Semester-VI

B.Sc (Honours) in Physics

DSE 4: **Experimental Techniques**

**Lecture**

**on**

**Curve fitting**

***Discussed by Dr. K R Sahu***

**Lecture- Va**

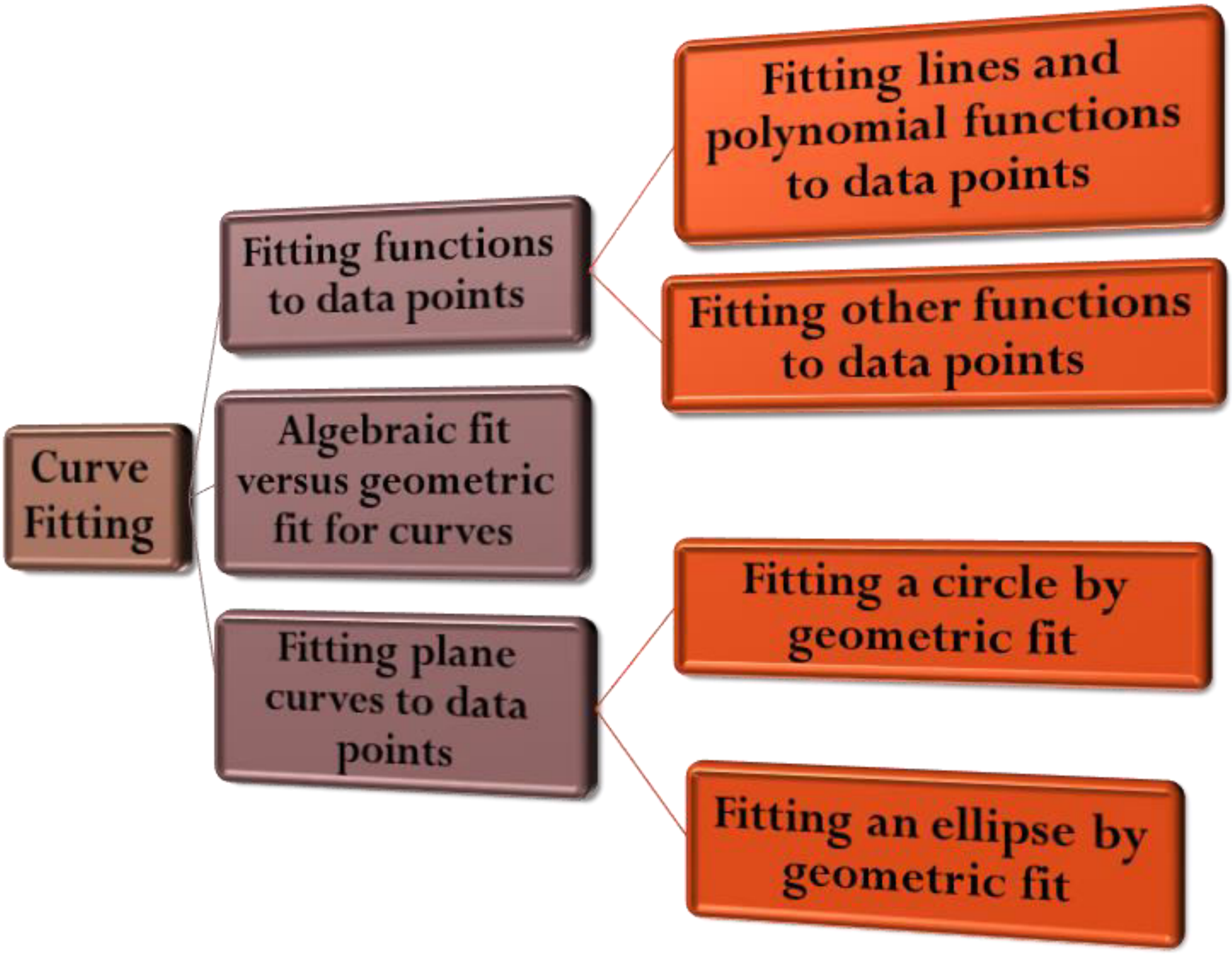
# Syllabus

- **Measurements**
  - Accuracy and precision and Significant figures.
  - Error and uncertainty analysis.
  - Types of errors:
    - Gross error,
    - Systematic error,
    - Random error.
  - Statistical analysis of data
    - Arithmetic mean,
    - Deviation from mean,
    - Average deviation,
    - Standard deviation,
    - Chi-square and
  - **Curve fitting.**
  - Guassian distribution.

# Curve fitting

**Curve fitting** is the process of constructing a curve, or mathematical function, that has the ***best fit to a series of data points***, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "**smooth**" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data.

# Types of Curve fitting



# Fitting functions to data points

Most commonly, one fits a function of the form $y = f(x)$.

**Fitting lines and polynomial functions to data points**

The first degree polynomial equation

**$y = ax+b$**

is a line with slope $a$. A line will connect any two points, so a first degree polynomial equation is an exact fit through any two points with distinct x coordinates.

If the order of the equation is increased to a second degree polynomial, the following results:

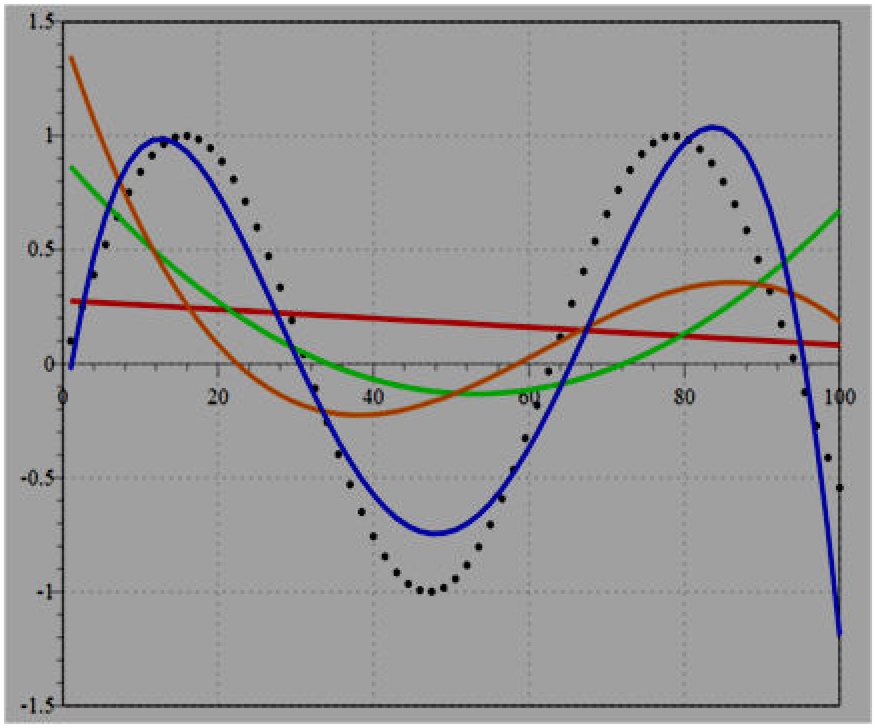
**$y = ax^2+bx+c$**

This will exactly fit a simple curve to three points.

If the order of the equation is increased to a third degree polynomial, the following is obtained:

**$y= ax^3+bx^2+cx+d$**

This will exactly fit four points.




Polynomial curves fitting points generated with a sine function. **Red line** is a **first degree polynomial**, **green line** is **second degree**, **orange line** is **third degree** and **blue** is **fourth degree**

A more general statement would be to say it will exactly fit four constraints. Each constraint can be a point, angle, or curvature (which is the reciprocal of the radius of an osculating circle). Angle and curvature constraints are most often added to the ends of a curve, and in such cases are called **end conditions**. Identical end conditions are frequently used to ensure a smooth transition between polynomial curves contained within a **single spline**. Higher-order constraints, such as "the change in the rate of curvature", could also be added. This, for example, would be useful in highway cloverleaf design to understand the rate of change of the forces applied to a car, as it follows the cloverleaf, and to set reasonable speed limits, accordingly.

The first degree polynomial equation could also be an exact fit for a single point and an angle while the third degree polynomial equation could also be an exact fit for two points, an angle constraint, and a curvature constraint. Many other combinations of constraints are possible for these and for higher order polynomial equations.

If there are more than $n + 1$ constraints (***n* being the degree of the polynomial**), the polynomial curve can still be run through those constraints. An exact fit to all constraints is not certain (but might happen, for example, in the case of a first degree polynomial exactly fitting three **collinear points**). In general, however, some method is then needed to evaluate each approximation. The **least squares** method is one way to compare the deviations.

The degree of the polynomial curve being higher than needed for an exact fit is undesirable for all the reasons listed previously for high order polynomials, but also leads to a case where there are an infinite number of solutions. For example, a first degree polynomial (a line) constrained by only a single point, instead of the usual two, would give an infinite number of solutions. This brings up the problem of how to compare and choose just one solution, which can be a problem for software and for humans, as well. For this reason, it is usually best to choose as low a degree as possible for an exact match on all constraints, and perhaps an even lower degree, if an approximate fit is acceptable.

**Fitting other functions to data points**

Other types of curves, such as **trigonometric functions** (such as sine and cosine), may also be used, in certain cases.

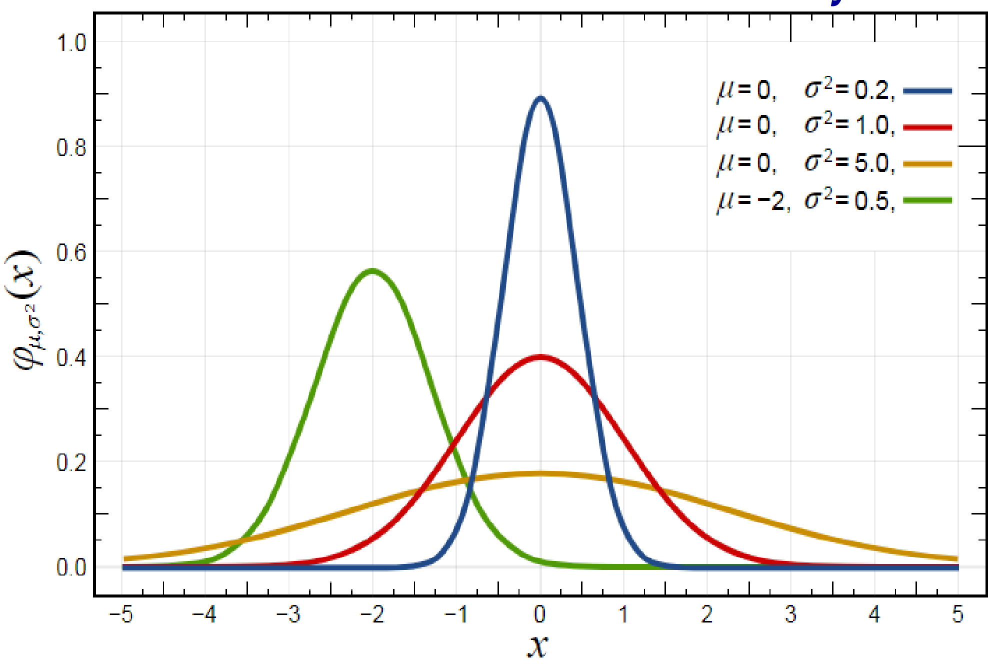
In spectroscopy, data may be fitted with **Gaussian, Lorentzian, Voigt** and related functions.

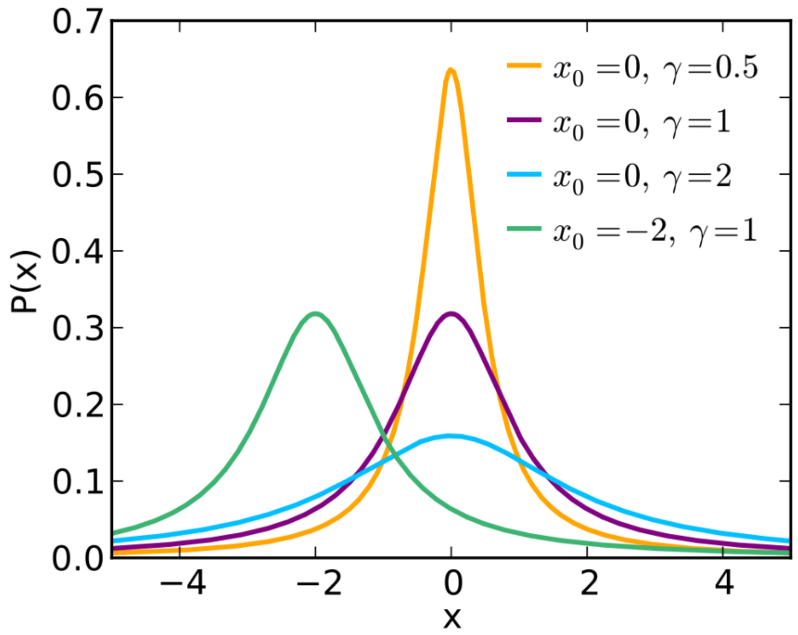
In **agriculture** the inverted logistic **sigmoid function** (S-curve) is used to describe the relation between crop yield and growth factors. The blue figure was made by a sigmoid regression of data measured in farm lands. It can be seen that initially, i.e. at low soil salinity, the crop yield reduces slowly at increasing soil salinity, while thereafter the decrease progresses faster.

**Gaussian distribution**

## Probability density function

**Lorentz distribution**



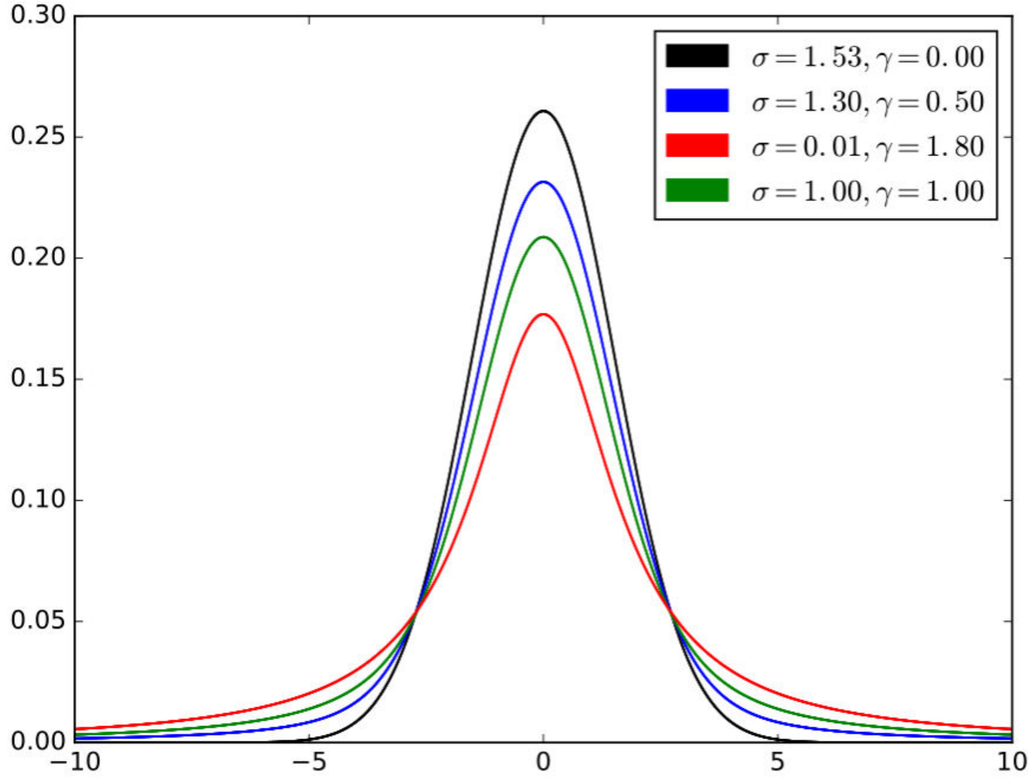




In probability theory, a normal (or **Gaussian** or Gauss or Laplace–Gauss) **distribution** is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The parameter **μ** is the **mean** or **expectation** of the distribution (and also its **median** and **mode**); and **σ** is its **standard deviation**. The **variance** of the distribution is . A random variable with a Gaussian distribution is said to be **normally distributed** and is called a **normal deviate**.

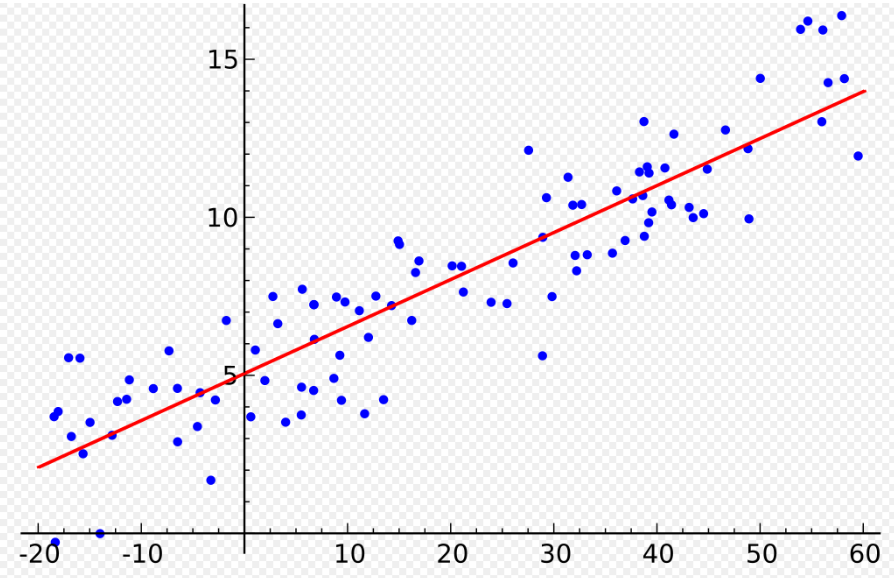
The **Cauchy distribution**, is a continuous probability distribution. It is also known, especially among physicists, as the **Lorentz distribution**, Cauchy–Lorentz distribution, Lorentz(ian) function, or Breit–Wigner distribution. The Cauchy distribution $f(x;x_0,\gamma)$ is the distribution of the x-intercept of a ray issuing from $(x_0,\gamma)$ with a uniformly distributed angle. It is also the distribution of the ratio of two independent normally distributed random variables with mean zero.

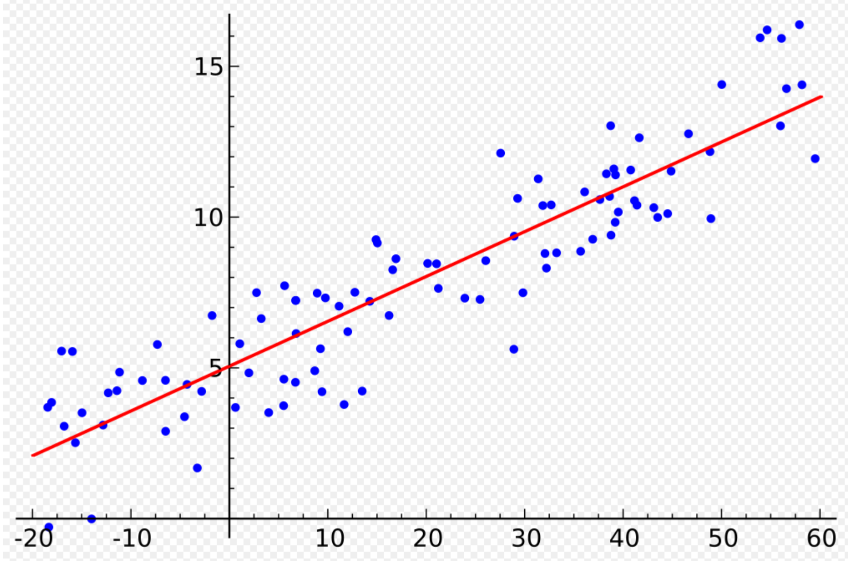



The **Voigt profile** is a probability distribution given by a convolution of a Cauchy-Lorentz distribution and a Gaussian distribution. It is often used in analyzing data from spectroscopy or diffraction.

**Algebraic fit versus geometric fit for curves**

For algebraic analysis of data, "fitting" usually means trying to find the curve that minimizes the vertical (*y*-axis) displacement of a point from the curve (e.g., **ordinary least squares**). However, for graphical and image applications geometric fitting seeks to provide the best visual fit; which usually means trying to minimize the orthogonal distance to the curve (e.g., **total least squares**), or to otherwise include both axes of displacement of a point from the curve. Geometric fits are not popular because they usually require non-linear and/or iterative calculations, although they have the advantage of a more aesthetic and geometrically accurate result.




In statistics, **ordinary least squares** (OLS) is a type of **linear least squares method** for estimating the **unknown parameters** in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function.

In applied statistics, **total least squares** is a type of **errors-in-variables regression**, a least squares data modeling technique in which **observational errors** on both **dependent and independent** variables are taken into account. It is a generalization of Deming regression and also of orthogonal regression, and can be applied to both **linear** and **non-linear** models.

The total least squares approximation of the data is generically equivalent to the best, in the Frobenius norm, low-rank approximation of the data matrix.

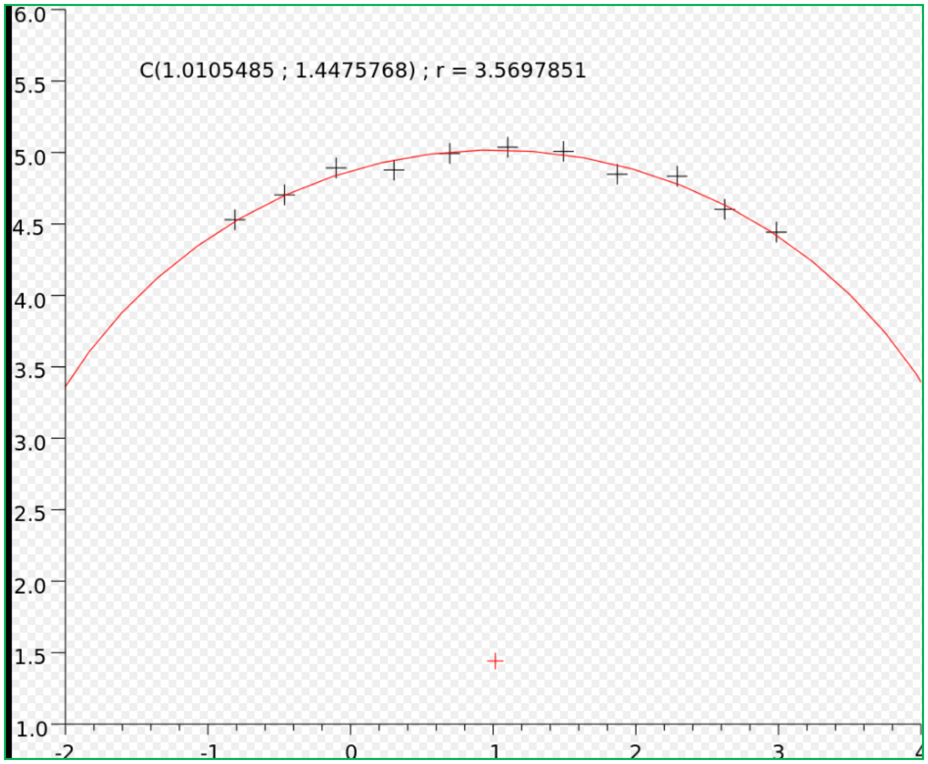
## Fitting plane curves to data points

If a function of the form y=f(x) cannot be postulated, one can still try to fit a **plane curve**. Other types of curves, such as **conic sections** (circular, elliptical, parabolic, and hyperbolic arcs) or **trigonometric functions** (such as sine and cosine), may also be used, in certain cases. For example, trajectories of objects under the influence of gravity follow a parabolic path, when air resistance is ignored. Hence, matching trajectory data points to a parabolic curve would make sense. Tides follow sinusoidal patterns, hence tidal data points should be matched to a sine wave, or the sum of two sine waves of different periods, if the effects of the Moon and Sun are both considered.

For a **parametric curve**, it is effective to fit each of its coordinates as a separate function of **arc length**; assuming that data points can be ordered, the **chord distance** may be used.

### Fitting a circle by geometric fit

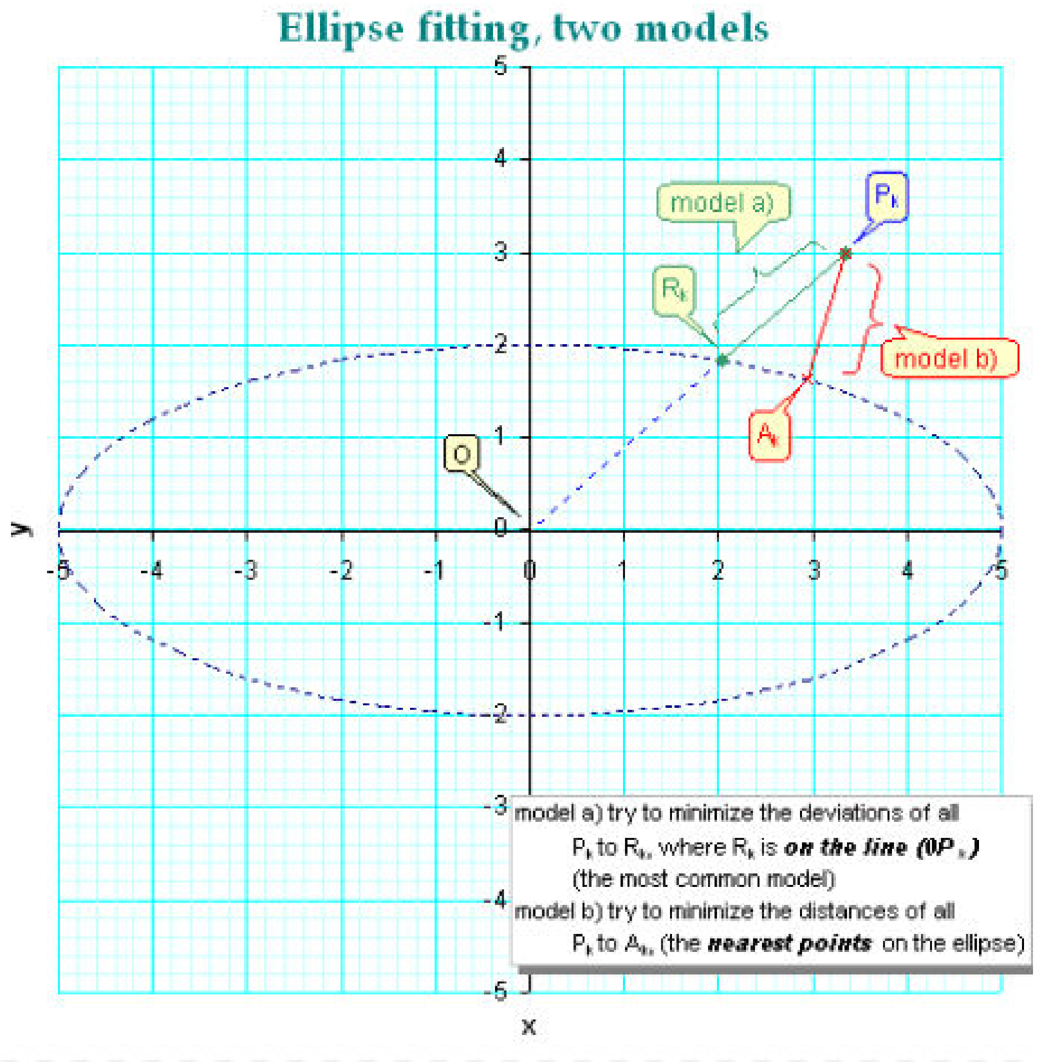
Coope approaches the problem of trying to find the best visual fit of circle to a set of 2D data points. The method elegantly transforms the ordinarily non-linear problem into a linear problem that can be solved without using iterative numerical methods, and is hence much faster than previous techniques.




Circle fitting with the Coope method, the points describing a circle arc, centre (1 ; 1), radius 4.

**Fitting an ellipse by geometric fit**

The above technique is extended to general ellipses by adding a non-linear step, resulting in a method that is fast, yet finds visually pleasing ellipses of arbitrary orientation and displacement.

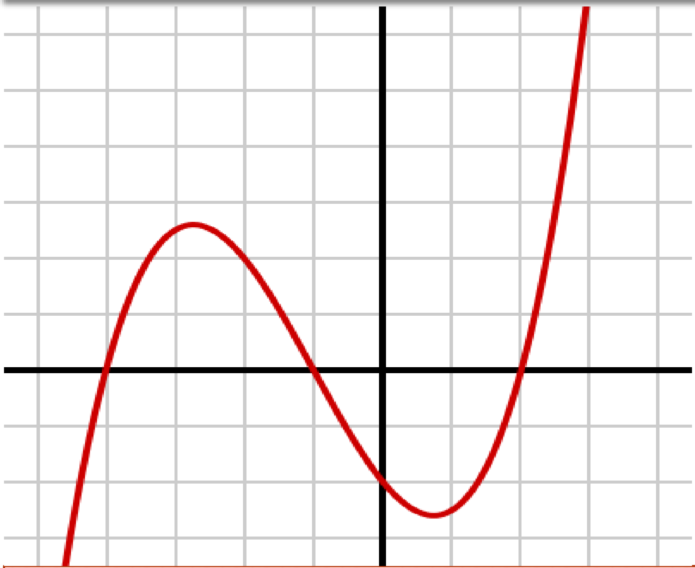



**Mathematical Calculation and with Examples has been discuss in the next lecture (Lecture-Vb)**

# Appendix-1

**Polynomial**

In mathematics, a ***polynomial*** is an expression consisting of variables (also called indeterminates) and coefficients, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponents of variables. An example of a polynomial of a single indeterminate, x, is $\mathbf{x^2 - 4x + 7}$. An example in three variables is $\mathbf{x^3 + 2xyz^2 - yz + 1}$.



**The graph of a polynomial function of degree 3,** with roots at 2, -1, and -4

$$\mathbf{Y = \tfrac{1}{4}\,(x+4)(x+1)(x-2)}$$
$$\mathbf{= x^3/4 + 3x^2/4 - 3x/2 - 2}$$

A polynomial in a single indeterminate *x* can always be written (or rewritten) in the form: $\mathbf{a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0}$ **;** where $\mathbf{a_0, a_1, a_2, a_3, \ldots, a_n}$ are constants and is the indeterminate. The word "indeterminate" means that represents no particular value, although any value may be substituted for it.

This can be expressed more concisely by using summation notation:

$$\sum_{k=0}^{n} a_k x^k$$

That is, a **polynomial** can either be **zero** or can be written as the **sum** of a **finite number of non-zero terms.** Each term consists of the product of a number – called the **coefficient** of the term – and a finite number of indeterminates, raised to nonnegative integer powers.

Polynomials appear in many areas of mathematics and science. For example, they are used to form **polynomial equations (form: P(x) = 0)**, which encode a wide range of problems, from elementary word problems to complicated scientific problems; they are used to define polynomial functions, which appear in settings ranging from basic chemistry and physics to economics and social science; they are used in calculus and numerical analysis to approximate other functions. In advanced mathematics, polynomials are used to construct polynomial rings and algebraic varieties, central concepts in algebra and algebraic geometry.

**Documents has been collected from**

https://en.wikipedia.org/wiki/Curve_fitting