Basic Math Refresher

William Stallings

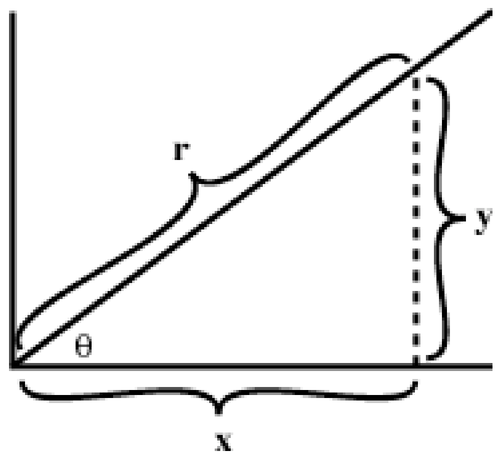
Here are some basic math concepts from high school that you might find helpful in reading some of my books and in doing some of the homework problems. If you have any suggestions for additions to this document, please contact me at ws@shore.net.

This document available at WilliamStallings.com/StudentSupport.html

Last updated: 10 March, 2005

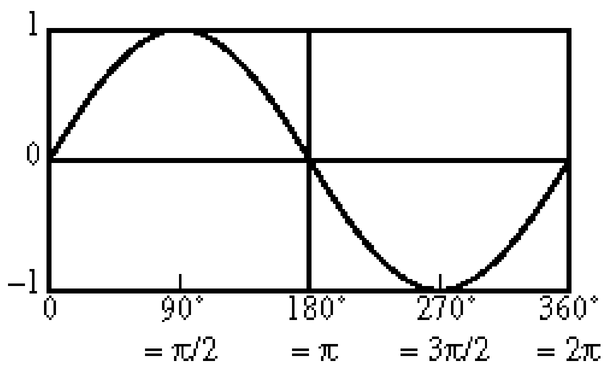
Trigonometric Identities

First, we define the trigonometric functions using this diagram:

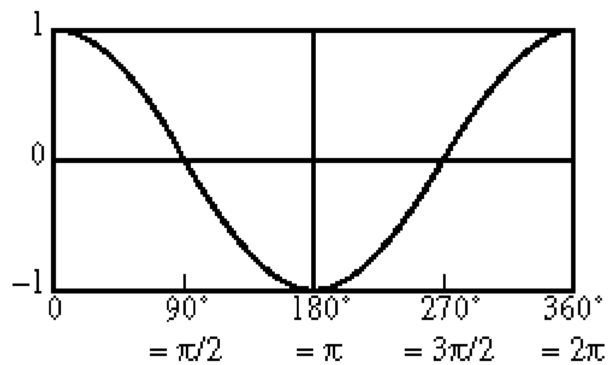


$\sin\theta = y/r$	$\csc\theta = r/y$
$\cos\theta = x/r$	$\sec\theta = r/x$
$\tan\theta = y/x$	$\text{ctn}\,\theta = x/y$

The sine and cosine graphs are as follows:



sin θ



cos θ

Angles may be expressed in degrees or radians, where $180^\circ = \pi$ radians.

$\sin(\theta) = \cos(\theta - 90^\circ) = \cos(\theta - \pi/2)$ $\quad$ $\cos(\theta) = \sin(\theta + 90^\circ) = \sin(\theta + \pi/2)$

$\sin(\theta) = \sin(\theta + 2n\pi)$ $\quad$ $\cos(\theta) = \cos(\theta + 2n\pi)$ $\quad$ n = any integer

Derivatives: $\frac{d}{dt}\sin(at) = a\cos(at) \quad \frac{d}{dt}\cos(at) = -a\sin(at)$

Reciprocal identities: $\sin\theta = \frac{1}{\csc\theta} \quad \cos\theta = \frac{1}{\sec\theta} \quad \tan\theta = \frac{1}{\cot\theta}$

Pythagorean identities: $\sin^2\theta + \cos^2\theta = 1$; $1 + \tan^2\theta = \sec^2\theta$; $1 + \cot^2\theta = \csc^2\theta$

Quotient identities: $\tan\theta = \frac{\sin\theta}{\cos\theta} \quad \cot\theta = \frac{\cos\theta}{\sin\theta}$

Even-odd identities: $\sin(-\theta) = -\sin(\theta)$; $\cos(-\theta) = \cos(\theta)$; $\tan(-\theta) = -\tan(\theta)$

$\csc(-\theta) = -\csc(\theta)$; $\sec(-\theta) = \sec(\theta)$; $\cot(-\theta) = -\cot(\theta)$

Sum-difference formulas: $\sin(\theta \pm \beta) = \sin\theta\cos\beta \pm \cos\theta\sin\beta$

$\cos(\theta \pm \beta) = \cos\theta\cos\beta \pm \sin\theta\sin\beta$

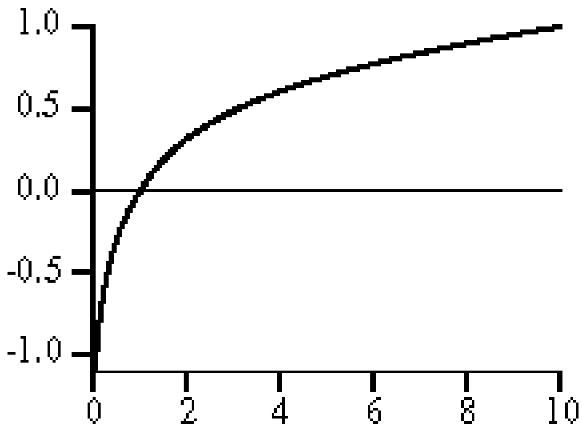
Double angle formulas: $\sin(2\theta) = 2\sin\theta\cos\theta$

$\cos(2\theta) = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$

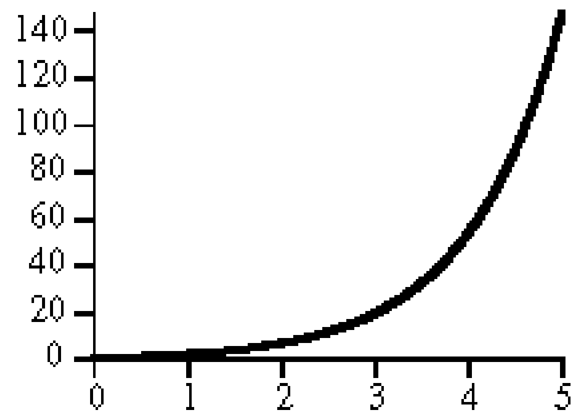
Product-to-sum formulas: $\sin\theta\sin\beta = (1/2)\,[\cos(\theta - \beta) - \cos(\theta + \beta)]$

$\cos\theta\cos\beta = (1/2)\,[\cos(\theta - \beta) + \cos(\theta + \beta)]$

Logarithms and Exponentials



$\log_{10}(x)$



10^x

The logarithm of x to the base a is written as $\log_a(x)$ and is equal to the power to which a must be raised in order to equal x:

$$\text{If } x = a^y, \text{ then } y = \log_a(x)$$

$\log(XY) = (\log X) + (\log Y)$

$\log(X/Y) = (\log X) - (\log Y)$

$\log(X^p) = p \log X$

$\log_{10} x = (\log_2 x)/(\log_2 10)$

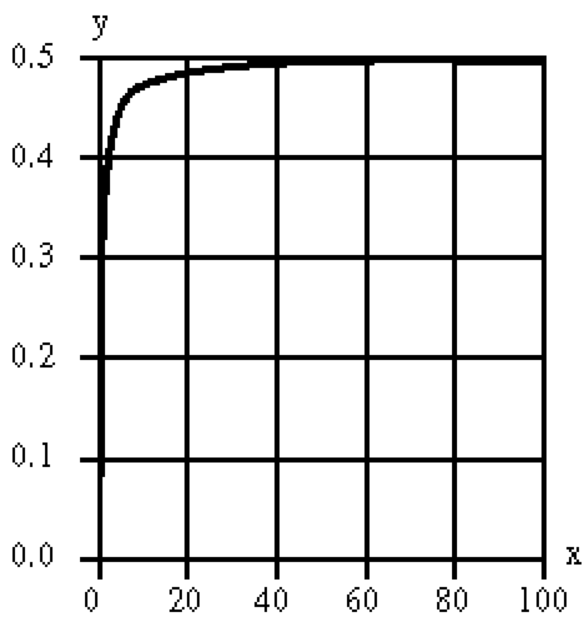
$\log_2 x = (\log_{10} x)/(\log_{10} 2)$

$b^X b^Y = b^{(X+Y)}$

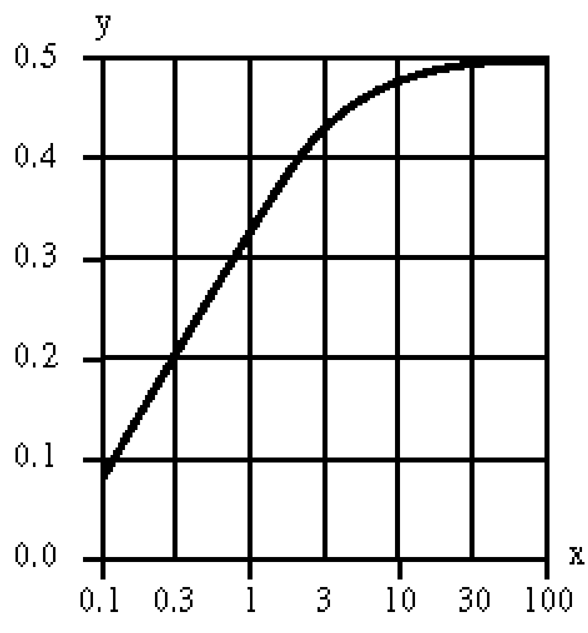
$b^X/b^Y = b^{(X-Y)}$

Log Scales

It is sometimes convenient to use what is called a log scale rather than a linear scale on either the x or y coordinate of a graph, or both. This is often useful if the range of values along the coordinate is several orders of magnitude (one order of magnitude = one factor of ten). Below are two plots of the same function y = x/(2x + 1). On the left-hand graph, both the x and y axes use a linear scale. This means, for example, that the distance on the x axis from 0 to 20 is the same as the distance from 20 to 40, and so on. The y axis is also linear: the distance from 0.0 to 0.1 is the same as from 0.1 to 0.2, and so on. It is hard to read values off this graph because the graph changes very rapidly for small values of x and very slowly for large values of x. What is needed is some way to "blow up" the part of the graph for small values of x without making the overall graph unmanageably big. The way to do this is to use a log scale on the x axis, in which distances are proportional to the log of the values on the x axis. This is shown in the right-hand graph. On the x axis, the distance from 10^{-1} to 10^{0} is the same as from 10^{0} to 10^{1}, which is the same as the distance from 10^{1} to 10^{2}. Now there is more emphasis on smaller values of x and it is easier to read useful values of y.



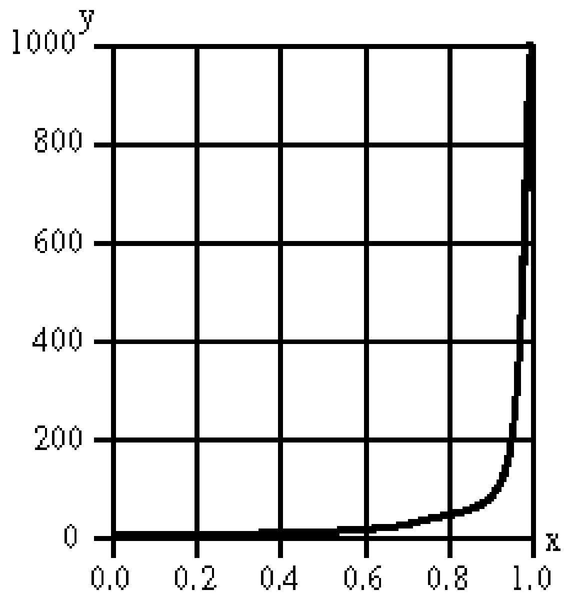
y = x/(2x + 1)



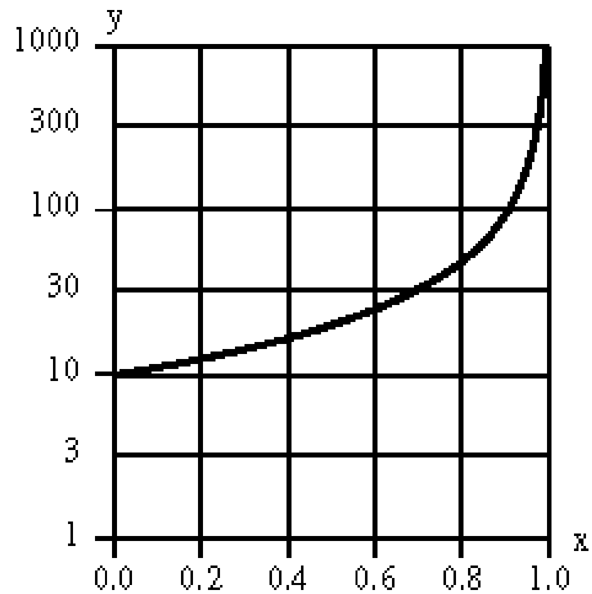
y = x/(2x + 1)

Note that you must be careful in interpolating between tick marks. For example, on a log scale, about half-way between 0.1 and 1 is 0.3, not 0.5; about half-way between 1 and 10 is 3, not 5; and so on.

The same technique can be applied to the y axis, as shown in the two graphs below. This function has y values that cover a wide range. In the left-hand graph, it is difficult to pick the y values out. In the right-hand graph, a log scale is used on the y axis and it is much easier to read the y values.



$y = (20 - x)/(2 - 2x)$



$y = (20 - x)/(2 - 2x)$

Vectors, Matrices, and Determinants

We use the following conventions:

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix} \qquad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{11} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{n2} & \cdots & a_{mn} \end{pmatrix}$$

row vector **X** column vector **Y** matrix **A**

Note that in a matrix, the first subscript of an element refers to the row and the second subscript refers to the column.

Arithmetic

Two matrices of the same dimensions can be added or subtracted element by element. Thus, for **C** = **A** + **B**, the elements of **C** are $c_{ij} = a_{ij} + b_{ij}$. To multiply a matrix by a scalar, every element of the matrix is multiplied by the scalar. Thus, for **C** = k**A**, we have $c_{ij} = k \times a_{ij}$. The product of a row vector of dimension m and a column vector of dimension m is a scalar:

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = x_1y_1 + x_2y_2 + \ldots + x_my_m$$

Two matrices **A** and **B** are conformable for multiplication, in that order, if the number of columns in **A** is the same as the number of rows in **B**. Let **A** be of order $m \times n$ (m rows and n columns) and **B** be of order $n \times p$. The product is obtained by multiply every row of **A** into every column of **B**, using the rules just defined for the product of a row vector and a column vector. Thus, for **C** = **AB**, we have $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, and the resulting matrix is of order $m \times p$. Notice that, by these rules, we can multiply a row vector by a matrix that has the same number of rows as the dimension of the vector; and we can multiply a matrix by a column vector if the matrix has the same number of columns as the dimension of the vector. Thus, using the notation at the beginning of this section: For **D** = **XA**, we end up with a row vector with elements $d_i = \sum_{k=1}^{m} x_k a_{ki}$. For **E** = **AY**, we end up with a column vector with elements $e_i = \sum_{k=1}^{m} a_{ik} y_k$.

Determinants

The determinant of the square matrix **A**, denoted by det(**A**), is a scalar value representing sums and products of the elements of the matrix. For details, see any text on linear algebra. Here, we simple report the results.

For a 2×2 matrix **A**, $\det(\mathbf{A}) = a_{11}a_{22} - a_{21}a_{12}$.

For a 3×3 matrix **A**, $\det(\mathbf{A}) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}$

In general, the determinant of a square matrix can be calculated in terms of its cofactors. A **cofactor** of **A** is denoted by $\mathrm{cof}_{ij}(\mathbf{A})$ and is defined as the determinant of the reduced matrix formed by deleting the *i*th row and *j*th column of **A** and choosing positive sign if $i + j$ is even and the negative sign if $i + j$ is odd. For example:

$$\mathrm{cof}_{23}\begin{pmatrix} 2 & 4 & 3 \\ 6 & 1 & 5 \\ -2 & 1 & 3 \end{pmatrix} = -\det\begin{pmatrix} 2 & 4 \\ -2 & 1 \end{pmatrix} = -10$$

The determinant of an arbitrary $n \times n$ square matrix can be evaluated as:

$$\det(\mathbf{A}) = \sum_{j=1}^{n}\left[a_{ij}\mathrm{cof}_{ij}(\mathbf{A})\right] \quad \text{for any } i$$

or

$$\det(\mathbf{A}) = \sum_{i=1}^{n}\left[a_{ij}\mathrm{cof}_{ij}(\mathbf{A})\right] \quad \text{for any } j$$

For example:

$$\det\begin{pmatrix} 2 & 4 & 3 \\ 6 & 1 & 5 \\ -2 & 1 & 3 \end{pmatrix} = a_{21}\mathrm{cof}_{21} + a_{22}\mathrm{cof}_{22} + a_{23}\mathrm{cof}_{23}$$

$$= 6(-9) + 1(12) + 5(-10) = -92$$

Inverse of a Matrix

If a matrix $\mathbf{A}$ has a nonzero determinant, then it has an inverse, denoted as $\mathbf{A}^{-1}$. The inverse has that property that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the matrix that is all zeros except for ones along the main diagonal from upper left to lower right. $\mathbf{I}$ is known as the identity matrix because any vector or matrix multiplied by $\mathbf{I}$ results in the original vector or matrix. The inverse of a matrix is calculated as follows. For $\mathbf{B} = \mathbf{A}^{-1}$,

$$b_{ij} = \frac{\mathrm{cof}_{ji}(\mathbf{A})}{\det(\mathbf{A})}$$

Some Basic Probability Concepts for Discrete Random Variables

Computations in probability are based on the concept of a random variable. In essence, a random variable reflects some quantity in the real world that can take on a number of different values with different probabilities. In the jargon of probability textbooks, a random variable is a mapping from the set of all possible events under consideration to the real numbers. That is, a random variable associates a real number with each event. This concept is sometimes expressed in terms of an experiment with many possible outcomes; a random variable assigns a value to each such outcome. A random variable is **continuous** if it takes on a noncountably infinite number of distinct values. A random variable is **discrete** if it takes on a finite or countably infinite number of values. This note is concerned with discrete random variables.

A discrete random variable *X* can be described by its **distribution function** $P_X(k)$:

$$\text{(Probability that X = k)} = P_X(k) = \Pr[X = k] \qquad \sum_{\text{all } k} P_X(k) = 1$$

We are often concerned with some characteristic of a random variable rather than the entire distribution, such as the mean value, also called expected value:

$$E[X] = \mu_X = \sum_{\text{all } k} k \Pr[X = k]$$

Other useful measures:

Second moment:
$$E\left[X^2\right] = \sum_{\text{all } k} k^2 \Pr[X = k]$$

Variance: $\mathrm{Var}[X] = E\left[(X - \mu_X)^2\right] = E\left[X^2\right] - \mu_X^2$

Standard deviation $\sigma_X = \sqrt{\mathrm{Var}[X]}$

The variance and standard deviation are measures of the dispersion of values around the mean. It is easy to show that for a constant *a*:

$$E[aX] = aE[X]; \qquad Var[aX] = a^2 Var[X]$$

The mean is known as a first-order statistic; the second moment and variance are second-order statistics. Higher-order statistics can also be derived from the probability density function.

For any two random variables *X* and *Y*, we have:

$$E[X + Y] = E[X] + E[Y]$$