSEMBLERS

There are two types of assemblers based on how many passes through the source are needed to produce the executable program.

1. **One-pass assemblers** go through the source code once. Any symbol used before it is defined will require "errata" at the end of the object code (or, at least, no earlier than the point

where the symbol is defined) telling the linker or the loader to "go back" and overwrite a placeholder which had been left where the as yet undefined symbol was used.

2. **Multi-pass assemblers** create a table with all symbols and their values in the first passes, then use the table in later passes to generate code.

In both cases, the assembler must be able to determine the size of each instruction on the initial passes in order to calculate the addresses of subsequent symbols.

This means that if the size of an operation referring to an operand defined later depends on the type or distance of the operand, the assembler will make a pessimistic estimate when first encountering the operation, and if necessary pad it with one or more "no-operation" instructions in a later pass or the errata. In an assembler with peephole optimization, addresses may be recalculated between passes to allow replacing pessimistic code with code tailored to the exact distance from the target.

The original reason for the use of one-pass assemblers was speed of assembly – often a second pass would require rewinding and rereading a tape or rereading a deck of cards.

With modern computers this has ceased to be an issue. The advantage of the multi-pass assembler is that the absence of errata makes the linking process (or the program load if the assembler directly produces executable code) faster.

APPLICATIONS OF ASSEMBLERS

1. Assembly language is typically used in a system's boot code, the low-level code that initializes and tests the system hardware prior to booting the operating system and is often stored inROM. (BIOS on IBM-compatible PC systems and CP/M is an example.)

2. Some compilers translate high-level languages into assembly first before fully compiling, allowing the assembly code to be viewed for debugging and optimization purposes.

3. Relatively low-level languages, such as C, allow the programmer to embed assembly language directly in the source code. Programs using such facilities, such as the Linux kernel, can then construct abstractions using different assembly language on each hardware platform. The system's portable code can then use these processor-specific components through a uniform interface.

4. Assembly language is useful in reverse engineering. Many programs are distributed only in machine code form which is straightforward to translate into assembly language, but more difficult to translate into a higher-level language. Tools such as the Interactive Disassembler make extensive use of disassembly for such a purpose.

5. Assemblers can be used to generate blocks of data, with no high-level language overhead, from formatted and commented source code, to be used by other code.

ADVANTAGES OF ASSEMBLER:

1. Very fast in translating assembly language to machine code as 1 to 1 relationship.
2. Assembly code is often very efficient (and therefore fast) because it is a low level language.
3. Assembly code is fairly easy to understand due to the use of English-like mnemonics.

DISADVANTAGES OF ASSEMBLERS:

1. Assembly language is written for a certain instruction set and/or processor.
2. Assembly tends to be optimised for the hardware it's designed for, meaning it is often incompatible with different hardware.
3. Lots of assembly code is needed to do relatively simple tasks, and complex programs require lots of programming time.

DIFFERENCE BETWEEN COMPILERS INTERPRETERS AND ASSEMBLERS

BASIS	COMPILERS	INTERPRETER	ASSEMBLER
1. DEFINITION	A compiler is a computer program that converts an entire program written in a high-level language (called source code) and translates it into an executable form (called object code).	An interpreter is a computer program that takes source code and converts each line in succession.	Assembler converts assembly languages to machine code than high-level programs languages.

2. INPUT	Compiler Takes Entire program as input	Interpreter Takes Single instruction as input .	Input source program in Assembly Language through an input device.
3. MEMORY REQUIREMENT	Memory Requirement : More(Since Object Code is Generated)	Memory Requirement is Less	
ERRORS	Errors are displayed after entire program is checked.	Errors are displayed for every instruction interpreted (if any)	Error messages generated during an assembly may originate from the assembler or from a higher level language such as C (many assemblers are written in C) or from the operating system environment

Computer language and types of Computer Language

Around the world language is a source of communication among human beings. Similarly, in order to communicate with computer user also needs to have a language, that should be understandable by the computers. For the purpose different languages are developed for performing different types of work on the computer.

Mainly there are two types of computer languages:

1) Low Level Computer Languages

These are machine codes or close to it. Computer cannot understand instruction given in high level languages or in English. It can only understand and execute instructions given in the form of machine languages i.e. the binary number 0 and 1.

There are two types of low level computer language.

Machine Language

The lowest and most elementary language and was the first type of programming language to be developed. Mache language is basically the only language which computer can understand. In fact, a manufacturer designs a computer to obey just one language, its machine code, which is represented inside the computer by a string of binary digits (bits) 0 and 1. The symbol 0 stands for the absence of an electric pulse and 1 for the presence of an electric pulse. Since a computer is capable of recognizing electric signals, therefore, it understands machine language.

The set of binary codes which can be recognize by the computer is known as the machine code instruction set. A machine language instruction consists of an operation code one or more operands. The operation code specifies that operation that is to be performed e.g. read, record etc. the operands identify the quantities to be operated on e.g. the numbers to be added or the locations where data are stored. But, it is almost impossible to write programs directly in machine code. For this reason, programs are normally written in assembly or high level languages and then are translated in the machine language by different translators.

Advantages

1. It makes fast and efficient use of the computer
2. It requires no translator to translate the code i.e. directly understood by the computer.

Disadvantages

1. All operation codes have to be remembered
2. All memory addresses have to be remembered
3. It is hard to amend or find errors in a program written in the machine language
4. These languages are machine dependent i.e. a particular machine language can be used on only one type of computer.

Assembly Languages

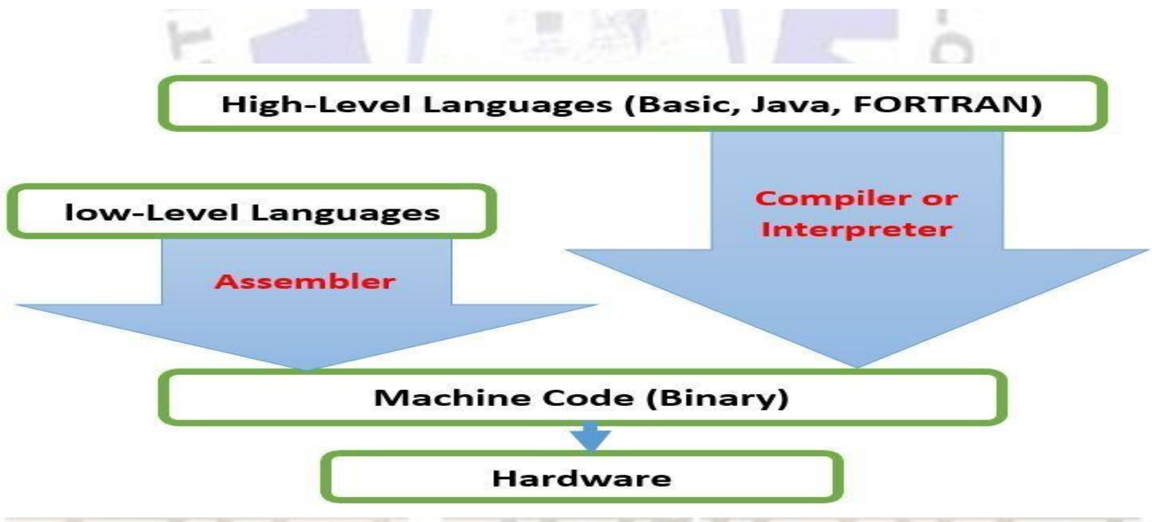
It was developed to overcome some of the many inconveniences of machine language. This is another low level but a very important language in which operation codes and operands are given in the form of alphanumeric symbols instead of 0's and 1's. These alphanumeric symbols will be known as mnemonic codes and can have maximum up to 5 letter combinations e.g. ADD for addition, SUB for subtraction, START LABEL etc. because of this feature it is also known as "Symbolic Programming Language". This language is very difficult and needs a lot of practice to master it because very small English support is given. This symbolic language helps in compiler orientations. The instructions of the assembly language will also be converted to machine codes by language translator to be executed by the computer

Advantages

1. It is easier to understand and use as compared to machine language
2. It is easy to locate and correct errors
3. It is modified easily

Disadvantages

1. Like machine language it is also machine dependent
2. Since it is machine dependent, there programmer should have the knowledge of he hardware also.



2) Computer High Level Languages

High level computer languages give formats close to English language and the purpose of developing high level languages is to enable people to write programs easily and in their own native language environment (English). High-level languages are basically symbolic languages that use English words and/or mathematical symbols rather than mnemonic codes. Each instruction in the high level language is translated into many machine language instructions thus showing one-to-many translation.

- **Problem-Oriented Language:** These are languages used for handling specialized types of data processing problems where programmer only specifies the input/output requirements and other relative information of the problem, that are to be solved. The programmer does not have to specify the procedure to be followed in solving that particular problem.
- **Procedural Language:** These are general purpose languages that are designed to express the logic of a data processing problem.
- **Non-procedural Language:** Computer Programming Languages that allow users and professional programmers to specify the results they want without specifying how to solve the problem.

Selection of a Computer Language

There are many high-level languages in use today. The choice of language depends on type and complexity of the problem.

1. Purpose e.g. business, educational, scientific.
2. Facilities provided e.g. meaningful variable names, control and data structures, error checking facilities.
3. Ease of learning and use.
4. Portability - if a program is to he used on or in more than one system.
5. Popularity - availability of compilers/interpreters.
6. Documentation provided.

Types of High Level Languages

Many languages have been developed for achieving different variety of tasks, some are fairly specialized others are quite general purpose. These are categorized according to their use as:

Algebraic Formula-Type Processing. These languages are oriented towards the computational procedures for solving mathematical and statistical problems. Examples are

1. BASIC (Basic All Purpose Symbolic Instruction Code)
2. FORTRAN (Formula Translation).
3. PL/1 (Programming Language, Version 1).
4. ALGOL (Algorithmic Language).
5. APL (A Programming Language).

Business Data Processing. These languages emphasize their capabilities for maintaining data processing procedures and files handling problems. Examples are:

1. COBOL (Common Business Oriented Language)
2. RPG (Report Program Generator).

String and List Processing. These are used for string manipulation including search for patterns, inserting and deleting characters. Examples are: LISP (List Processing).

Multipurpose Language. A general purpose language used for algebraic procedures, data and string processing. Examples are:

1. Pascal (after the name of Blaise Pascal).
2. PL/1 (Programming Language, version 1).
3. C language.

Simulation: These may be written in algebraic or multipurpose languages. Examples are:

1. SPSS (Statistical Package System Simulator).
2. GPSS (General Purpose System Simulator).

Advantages: Following are the advantages of a high level language:

1. User-friendly (people based)
2. Similar to English with vocabulary of words awl symbols therefore it is easier to learn.
3. They require less time to write
4. They are easier to maintain
5. Problem oriented' rather than 'machine' based
6. Shorter than their low-level equivalents. One statement translates into many machine code instructions.
7. Program written in a high-level equivalent can he translated into many machine language and therefore can run on every computer for which there exists an appropriate translator.
8. It is independent of the machine on which it used i.e. programs developed in high level language can be run on any computer.

Disadvantages: There are certain disadvantages also Inspite these disadvantages high-level languages have proved their worth. The advantages out-weigh the disadvantages by far, for most applications. These are:

1. A high-level language has to be translated into the -machine language by a translator and thus a price in computer time is paid.
2. The object code generated by a translator might be inefficient compared to an equivalent assembly language program.

INTRODUCTION TO GUI

A **graphical user interface** is fondly called "GUI" pronounced "gooey." The word "graphical" means pictures; "user" means the person who uses it; "interface" means what you see on the screen and how you work with it. So a graphical user interface, then, means that you (the user) get to work with little pictures on the screen to boss the computer around, rather than type in lines of codes and commands.

(GUI) An INTERACTIVE outer layer presented by a computer software product (for example an operating system) to make it easier to use by operating through pictures as well as words. Graphical user interfaces employ visual metaphors, in which objects drawn on the computer's screen mimic in some way the behaviour of real objects, and manipulating the screen object controls part of the program.

A graphical user interface uses *menus* and *icons* (pictorial representations) to choose commands, start applications, make changes to documents, store files, delete files, etc. You can use the mouse to control a cursor or pointer on the screen to do these things, or you can alternatively use the keyboard to do most actions. A graphical user interface is considered *user-friendly*.

The most popular GUI metaphor requires the user to point at pictures on the screen with an arrow pointer steered by a MOUSE or similar input device. Clicking the MOUSE BUTTONS while pointing to a screen object selects or activates that object, and may enable it to be moved across the screen by dragging as if it were a real object.

GUIs have many advantages and some disadvantages. They make programs much easier to learn and use, by exploiting natural hand-to-eye coordination instead of numerous obscure command sequences. They reduce the need for fluent typing skills, and make the operation of software more comprehensible and hence less mysterious and anxiety- prone. For visually-oriented tasks such as word processing, illustration and graphic design they have proved revolutionary.

GUIs can also present great difficulties for people with visual disabilities, and their interactive nature makes it difficult to automate repetitive tasks by batch processing. Neither do GUIs *automatically* promote good user interface design. Hiding 100 poorly-chosen commands behind the tabs of a property sheet is no better than hiding them among an old-fashioned menu hierarchy - the point is to reduce them to 5 more sensible ones.

Historically, the invention of the GUI must be credited to Xerox PARC where the first GUI based workstations - the XEROX STAR and XEROX DORADO - were designed in the early 1970s. These proved too expensive and too radical for commercial exploitation, but it was following a visit to PARC by Steve Jobs in the early 1980s that Apple released the LISA, the first commercial GUI computer, and later the more successful MACINTOSH. It was only following the 1990 release of Windows version 3.0 that GUIs became ubiquitous on IBM-compatible PCs.

UNIT III

INTRODUCTION TO WORD PROCESSOR

Word processing is an application program that allows you to create letters, reports, newsletters, tables, form letters, brochures, and Web pages. Using this application program you can add pictures, tables, and charts to your documents. You can also check spelling and grammar. A word processor is an electronic device or computer application software that performs word processing: the composition, editing, formatting and sometimes printing of any sort of written material. Word processing can also refer to advanced shorthand techniques, sometimes used in specialized contexts with a specially modified typewriter. The term was coined at IBM's Böblingen, West Germany Laboratory in the 1960s. Typical features of a word processor include font application, spell checking, grammar checking, a built-in thesaurus, automatic text correction, Web integration and HTML exporting, among others.

The word processor emerged as a stand-alone office machine in the 1970s and 1980s, combining the keyboard text-entry and printing functions of an electric typewriter with a dedicated computer processor for the editing of text. Although features and designs varied among manufacturers and models, and new features were added as technology advanced, word processors typically featured a monochrome display and the ability to save documents on memory cards or diskettes. Later models introduced innovations such as spell-checking programs, improved formatting options, and dot-matrix printing. As the more versatile combination of personal computers and printers became commonplace, and computer software applications for word processing became popular, most business machine companies stopped manufacturing word processor machines. As of 2009 there were only two U.S. companies, Classic and Alpha Smart, which still made them. Many older machines, however, remain in use. Since 2009, Sentinel has offered a machine described as a "word processor", but it is more accurately a highly specialized microcomputer used for accounting and publishing.

Word processors are descended from early text formatting tools (sometimes called "text justification" tools, from their only real capability). Word processing was one of the earliest applications for the personal computer in office productivity. Although early processors take advantage of a graphical user interface providing some form of what you-see-is-what-you-get editing. Most are powerful systems consisting of one or more programs that can produce any arbitrary combination of images, graphics and text, the latter handled with type-setting capability.

Microsoft Word is the most widely used word processing software. Microsoft estimates that over 500,000,000 people use the Microsoft Office suite, which includes Word. Many other word processing applications exist, including WordPerfect (which dominated the market from the mid-1980s to early-1990s on computers running Microsoft's MS-DOS operating system) and open source applications OpenOffice.org Writer, Libre Office Writer, Abi Word, KWord, and LyX. Web-based word processors, such as Office Web Apps or Google Docs, are a relatively new category.

Main features of word processing applications:

- Create professional documents fast, using built-in and custom templates.
- Easily manage large documents using various features like the ability to create table of contents, index, and cross-references.
- Work on multiple documents simultaneously.
- With the help of mail merge, you can quickly create merge documents like mass mailings or mailing labels.

- AutoCorrect and AutoFormat features catch typographical errors automatically and allow you to use predefined shortcuts and typing patterns to quickly format your documents.
- The print zoom facility scales a document on different paper sizes, and allows you to print out multiple pages on a single sheet of paper.
- The nested tables feature supports putting one table inside another table.
- Export and save your word documents in PDF and XPS file format.
- Batch mailings using form letter template and an address database (also called mail merging);
- Indices of keywords and their page numbers;
- Tables of contents with section titles and their page numbers;
- Tables of figures with caption titles and their page numbers;
- Cross-referencing with section or page numbers;
- Footnote numbering;
- New versions of a document using variables (e.g. model numbers, product names, etc.)

Other word processing functions include spell checking (actually checks against wordlists), "grammar checking" (checks for what seem to be simple grammar errors), and a "thesaurus" function (finds words with similar or opposite meanings). Other common features include collaborative editing, comments and annotations, support for images and diagrams and internal cross-referencing.

Features of word:

MS Word has useful features and tools introduced to produce professionally created documents. You can easily create, format, edit professional-looking user document using comprehensive set of easy to use tools provided by MS Word. It uses the MS Office Fluent user Interface concept. This interface uses a new component called Ribbon to group the tools by task, within task by sub tasks and related commands that are used more frequently. The new user result oriented interface presents the tools to you in a more organised and efficient manner, which are easy to locate.

1. Tabs are more task oriented such as Home, Insert, Page Layout
2. Within each tab, the related sub-tasks are grouped together
3. Related command buttons are also grouped together to execute a command or to display a command menu

Microsoft Office Word helps you produce professional-looking documents by providing a comprehensive set of tools for creating and formatting your document in a new interface. Rich review, commenting, and comparison capabilities help you quickly gather and manage feedback from colleagues. Advanced data integration ensures that documents stay connected to important sources of business information.

The MS Word 2007 provides a lot of pre-formatted template to produce documents, reports etc. While using the pre-formatted template, you can select already available cover page, header and footer to give the documents a professional look without spending much time in formatting a new one. MS Word 2007 also provides features for creating chart and diagram which include three-dimensional shapes, transparency, drop shadows, and other effects. This helps create highly professional documents with flexibility in representing data more efficiently and professionally. Before sharing a document which is in its final form with others, you can use MS Word 2007 "Mark As Final" features to protect the document from any changes. "Mark as Final" command makes the document "read-only" making the typing, editing and proofing command disabled. MS Word 2007 also provides the feature and tools to export your document to either PDF (Portable Document Format) or XPS (XML Paper Specification) format.

ADVANTAGES OF WORD PROCESSING

The advantages of Word processing are synonymous with the benefits provided MS-Word. The most prominent ones are enlisted below:

Create professional-looking documents

Office Word 2007 provides editing and reviewing tools for creating polished documents more easily than ever before.

Spend more time writing, less time formatting

A new, results-oriented interface presents tools to you when you need them, in a clear and organized fashion:

- Save time and get more out of the powerful Word capabilities by selecting from galleries of predefined styles, table formats, list formats, graphical effects, and more.

- Word eliminates the guesswork when you apply formatting to your document. The galleries of formatting choices give you a live visual preview of the formatting in your document before you commit to making a change.

Add preformatted elements with just a few clicks

Office Word 2007 introduces building blocks for adding preformatted content to your documents:

- When you are working on a document from a particular template type, such as a report, you can select from a gallery of preformatted cover pages, pull quotes, and headers and footers to make your document look more polished.
- If you want to customize the preformatted content, or if your organization often uses the same piece of content, such as legal disclaimer text or customer contact information, you can create your own building blocks that you select from the gallery with a single click.

Communicate more effectively with high-impact graphics

New charting and diagramming features include three-dimensional shapes, transparency, drop shadows, and other effects.

Instantly apply a new look and feel to your documents

When your company updates its look, you can instantly follow suit in your documents. By using Quick Styles and Document Themes, you can quickly change the appearance of text, tables, and graphics throughout your document to match your preferred style or color scheme.

Easily avoid spelling errors

The following are some new features of the spelling checker:

- The spelling checker has been made more consistent across the 2007 Microsoft Office system programs. Examples of this change include:
- Several spelling checker options are now global. If you change one of these options in one Office program, that option is also changed for all the other Office programs.
- In addition to sharing the same custom dictionaries, all programs can manage them using the same dialog box. The 2007 Microsoft Office system spelling checker includes the post-reform French dictionary. In Microsoft Office 2003, this was an add-in that had to be separately installed.
- An exclusion dictionary is automatically created for a language the first time that language is used. Exclusion dictionaries let you force the spelling checker flag words you want to avoid using. They are handy for avoiding words that are obscene or that don't match your style guide

- The spelling checker can find and flag some contextual spelling errors. Have you ever typed a mistake similar to the following? I will see you their. In Office Word 2007, you can enable the Use contextual spelling option to get help with finding and fixing this type of mistake. This option is available when checking the spelling of documents in English, German or Spanish.
- You can disable spelling and grammar checking for a document or for all documents you create.

Share documents confidently

When you send a draft of a document to your colleagues for their input, Office Word 2007 helps you efficiently collect and manage their revisions and comments. When you are ready to publish the document, Office Word 2007 helps you ensure that any unresolved revisions and comments aren't still lurking in the published document.

Quickly compare two versions of a document

Office Word 2007 makes it easy to find out what changes were made to a document. When you compare and combine documents, you can see both versions of the document — with the deleted, inserted, and moved text clearly marked in a third version of the document.

Find and remove hidden metadata and personal information in documents

Before you share your document with other people, you can use the Document Inspector to check the document for hidden metadata, personal information, or content that may be stored in the document. The Document Inspector can find and remove information like comments, versions, tracked changes, ink annotations, document properties, document management server information, hidden text, custom XML data, and information in headers and footers. The Document Inspector can help you ensure that the documents you share with other people do not contain any hidden personal information or any hidden content that your organization might not want distributed. Additionally, your organization can customize the Document Inspector to add checks for additional types of hidden content.

Add a digital signature or signature line to your documents

You can help provide assurance as to the authenticity, integrity, and origin of your document by adding a digital signature to the document. In Office Word 2007 you can either add an invisible digital signature to a document, or you can insert a Microsoft Office Signature Line to capture a visible representation of a signature along with a digital signature.

The ability to capture digital signatures by using signature lines in Office documents makes it possible for organizations to use paperless signing processes for documents like contracts or other agreements. Unlike signatures on paper, digital signatures provide a record of exactly what was signed and they allow the signature to be verified in the future.

Convert your Word documents to PDF or XPS

Office Word 2007 supports exporting your file to the following formats:

- **Portable Document Format (PDF)** PDF is a fixed-layout electronic file format that preserves document formatting and enables file sharing. The PDF format ensures that when the file is viewed online or printed, it retains exactly the format that you intended, and that data in the file cannot be easily changed. The PDF format is also useful for documents that will be reproduced by using commercial printing methods
- **XML Paper Specification (XPS)** XPS is an electronic file format that preserves document formatting and enables file sharing. The XPS format ensures that when the file is viewed online or printed, it retains exactly the format that you intended, and that data in the file cannot be easily changed.

Instantly detect documents that contain embedded macros

Office Word 2007 uses a separate file format (.docm) for macro-enabled documents, so you can instantly tell whether a file is capable of running any embedded macros.

Prevent changes to a final version of a document

Before you share a final version of a document with other people, you can use the **Mark As Final** command to make the document read-only and communicate to other people that you are sharing a final version of the document. When a document is marked as final, typing, editing commands, and proofing marks are disabled, and people who view the document cannot inadvertently change the document. The **Mark As Final** command is not a security feature. Anyone can edit a document that is marked as final by turning off **Mark As Final**.

Reduce file sizes and improve corruption recovery

The new Word XML format is a compressed, segmented file format that offers a dramatic reduction in file size and helps ensure that damaged or corrupted files can be easily recovered.

Connect your documents to business information

In your business, you create documents to communicate important business data. You can save time and reduce the risk of error by automating the process of this communication. Create dynamic smart documents that update themselves by using new document controls and data binding to connect to your back-end systems.

Manage document properties in the Document Information Panel

The Document Information Panel makes it easy to view and edit document properties while you work on your Word document. The Document Information Panel displays at the top of your document in Word. You can use the Document Information Panel to view and edit both standard Microsoft Office document properties and properties for files that are saved to a document management server. If you use the Document Information Panel to edit the document properties for a server document, the updated properties will be saved directly to the server.

For example, you may have a server that keeps track of a document's editorial status. When you put the finishing touches on a document, you can open the Document Information Panel to change the document's editorial status from Draft to Final. When you save the document back on the server, the change in editorial status is updated on the server.

If you store document templates in a library on a Microsoft Windows SharePoint Services 3.0 server, the library might include custom properties that store information about the templates. For example, your organization may require you to categorize documents in the library by filling in a Category property. Using the Document Information Panel, you can edit properties like this directly within the Word environment.

Recover from computer problems

2007 Microsoft Office system provides improved tools for recovering your work in the event of a problem in Office Word 2007.

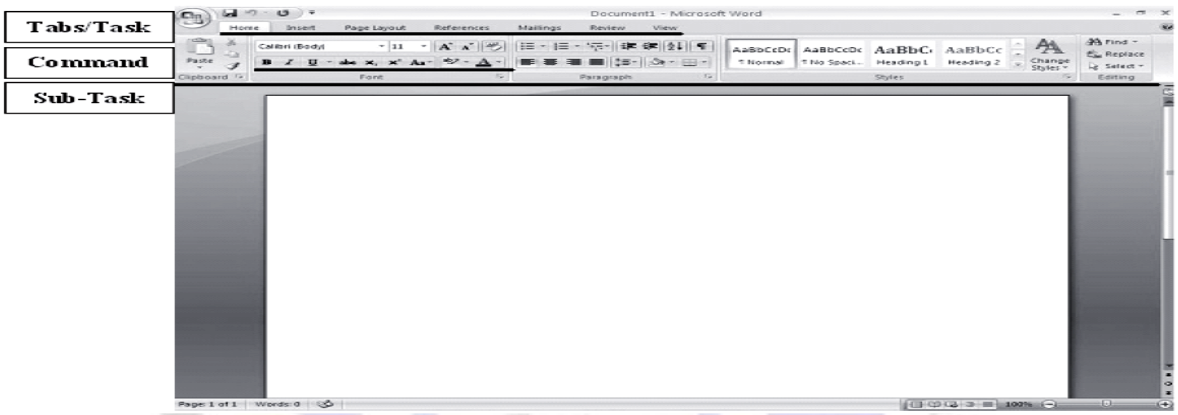
Office Diagnostics: Microsoft Office Diagnostics is a series of diagnostic tests that can help you to discover why your computer is crashing. The diagnostic tests can solve some problems directly and may identify ways that you can solve other problems. Microsoft Office Diagnostics replaces the following Microsoft Office 2003 features: Detect and Repair and Microsoft Office Application Recovery.

Program recovery: Office Word 2007 has improved capabilities to help avoid losing work when the program closes abnormally. Whenever possible, Word tries to recover some aspects of the state of the program after it restarts. For example, you are working on several files at the same time. Each file is open in a different window with specific data visible in each window.

Word crashes. When you restart Word, it opens the files and restores the windows to the way they were before Word crashed.

THE WORD SCREEN LAYOUT

The Word screen (Window) contains a number of objects such as Tabs, Menus, Sub menus, short-cut commands etc.



ADVANCED EXCEL

Excel is a software application of almost unlimited depth and complexity. There is no way to count the number of features it offers, but they must run into the thousands, if not tens of thousands. Fortunately, most business analysts do not need to become experts in all aspects of Excel, but they often find that they need to develop expertise in a few specialized domains within Excel. Which domains are relevant depends on the job they perform as well as on their level of interest in Excel and spreadsheet modeling generally.

1) CONTROLS

Excel controls allow the user to change the contents or behaviour of a spreadsheet without interacting directly with individual cells. Controls such as boxes and buttons are familiar because they appear frequently in commonly used windows in the Excel user interface. Among the Excel options, for example, the window for formula options uses a button to select a Workbook Calculation Mode (only one of the three available choices is allowed) and a check box (on or off) to Enable Iterative Calculation.

To place a control on a spreadsheet requires a sequence of steps. First, click on the desired control icon from the toolbar. Using the cursor, which now appears as a cross, drag and drop the control to the desired location in the spreadsheet. In doing so, use the cursor to let the size of the control as well as its location.

	For moving and scrolling
Ctrl +arrow key	Move to the edge of the current data region
Home	Move to the beginning of the row
Ctrl +Home	Move to the beginning of the worksheet (A1)
Ctrl +End	Move to the bottom-right corner of the used area of the worksheet
PgDn	Move down one screen
PgUp	Move up one screen
Alt +PgDn	Move one screen to the right
Alt +PgUp	Move one screen to the left
F5	Display the Go To dialog box
	For entering data on a worksheet
Alt + Enter	Opens a new workbook, if one is already open
Shift + Enter	Complete a cell entry and move up one cell
Tab	Complete a cell entry and move to the right cell
Shift +Tab	Complete a cell entry and move to the left cell
Ctrl + Delete	Delete text to the end of the line
Shift +F2	Edit a cell comment
Ctrl +D	Fill down (a selected column of cells with the content of the first cell)
Ctrl +R	Fill to the right (a selected row of cells with the content of the first cell)
Ctrl +F3	Open the Define Name dialog box
	For working in cells or the formula bar
Ctrl +Shift +Enter	Enter a formula as an array formula
F2	Edit the active cell
F3	Open the Paste Name window
Shift +F3	Open the Insert Function (or Function Arguments) window
F9	Calculate all sheets in all open workbooks
Ctrl + Alt +F9	Calculate all worksheets in the active workbook
Shift + F9	Calculate the active worksheet

Ctrl +; (semicolon)	Enter the current date
Ctrl + Shift + : (colon)	Enter the current time
Ctrl +~	Display all formulas
	For inserting, deleting and copying selection
Ctrl +C	Copy the selection
Ctrl +X	Cut the selection
Ctrl +V	Paste the selection
Delete	Clear the contents of the selection
Ctrl + -(hyphen)	Delete (dialog box)
Ctrl +Z	Undo the last action
Ctrl +Shift + Plus sign	Insert (dialog box)
	For selecting cells, columns or rows
Shift +arrow key	Extend the selection by one cell
Ctrl + Shift +arrow key	Extend the selection to the last nonblank cell in the same column or row
Ctrl +space bar	Select the entire column
Ctrl +A	Select the entire worksheet
	For working with worksheets and macros
Shift +F11	Insert a new worksheet
Alt + F8	Display the Macro dialog box
Alt + F11	Display the Visual Basic Editor (VBE)
Ctrl +PgDn	Move to the next sheet in the workbook
Ctrl + PgUp	Move to the previous sheet in the workbook
	Miscellaneous
Ctrl +S	Save an active workbook

Ctrl +N	Open a new workbook
Ctrl +O	Open an existing workbook
Shift +F5 or Ctrl +F	Display the find dialog box
Ctrl +H	Display the Replace dialog box

UNIT IV

Data Communication

When we communicate, we are sharing information. This sharing can be local or remote. Between individuals, local communication usually occurs face to face, while remote communication takes place over distance. The term **Telecommunication**, which includes Telephony, Telegraphy, and television, means communication at a distance. The data refers to facts, concepts and instruction presented in whatever form is agreed upon by the parties creating and using the data. In the context of computer information system, data represented by binary information units produced and consumed in the form of 0s and 1s.

Data Communications is the transfer of data or information between a source and a receiver. The source transmits the data and the receiver receives it. The actual generation of the information is not part of Data Communications nor is the resulting action of the information at the receiver. Data Communication is interested in the transfer of data, the method of transfer and the preservation of the data during the transfer process.

The purpose of Data Communications is to provide the rules and regulations that allow computers with different disk operating systems, languages, cabling and locations to share resources. The rules and regulations are called protocols and standards in Data Communications.

For data communication to occur, the communicating devices must be part of a communication system made up of a combination of hardware and software. The effectiveness of a data communication system depends on the three fundamental characteristics:

1. Delivery: The System must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user

2. Accuracy: The system must deliver data accurately. Data that have been altered in transmission and left uncorrected are rustles

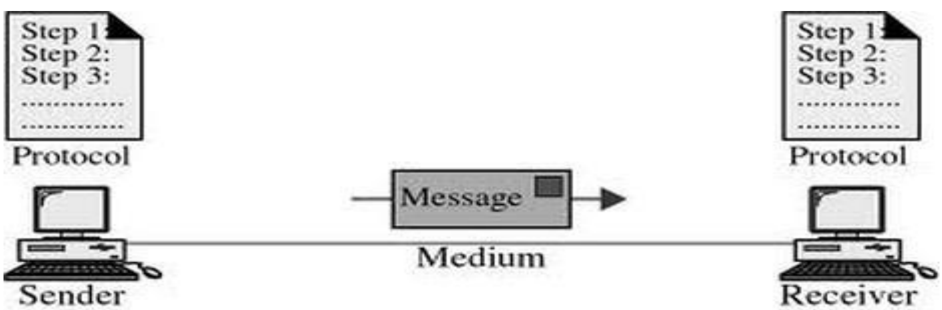
3. Timeliness: The system must deliver data in a timely manner. Data delivered late are useless. In the case of video, audio, and voice data, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery id called real-time transmission.

Components

Basic Components of a Communication System

The following are the basic requirements for working of a communication system.

1. The sender (source) who creates the message to be transmitted
2. A medium that carries the message
3. The receiver (sink) who receives the message



Data Communication system components

1. Message: A **message** in its most general meaning is an object of communication. It is a vessel which provides information. Yet, it can also be this information. Therefore, its meaning is dependent upon the context in which it is used; the term may apply to both the information and its form.

2. Sender: The sender will have some kind of meaning she wishes to convey to the receiver. It might not be conscious knowledge, it might be a sub-conscious wish for communication. What is desired to be communicated would be some kind of idea, perception, feeling, or datum. It will be a part of her reality that she wishes to send to somebody else.

3. Receiver: These messages are delivered to another party. No doubt, you have in mind a desired action or reaction you hope your message prompts from the opposite party. Keep in mind, the other party also enters into the communication process with ideas and feelings that will undoubtedly influence their understanding of your message and their response. To be a successful communicator, you should consider these before delivering your message, then acting appropriately.

4. Medium: Medium is a means used to exchange / transmit the message. The sender must choose an appropriate medium for transmitting the message else the message might not be conveyed to the desired recipients. The choice of appropriate medium of communication is essential for making the message effective and correctly interpreted by the recipient. This choice of communication medium varies depending upon the features of communication. For instance - Written medium is chosen when a message has to be conveyed to a small group of people, while an oral medium is chosen when spontaneous feedback is required from the recipient as misunderstandings are cleared then and there.

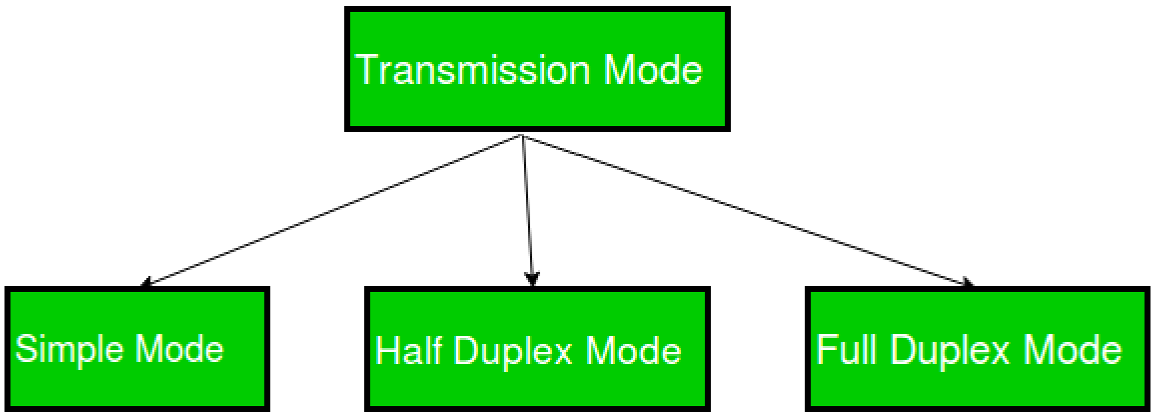
5. Protocol: A **protocol** is a formal description of digital message formats and the rules for exchanging those messages in or between computing systems and in telecommunications. Protocols may include signaling, authentication and error detection and correction syntax, semantics, and synchronization of communication and may be implemented in hardware or software, or both.

6. Feedback: Feedback is the main component of communication process as it permits the sender to analyze the efficacy of the message. It helps the sender in confirming the correct interpretation of message by the decoder. Feedback may be verbal (through words) or non-verbal (in form of smiles, sighs, etc.). It may take written form also in form of memos, reports, etc.

<u>Transmission Modes in Computer Networks (Simplex, Half-Duplex and Full-Duplex)</u>

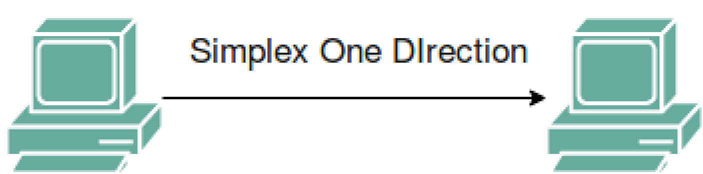
Transmission mode means transferring of data between two devices. It is also known as communication mode. Buses and networks are designed to allow communication to occur between individual devices that are interconnected. There are three types of transmission mode:-

- **Simplex Mode**
- **Half-Duplex Mode**
- **Full-Duplex Mode**



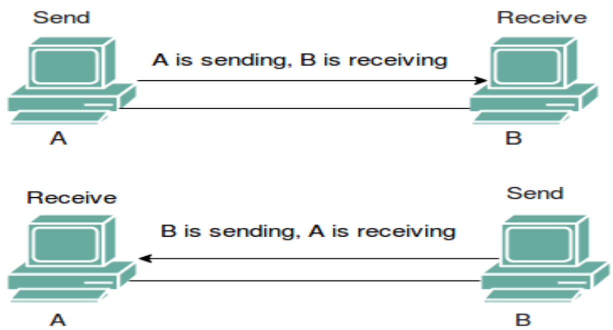
Simplex Mode

In Simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit, the other can only receive. The simplex mode can use the entire capacity of the channel to send data in one direction. Example: Keyboard and traditional monitors. The keyboard can only introduce input, the monitor can only give the output.



Half-Duplex Mode

In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa. The half-duplex mode is used in cases where there is no need for communication in both direction at the same time. The entire capacity of the channel can be utilized for each direction. Example: Walkie- talkie in which message is sent one at a time and messages are sent in both the directions.

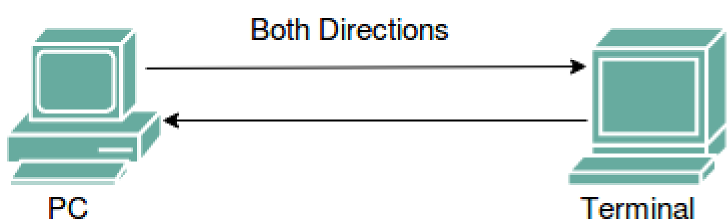


Full-DuplexMode

In full-duplex mode, both stations can transmit and receive simultaneously. In full_duplex mode, signals going in one direction share the capacity of the link with signals going in other direction, this sharing can occur in two ways:

- Either the link must contain two physically separate transmission paths, one for sending and other for receiving.
- Or the capacity is divided between signals travelling in both directions.

Full-duplex mode is used when communication in both direction is required all the time. The capacity of the channel, however must be divided between the two directions. Example: Telephone Network in which there is communication between two persons by a telephone line, through which both can talk and listen at the same time.

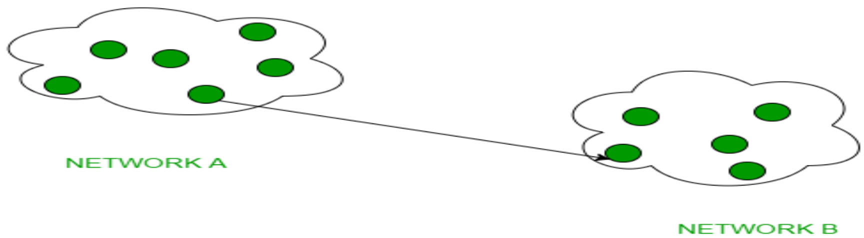


Difference between Unicast, Broadcast and Multicast

The **cast** term here signifies some data(stream of packets) is being transmitted to the recipient(s) from client(s) side over the communication channel that help them to communicate. Let's see some of the "cast" concepts that are prevailing in the computer networks field.

1. Unicast –

This type of information transfer is useful when there is a participation of single sender and single recipient. So, in short you can term it as a one-to-one transmission. For example, a device having IP address 10.1.2.0 in a network wants to send the traffic stream (data packets) to the device with IP address 20.12.4.2 in the other network, then unicast comes into picture. This is the most common form of data transfer over the networks.



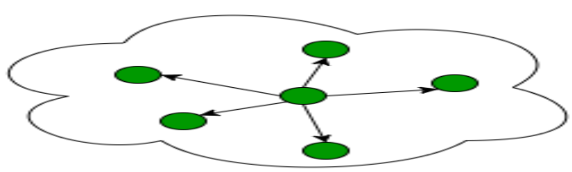
UNICAST EXAMPLE

2. Broadcast –

Broadcasting transfer (one-to-all) techniques can be classified into two types :

- **LimitedBroadcasting** –

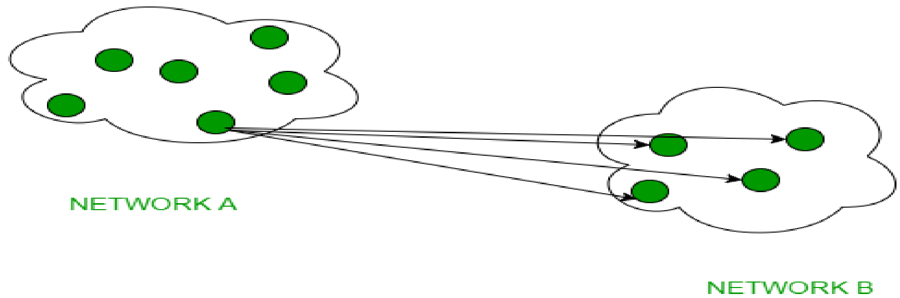
Suppose you have to send stream of packets to all the devices over the network that you reside, this broadcasting comes handy. For this to achieve,it will append 255.255.255.255 (all the 32 bits of IP address set to 1) called as **Limited Broadcast Address** in the destination address of the datagram (packet) header which is reserved for information tranfer to all the recipients from a single client (sender) over the network.



NETWORK CLUSTER

- **Direct Broadcasting** –

This is useful when a device in one network wants to transfer packet stream to all the devices over the other network. This is achieved by translating all the Host ID part bits of the destination address to 1,referred as **Direct Broadcast Address** in the datagram header for information transfer.



This mode is mainly utilized by television networks for video and audio distribution. One important protocol of this class in Computer Networks is Address Resolution Protocol (ARP) that is used for resolving IP address into physical address which is necessary for underlying communication.

3. Multicast –

In multicasting, one/more senders and one/more recipients participate in data transfer traffic. In this method traffic recline between the boundaries of unicast (one-to-one) and broadcast (one-to-all). Multicast lets server's direct single copies of data streams that are then simulated and routed to hosts that request it. IP multicast requires support of some other protocols like **IGMP (Internet Group Management Protocol), Multicast routing** for its working. Also in Classful IP addressing **Class D** is reserved for multicast groups.

Network: Network is a communication system which supports many users. Network is a broad term similar to 'system'.

Computer Network: Computer network is a system which allows communication among the computers connected in the network.

Protocol:

- For successful communication to occur, it is not enough for the 'sender' to simply transmit the message and assume that the 'receiver' will receive it properly.
- There are certain rules that must be followed to ensure proper communication.
- A set of such rules is known as a 'protocol' of the data communication system.
- Many different protocols are used in the modern data communication system.
- The interconnection of one station to many stations is called as 'networking'.
- A network is any interconnection of two or more stations that wish to communicate.

Node: Each station in a communication network is called as a node. The nodes are connected in different way to each other to form a network.

Many other forms of interconnection are possible. The most familiar network is the Telephone System. It is the largest and most sophisticated network of all.

Introduction to Computer Networks:

- During 20th century the most important technology has been the information gathering, its processing and distribution.
- The computers and communication have been merged together and their merger has had a profound effect on the manner in which computer systems are organized.
- The old computer in which a single computer used to serve all the computational needs of an organisation has been replaced by a new one in which a large number of separate but interconnected computers do the job.
- Such systems are called as Computer Networks.
- Two systems are said to be interconnected if they interchange information. The connection between the separate computers can be done via a copper wire, fiber optics, microwaves or communication satellite.

Distributed Systems:

- If one computer can forcibly start, stop or control another the computer are not autonomous. A system with one control unit and many slaves, or a large computer with remote printers and terminals is not called a computer network, it is called a Distributed System.
- In distributed system, the existence of multiple autonomous computers is not visible to the user. The user can type a command to run a program and it runs.
- It is up to operating system to select the best processor, find and transport all the input files to that processor and give the results to the user.
- With a computer network, the user must explicitly (definitely) log onto one machine, explicitly submit jobs remotely, explicitly moves files around and generally handle all the network management personally.
- With a distributed system, nothing has to be done explicitly; it is all automatically done by the system without the user's knowledge.
- **Basically, a distributed system is a software system built on top of a network. The software gives it a high degree of cohesiveness (sticking together) and transparency.**

Internet:

The network formed by the co-operative interconnection of a large number of computer networks.

- Network of Networks
- No one owns the Internet
- Every person who makes a connection owns a slice of the Internet.
- There is no central administration of the Internet.

Internet is comprises of:

A community of people: who use and develop the network.

A collection of resources: that can be reached from those networks.

A setup to facilitate collaboration: Among the members of the research and educational communities worldwide.

The connected networks use the TCP/IP protocols:

Important Internet applications:

World wide web(WWW)

File Transfer Protocol(FTP)

Electronic Mail

Internet Relay Chat

Intranet:

A private TCP/IP internetwork within an organization that uses Internet technologies such as Web servers and Web browsers for sharing information and collaborating. Intranets can be used to publish company policies and newsletters, provide sales and marketing staff with product information, provide technical support and tutorials, and just about anything else you can think of that fits within the standard Web server/Web browser environment.

Intranet Web servers differ from public Web servers in that the public must have the proper permissions and passwords to access the intranet of an organization. Intranets are designed to permit users who have access privileges to the internal LAN of the organization. Within an intranet, Web servers are installed in the network. Browser technology is used as the common front end to access information on servers such as financial, graphical, or text-based data.

Extranet:

Extranets refer to applications and services that are Intranet based, and use extended, secure access to external users or enterprises. This access is usually accomplished through passwords, user IDs, and other application level security. An extranet is the extension of two or more intranet strategies with a secure interaction between participant enterprises and their respective intranets.

Part of a Company's Intranet that is extended to users outside the company (eg. Normally over the Internet). In its simplest form, a private TCP/IP network that securely shares information using Hypertext Transfer Protocol (HTTP) and other Internet protocols with business partners such as vendors, suppliers, and wholesale customers. An extranet is thus a corporate intranet that is exposed over the Internet to certain specific groups that need
access to it. Extranets built in this fashion follow the client/server paradigm, with Web servers such as Apache.

Extranets are a powerful tool because they let businesses share resources on their own private networks over the Internet with suppliers, vendors, business partners, or customers. Extranets are typically used for supporting realtime supply chains, for enabling business partners to work together, or to share information such as catalogs with
customers. The power of the extranet is that it leverages the existing technology of the Internet to increase the power, flexibility, and competitiveness of businesses utilizing well-known and easily used tools such as Web servers and Web browsers. Extranets also save companies money by allowing them to establish business-to business connectivity over the Internet instead of using expensive, dedicated leased lines. Extranets can also save money by reducing phone and fax costs.

Network Topology:

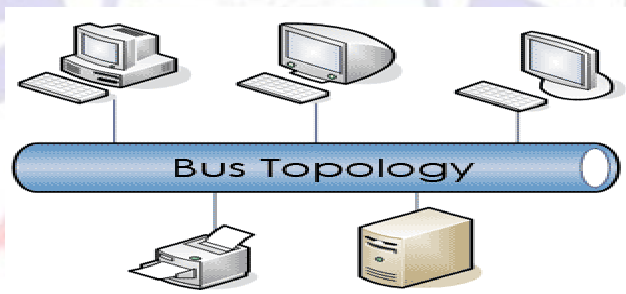
- The meaning of physical topology indicates the way in which a network is physically laid out.
- Two or more devices connect to a link; two or more links form a topology.
- The topology of a network is the geometric representation of the relationship of all the links connecting the devices (or nodes).

- There are five basic network topologies-

i. Mesh Topology
ii. Star Topology
iii. Bus Topology
iv. Ring Topology
v. Tree Topology

- While selecting one of the above five topologies we have to consider the relative status of the device to be linked.
- Two relationships on the basis are possible as follows:
- ✓ **Peer to peer:** It is the relationship where the devices share the link equally. The examples are ring and mesh topologies.
- ✓ **Primary-secondary:** It is the relationship in which one device controls and the other devices have to transmit through it. The examples are star and tree topology.

1. Bus Topology



- The bus topology is usually used when a network installation is small, simple or temporary.
- When one computer sends a signal up the cable; all the computers on the network receive the information, but the one with the address that matches the one encoded in the message accepts the information while all the others reject the message.
- The speed of the bus topology is slow because only one computer can send a message at a time. A computer must wait until the bus is free before it can transmit.
- The bus topology requires a proper termination at both the ends of the cable.
- Since the bus is a passive topology, the electrical signal from a transmitting computer is free to travel the entire length of the cable.

- Without termination when the signal reaches the end of the cable, it returns back and travels back up the cable.
- The transmitted waves and reflected waves, if they are in phase add and if they are out of phase cancel.
- Thus adding and cancellation of the wave leads to what is called a standing wave.
- The standing waves can distort the normal signals which are travelling along the cable.
- The terminators absorb the electrical energy and stop reflections.

Characteristics of the bus topology:

- ✓ The signal strength of the transmitted signal should be adequately high so as to meet the minimum signal strength requirements of the receiver.
- ✓ Adequate signal to noise ratio should be maintained.
- ✓ This is a multipoint configuration. There are more than two devices connected to the medium and they are capable of transmitting on the medium. Hence the medium access control is essential for the bus topology.
- ✓ The signal should not be too strong. This is necessary to avoid the overloading of transmitter and hence the possibility of signal distortion.
- ✓ This is called as signal balancing which is not an easy task at all. Specially the signal balancing becomes increasingly difficult with increase in the number of stations.
- ✓ The solution to this problem is to divide the transmission medium into small segments and within each such segment a pair wise balancing is done using amplifiers or repeaters between the segments.

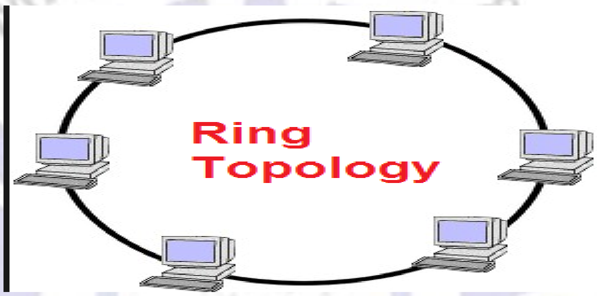
Advantages of bus topology:

- ✓ The bus topology is easy to understand, install and use for small networks.
- ✓ The cabling cost is less as the bus topology requires the least amount of cable to connect the computers.
- ✓ The bus topology is easy to expand by joining two cables with a BNC barrel connector.
- ✓ In the expansion of a bus topology repeaters can be used to boost the signal and increase the distance.

Disadvantages of bus topology:

- ✓ Heavy network traffic slows down the bus speed. In bus topology only one computer can transmit and other have to wait till their turn comes and there is no co-ordination between computers for reservation of transmitting time slot.
- ✓ The BNC connectors used for expansion of the bus attenuates the signal considerably.
- ✓ A cable break or loose BNC connector will cause reflections and bring down the whole network causing all network activity to stop.

2. Ring topology



- ➢ In a ring topology, each computer is connected to the next computer, with the last one connected to the first.
- ➢ Rings are used in high performance networks where large bandwidth is necessary.
- ➢ Every computer is connected to the next computer in the ring and each retransmits what it receives from the previous computer hence the ring is an active network.
- ➢ The message flow around the ring in one direction. There is no termination because there is no end to the ring.
- ➢ **Some ring networks do token passing. A short message called a token is passed around the ring until a computer wishes to send information to another computer.**
- ➢ That computer modifies the token, adds an electronic address and data and sends it around the ring.
- ➢ The receiving computer returns a message to the originator indicating that the message has been received.
- ➢ The sending computer then creates another token and places it on the network, allowing another station to capture the token and begin transmitting.
- ➢ The token circulates until a station is ready to send and capture the token. Faster networks circulate several tokens at once.

- Some ring networks have two counter-rotating rings that help them recover from network faults.

Characteristics of ring LANs:

- The data is transferred in a sequential manner bit by bit around the ring. Each repeater will regenerate and retransmit each bit.

Functions of a ring:

- ✓ A ring can operate as a communication network if it performs the following three functions:

1) Data insertion
2) Data reception
3) Data removal

These functions are actually provided by the repeaters.

- ✓ Each repeater also acts as the device attachment point. Hence the function of data insertion is accomplished by the repeaters.
- ✓ Data is transmitted in the form of packets.
- ✓ Each packet consists of a destination address field. As this packet by a repeater, the destination address field is copied by the repeater.
- ✓ If the destination address field corresponds to the address of a device then that repeater copies the remaining contents of packet as well.
- ✓ Data insertion and reception can be done easily by the repeaters but data removal is more difficult on a ring.
- ✓ As the ring is a closed loop, a packet will circulate on it indefinitely if it is not removed.
- ✓ A packet can be removed by the addressed repeater or each packet can be removed by the transmitting repeater itself after the packet has made one trip around the ring.
- ✓ The second approach is more desirable.

Problems faced in the ring topology:

- If any link breaks or if any repeater fails then the entire network will be disabled.
- To install a new repeater for supporting a new device, it is necessary to have the identification of two nearby, topologically adjacent repeaters.
- It is necessary to take preventive measures to deal with the time jitter.
- Due to the closed nature of the ring topology it is necessary to remove the circulating packets. These problems except for the last one can be rectified by refinements of the ring topology.

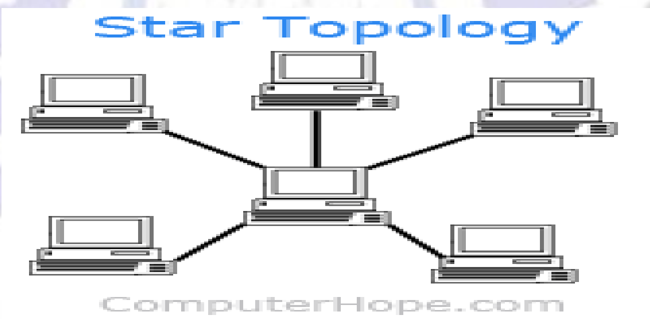
Advantages of ring topology:

- No one computer can monopolise the network because every computer is given equal across to the token.
- The fair sharing of the network allows the network to continue function in a useful, if slower, manner rather than fail once capacity is exceeded as more users are added.

Disadvantages of ring topology:

- Failure of one computer on the ring can affect the whole network.
- It is difficult to trouble short the ring.
- Adding or removing the computers disturbs the network activity.

3. Star topology



- In a star topology all the cables run from the computers to a central location where they are all connected by a device called a hub.
- Each computer on a star network communicates with a central hub that resends the message either to all the computers in a broadcast star network or only to the destination computer in a switched star network.
- The hub in a broadcast star network can be active or passive. An active hub generates the electrical signal and sends it to all the computers connected to it.
- This type of hub is usually called a multiport repeater. Active hubs require external power supply.
- A passive hub is a wiring panel or punch down block which acts as a connection point. It does not amplify or regenerate the signal. Passive hubs do not require electrical power supply.

- Several types of cables can be used to implement a star network. A hybrid hub can use different types of cable in the same star network.
- A star network can be expanded by placing another star hub.
- This arrangement allows several more computers or hubs to be connected to that hub. This creates a hybrid star network.

Disadvantages of star topology:

- ✓ If the central hub fails, the whole network fails to operate.
- ✓ Many star networks require a device at the central point to rebroadcast or switch the network traffic.
- ✓ The cabling cost is more since cables must be pulled from

4. Mesh Topology

- In a mesh topology every device has a dedicated point-to-point link to every other device.
- The term dedicated means that the link carries traffic only between two devices it connects.
- A fully connected mesh network therefore has n(n-1)/2 physical channels to link n devices.

Advantages of mesh topology:

- ✓ The use of dedicated links guarantees that each connection can carry its own data load, thus eliminating traffic problems.
- ✓ A mesh topology is robust because the failure of any one computer does not bring down the entire network.
- ✓ Point to point links make fault diagnose easy.
- ✓ It provides security and privacy because every message sent travels along a dedicated line.

Disadvantages of mesh topology:

- ✓ Cabling cost is more.
- ✓ The hardware required to connect each link input/output and cable is expensive.
- ✓ Since every computer must be connected to every other computer installation and reconfiguration is difficult.

5. Tree topology

- ➢ A tree topology is a variation of a star. As in a star, nodes in a tree are linked to a central hub that controls the traffic to the network.
- ➢ However, not every computer plugs into the central hub, majority of them are connected to a secondary hub which in turn is connected to the central hub.
- ➢ The central hub in the tree is an active hub which contains repeater. The repeater amplifies the signal and increase the distance a signal can travel.
- ➢ The secondary hub may be active or passive. A passive hub provides a simple physical connection between the attached devices.

Advantages of tree topology:

- ✓ It allows more devices to be attached to a single hub and can therefore increase the distance of a signal can travel between devices.
- ✓ It allows the network to isolate and prioritise connections from different computers.

Disadvantages of tree topology:

- ✓ If the central hub fails the system breaks down.
- ✓ The cabling cost is more.

Comparison Chart

BASIS FOR COMPARISON	**STAR TOPOLOGY**	**RING TOPOLOGY**
Architecture structure	Peripheral nodes are linked to the central device known as a hub.	Every node has two branches connected to a node either side of it.
Amount of cabling required	Larger	Less as compared to star topology
Point of failure	Hub	Every node in the ring
Data traversal	All data passes through the central network connection.	Data moves in only one direction around the ring till it arrives the destination.
Network expansion	A new cable is plugged in from the new node to the hub.	In order to add a new node, a connection must be broken which turns down the network.
Fault isolation	Easy	Difficult
Troubleshooting	The other nodes are affected only in the case of a hub failure.	When a node goes down the information continues to transfer till the damaged node.
Cost	High	Low

Key Differences between Star and Ring Topology

1. In the star topology, each device is connected to a central node which sends the information received from one device to the other and act as a mediator. On the other hand, in the ring topology, each device has two nodes connected to either side of it, and the last node is connected to the first one.

2. The star topology requires more cable than ring topology.

3. Hub in the star topology is considered as a point of failure because the failure of any device would not affect the whole network, but if hub goes down, no data is transmitted across it. In contrast, each node in the ring topology is considered to be a point of failure as the failure of any device could significantly affect whole ring network.

4. In a star topology, all the data travels through the central hub. As against, in the ring topology, the data passes through each node unidirectionally until it reaches the destination.

5. To add new nodes to the ring network, a cable is used to connect the new device to the hub without influencing the rest of the network. On the contrary, the addition of new devices is done by breaking a connection which results in temporary unserviceable network till the new device is activated.

6. Fault isolation is easier in star topology while it is quite difficult in the ring topology.

7. Troubleshooting in the ring topology is simple, as the information continues to transfer through the rest of ring until reaching the point of failure. Conversely, in the star topology, the other devices are affected only when the connecting device goes down (Hub).

8. Star topology is expensive than the ring because it requires central connecting device usually hub.

<u>Types of area networks – LAN, MAN and WAN</u>

The **Network** allows computers to **connect and communicate** with different computers via any medium. LAN, MAN and WAN are the three major types of the network designed to operate over the area they cover. There are some similarities and dissimilarities between them. One of the major differences is the geographical area they cover, i.e. **LAN** covers the smallest area; **MAN** covers an area larger than LAN and **WAN** comprises the largest of all. There are other types of Computer Networks also, like :

- PAN (Personal Area Network)
- SAN (Storage Area Network)
- EPN (Enterprise Private Network)
- VPN (Virtual Private Network)

1) **Local Area Network (LAN) –**

LAN or Local Area Network connects network devices in such a way that personal computer and workstations can share data, tools and programs. The group of computers and devices are connected together by a switch, or stack of switches, using a private addressing scheme as defined by the TCP/IP protocol. Private addresses are unique in relation to other computers on the local network. Routers are found at the boundary of a LAN, connecting them to the larger WAN.

Data transmits at a very fast rate as the numbers of computers linked are limited. By definition, the connections must be high speed and relatively inexpensive hardware (Such as hubs, network adapters and Ethernet cables). LANs cover smaller geographical area (Size is limited to a few kilometers) and are privately owned. One can use it for an office building, home, hospital, schools, etc. LAN is easy to design and maintain. A Communication medium used for LAN has twisted pair cables and coaxial cables. It covers a short distance, and so the error and noise are minimized.

Early LAN's had data rates in the 4 to 16 Mbps range. Today, speeds are normally 100 or 1000 Mbps. Propagation delay is very short in a LAN. The smallest LAN may only use two computers, while larger LANs can accommodate thousands of computers. A LAN typically relies mostly on wired connections for increased speed and security, but wireless connections can also be part of a LAN. The fault tolerance of a LAN is more and there is less congestion in this network. For example : A bunch of students playing Counter Strike in the same room (without internet).

2) **Metropolitan Area Network (MAN) –**

MAN or Metropolitan area Network covers a larger area than that of a LAN and smaller area as compared to WAN. It connects two or more computers that are apart but resides in the same or different cities. It covers a large geographical area and may serve as an ISP (Internet Service Provider). MAN is designed for customers who need a high-speed connectivity. Speeds of MAN ranges in terms of Mbps. It's hard to design and maintain a Metropolitan Area Network.

The fault tolerance of a MAN is less and also there is more congestion in the network. It is costly and may or may not be owned by a single organization. The data transfer rate and the propagation delay of MAN is moderate. Devices used for transmission of data through MAN are:

Modem and Wire/Cable. Examples of a MAN are the part of the telephone company network that can provide a high-speed DSL line to the customer or the cable TV network in a city.

3) Wide Area Network (WAN) –

WAN or Wide Area Network is a computer network that extends over a large geographical area, although it might be confined within the bounds of a state or country. A WAN could be a connection of LAN connecting to other LAN's via telephone lines and radio waves and may be limited to an enterprise (a corporation or an organization) or accessible to the public. The technology is high speed and relatively expensive.

There are two types of WAN: **Switched WAN and Point-to-Point WAN**. WAN is difficult to design and maintain. Similar to a MAN, the fault tolerance of a WAN is less and there is more congestion in the network. A Communication medium used for WAN is PSTN or Satellite Link. Due to long distance transmission, the noise and error tend to be more in WAN.

WAN's data rate is slow about a 10th LAN's speed, since it involves increased distance and increased number of servers and terminals etc. Speeds of WAN ranges from few kilobits per second (Kbps) to megabits per second (Mbps). Propagation delay is one of the biggest problems faced here. Devices used for transmission of data through WAN are: Optic wires, Microwaves and Satellites. Example of a Switched WAN is the asynchronous transfer mode (ATM) network and Point-to-Point WAN is dial-up line that connects a home computer to the Internet.

Conclusion

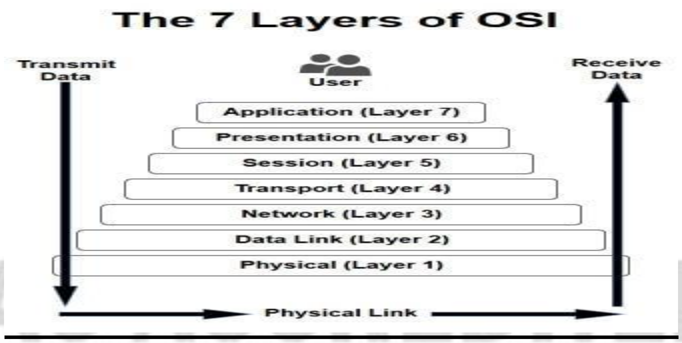
There are many advantages of LAN over MAN and WAN, such as LAN's provide excellent reliability, high data transmission rate, they can easily be managed, and shares peripheral devices too. Local Area Network cannot cover cities or towns and for that Metropolitan Area Network is needed, which can connect city or a group of cities together. Further, for connecting Country or a group of Countries one requires Wide Area Network.

<u>OSI MODEL</u>

- This model is based on a proposal developed by the international standards of organisation in 1983.
- It was revised in 1995.
- The model is called the ISO OSI (open systems interconnection) reference model because it deals with connecting open systems that is, systems that are open for communication with other systems.
- The OSI model has seven layers. The principles that were applied to arrive at the seven layers can be summarized as:

i. A layer should be created where a different abstraction is needed.
ii. Each layer should perform a well-defined function.
iii. The function of each layer should be choosen with an eye toward defining internationally standardized protocols.
iv. The layer boundaries should be choosen to minimize the information flow across the interfaces.
v. The number of layers should be large enough that distinct functions need not be thrown together in the same layer out of necessity and small enough that the architecture does not become unwieldy.

- The OSI model itself is not a network architecture because it does not specify the exact services and protocols to be used in each layer. It just tells what each layer should do. However, ISO has also produced standards for all the layers, although these are not part of the reference model itself. Each one has been published as a separate international standard.
- All the applications need not use all the seven layers. The lower three layers are enough for most of the applications. Each layer is built from electronic circuits and/or software and has a separate existence from the remaining layers.
- Each layer is supposed to handle messages or data from the layers which are immediately above or below it.
- This is done by following the protocol rules. Thus each layer takes data from the adjacent layer, handles it according to these rules and then passes the processed data to the next layer on the other side.



<u>LAYERS IN THE OSI MODEL</u>

1) **Physical layer:**

The physical layer coordinates the functions required to carry a bit stream over a physical medium. It deals with the mechanical and electrical specifications of the interface and transmission medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to occur.

The physical layer is responsible for movements of individual bits from one hop (node) to the next.

Other functions of physical layer are as follows:

a. **Physical characteristics of interfaces and medium:** The physical layer defines the characteristics of the interface between the devices and the transmission medium. It also defines the type of transmission medium.
b. **Representation of bits:** The physical layer data consists of a stream of bits (sequence of 0s or 1s) with no interpretation. To be transmitted, bits must be encoded into signals- electrical or optical. The physical layer defines the type of encoding (how 0s and 1s are changed to signals).
c. **Data rate:** The transmission rate- the number of bits sent each second-is also defined by the physical layer. In other words, the physical layer defines the duration of a bit, which is how long it lasts.
d. **Synchronization of bits:** The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level. In other words, the sender and the receiver clocks must be synchronized.
e. **Line configuration:** The physical layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.

f. **Physical topology:** The physical topology defines how devices are connected to make a network. Devices can be connected by using a mesh topology (every device is connected to every other device), a star topology (devices are connected through a central device), a ring topology (each device is connected to the next, forming a ring), a bus topology (every device is on a common link), or a hybrid topology (this is a combination of two or more topologies).

g. **Transmission mode:** The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex. In simplex mode, only one device can send; the other can only receive. The simplex mode is a one-way communication. In the half-duplex mode, two devices can send and receive, but not at the same time. In a full-duplex mode, two devices can send and receive at the same time.

2) Data link layer:

The data link layer transforms the physical layer, a raw transmission facility, to a reliable link. It makes the physical layer appear error-free to the upper layer (network layer).

The data link layer is responsible for moving frames from one hop (node) to the next.

Other responsibilities of the data link layer include the following:

a. **Framing:** the data link layer divides the stream of bits received from the network layer into manageable data units called frames.

b. **Physical Addressing:** If frames are to be distributed to different systems on the network, the data link layer adds a header to the frame to define the sender and/or receiver of the frame. If the frame is intended for a system outside the sender's network, the receiver address is the address of the device that connects the network to the next one.

c. **Flow control:** If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.

d. **Error control:** The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames. It also uses a mechanism to recognize duplicate frames. Error control is normally achieved through a trailer added to the end of the frame.

e. **Access control:** When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at a given time.

3) **Network layer:**

The network layer is responsible for the source to destination delivery of a packet, possibly across multiple network (links). Whereas the data link layer oversees the delivery of the packet between two systems on the same network (links), the network layer ensures that each packet gets from its point of origin to its final destination.

If two systems are connected to the same link, there is usually no need for a network layer. However, if the two systems are attached to different networks with connecting devices between the networks, there is often a need for the network layer to accomplish source-to-destination delivery.

The network layer is responsible for the delivery of individual packets from the source host to the destination host.

Other responsibilities of the network layer include:

a. **Logical addressing:** The physical addressing implemented by the data link layer handles the addressing problem locally. If a packet passes the network boundary, we need another addressing system to help distinguish the source and destination systems. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses of the sender and receiver.

b. **Routing:** When independent networks or links are connected to create internetworks (network of networks) or a large network, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism.

4) Transport layer:

The transport layer is responsible for process-to-process delivery of the entire message. A process is an application program running on a host. Whereas the network layer oversees source-to-destination delivery of individual packets, it does not recognize any relationship between those packets. It treats each one independently, as though each piece belonged to a separate message, whether or not it does.

The transport layer on the other hand, ensures that the whole message arrives intact and in order, overseeing both error control and flow control at the source-to-destination level.

The transport layer is responsible for the delivery of a message from one process to another.

Other responsibilities of the transport layer include the following:

a. **Service-point addressing:** Computers often run several programs at the same time. For this reason, source-to-destination delivery means delivery not only from one computer to the next but also from a specific process (running program) on one computer to a specific process on the other. The transport layer header must therefore include a type of address called a service-point address (or port address). The network layer gets each packet to the correct computer; the transport layer gets the entire message to the correct process on that computer.

b. **Segmentation and reassembly:** A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

c. **Connection control:** The transport layer can be either connectionless or connection-oriented. A connectionless transport layer treats each segment as an independent packet and delivers it to the transport layer at the destination machine. A connection-oriented transport layer makes a connection with the transport layer at the destination machine first before delivering the packets. After all the data are transferred, the connection is terminated.

d. **Flow control:** Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.

e. **Error control**: Error control at this layer is performed process-to-process rather than across a single link. The sending transport layer makes sure that the entire message arrives at the receiving transport layer without error (damage, loss or duplication). Error correction is usually achieved through retransmission.

5) Session layer:

The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes. The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems.

The session layer is responsible for dialog controller and synchronization.

Responsibilities of the session layer include the following:

a. **Dialog control:** The session layer allows two systems to enter into a dialog. It allows the communication between two processes to take place in either half-duplex (one way at a time) or full-duplex (two ways at a time) mode.

b. **Synchronization:** The session layer allows a process to add checkpoints, or synchronization points, to a stream of data. For example, if a system is sending a file of 2000 pages, it is advisable to insert checkpoints after every 100 pages to ensure that each 100-page unit is received and acknowledged independently. In this case, if a crash happens during the transmission of page 523, the only pages that need to be resent after system recovery are pages 501 to 523. Pages previous to 501 need not be resent.

6) Presentation layer:

The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems.

The presentation layer is responsible for translation, compression, and encryption.

Specific responsibilities of the presentation layer include the following:

a. **Translation:** The processes in two systems are usually exchanging information in the form of character strings, numbers and so on. The information must be changed to bit streams before being transmitted. Because different computers use different encoding systems, the presentation layer is responsible for interoperability between these different encoding methods. The presentation layer at the sender changes the information from its sender-dependent format into a common format. The presentation layer at the receiving machine changes the common format into its receiver-dependent format.

b. **Encryption:** To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverses the original process to transform the message back to its original form.

c. **Compression:** Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

7) Application layer:

The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, remote file access and transfer, shared database management, and other types of distributed information services.

The application layer is responsible for providing services to the user.

Specific responsibilities provided by the application layer include the following:

a. **Network virtual termination:** A network virtual terminal is a software version of a physical terminal, and it allows a user to log on to a remote host. To do so, the application creates a software emulation of a terminal at the remote host. The user's computer talks to the software terminal which, in turn, talks to the host, and vice versa. The remote host believes it is communicating with one of its own terminals and allows the user to log on.

b. **File transfer, access, and management:** This application allows a user to access files in a remote host (to make changes or read data), to retrieve files from a remote computer for use in the local computer, and to manage or control files in a remote computer locally.

c. **Mail services:** This application provides the basis for e-mail forwarding and storage.

d. **Directory services:** This application provides distributed database sources and access for global information about various objects and services.

Merits of OSI model:

- ✓ It distinguishes very clearly between the services, interfaces and protocols.
- ✓ The protocols in OSI model are better hidden. So they can be easily replaced by new protocols as the technology changes.
- ✓ OSI model is truly a general model.
- ✓ This model supports connection oriented as well as connectionless services.

Demerits of OSI model:

- ✓ Sessions and presentation layers are not of much use.
- ✓ This model was devised before the protocols were invented. So in real life there is a problem of fitting protocol into a model.