An Introduction to the Mathematics of Digital Signal Processing

Part I: Algebra, Trigonometry, and The Most Beautiful Formula in Mathematics

F. R. Moore
Bell Laboratories
Murray Hill, New Jersey 07974

Introduction

As it says in the front of the *Computer Music Journal* number 4, there are many musicians with an interest in musical signal processing with computers, but only a few have much competence in this area. There is of course a huge amount of literature in the field of digital signal processing, including some first-rate textbooks (such as Rabiner and Gold's *Theory and Application of Digital Signal Processing*, or Oppenheim and Schafer's *Digital Signal Processing*), but most of the literature assumes that the reader is a graduate student in engineering or computer science (why *else* would he be interested?), that he wants to know *everything* about digital signal processing, and that he already knows a great deal about mathematics and computers. Consequently, much of this information is shrouded in mathematical mystery to the musical reader, making it difficult to distinguish the wheat from the chaff, so to speak. Digital signal processing *is* a very mathematical subject, so to make past articles clearer and future articles possible, the basic mathematical ideas needed are presented in this two-part tutorial. In order to prevent this presentation from turning into several fat books, only the main ideas can be outlined; and mathematical proofs are of course omitted. But keep in mind that learning mathematics is much like learning to play a piano: no amount of reading will suffice—it is necessary to actually practice the techniques described (in this case, by doing the problems) before the concepts become useful in the "real" world. Therefore some problems are provided (without answers) to give the motivated reader an opportunity both to test his understanding and to acquire some skill.

Part I of the tutorial (this part) provides a general review of algebra and trigonometry, including such areas as equations, graphs, polynomials, logarithms, complex numbers, infinite series, radian measures, and the basic trigonometric functions. Part II will discuss the application of these concepts and others in transforms, such as the Fourier and z-transforms, transfer functions, impulse response, convolution, poles and zeroes, and elementary filtering. Insofar as possible, the mathematical treatment always stops just short of using calculus, though a deep understanding of many of the concepts presented requires understanding of calculus. But digital signal processing inherently requires less calculus than analog signal processing, since the integral signs are replaced by the easier-to-understand discrete summations. It is an experimental goal of this tutorial to see how far into digital signal processing it is possible to explore *without* calculus.

Algebra

To most people, mathematics *means* formulas and equations, which are expressions describing the relationships among quantities. As long as the relationships do not use the integration or differentiation ideas of calculus, they usually fall into the general domain of algebra, named after the arabic best-seller of the 9th century, *Kitab al jabr w'al-muqabala* ("Rules of Restoration and Reduction") by Abu Ja'far Mohammed ibn Mûsâ al-Khowârizmi (from whose name the word *algorithm* is derived).

Algebra is, in fact, merely a systematic notation of quantitative relationships among numerical quantities, usually called variables, since with algebra we can manipulate the

relationships into various forms without specifying the particular quantities we are manipulating. For example, the equation:

$$y = x+1$$

"says" that y is an arbitrary name given to a quantity which is one greater than another quantity, x. If we were to write

$$y - 1 = x$$

we would be "saying" exactly the same thing, just as we would if we wrote any of the following:

$$\begin{aligned} 16y &= 16 + 16x \\ y/2 &= \tfrac{1}{2}(x + 1) \\ \pi(y - \pi) &= \pi(1 - \pi) + \pi x \end{aligned}$$

The basic notion here is that whatever is on the left hand side of the equal sign (=) is just *another name* for what is on the right hand side. Of course, as the last example above shows, there are simple ways and complicated ways to say the same thing, and it is usually the task of the algebraist to find the simplest way of expressing a relationship so that it can be easily understood.

Functions, Numbers, and Graphs

Sometimes it is desirable to give a name to an entire relationship, rather than just to the variables in a relationship. Mathematicians have a keen sense of brevity, so these names are usually single letters as well, but they serve quite a different purpose. For example, the notation

$$f(x) = x + 1$$

means that "f" is being defined as a *function of* x, where x is called the *independent variable*, since it can take on any value whatsoever. We can now write

$$y = f(x)$$
(read: "y equals f of x")

to mean that the value of y (which is called a *dependent variable* since its value depends on the value chosen for x) is a function of x, and the function is named f. Remember that $f(x)$ is just another name for $x + 1$, so the last equation above is still saying the same thing as all of the previous examples. The advantages of the function notation are that it a) explicitly states the name of the varying quantity (the independent variable or *argument* of the function), and b) it gives a short name to what may be a complicated expression, allowing its further manipulation. For example:

let $f(x) = x + 1$ (as above), and
$g(x) = 2x + 3$

We might now define:

$$a = f(x) + g(x) \quad \text{and}$$
$$b = f(x) - g(x)$$

Of course, this "says" the same thing as

$$a = 3x + 4 \quad \text{and}$$
$$b = -x - 2$$

but the latter form doesn't show explicitly where these relationships come from.

What do we mean when we say that x can have *any* value? In fact, what does value mean? Without going too far afield into the theory of numbers, we should note that in many cases, the value of the independent variable in a particular function is restricted to the set of all natural numbers, or integers, or reals. Briefly, the set of natural numbers (denoted here as $\mathbf{N}$) is the set of numbers used for counting:

$$\mathbf{N} = \{ 0, 1, 2, 3 \ldots \}$$

(the curly braces "{ }" denote a set, and the ellipsis "..." means here that the set has an infinite number of elements). To indicate that the independent variable must be chosen from this set, we write

$$f(x) = x - 1 \qquad x \in \mathbf{N}$$

where "$\in \mathbf{N}$" means "is an element of $\mathbf{N}$", the set of all natural numbers. Suppose we choose x equal to 0; what is $f(x)$ equal to? Our Pavlovian response is, of course, minus one, but note that this number is *not* a natural number as defined above.

So even though x might always be a natural number, $f(x)$ might not be. Other sets of numbers frequently encountered are $\mathbf{I}$, the set of all integer numbers,

$$\mathbf{I} = \{ 0, \pm 1, \pm 2, \pm 3, \ldots \}$$

and $\mathbf{R}$, the set of all real numbers. Real numbers are those which can be written as a (possibly unending) decimal expression, such as π, 2, and 1/3, since π = 3.14159. . ., 2 = 2.000. . ., and 1/3 = .333. . . . Sometimes $\mathbf{R}^+$ is used to denote the positive reals, $\mathbf{R}^2$ for the set of all *ordered pairs* of real numbers, etc. Just as the integers include all of the natural numbers, the reals include the integers, as well as the rationals (numbers formed by the ratio of two integers, such as 1/3 or 22/7), and the irrationals, like π (which is *approximately* equal to 22/7, but is not exactly equal to *any* ratio of two integers). It is a fundamental mystery that the ratio of the diameter of a circle to its circumference should so transcend our ability to compute it exactly on any number of fingers, but that's just the way our particular universe is arranged! π and e are also called trancendental numbers for such metaphysical reasons (more about e later).

So if we are permitted to use the integers, we can completely solve $f(x) = x - 1, x \in \mathbf{N}$ for all allowed values of x. It is clear that the equation

$$3x = 2 \qquad x \in \mathbf{I}$$

has *no* solution, since no integer has the value 2/3. There is another type of number needed to solve such equations as $x^2 + 1 = 0$, since no real number when multiplied by itself is equal to -1. Mathematicians simply *define* the square root of minus one as i, the imaginary unit. (Engineers use j, since i was already used to stand for current in the engineering

literature. In Part I of this tutorial we shall stick with i; Part II will use j, since signal processing is a branch of engineering.) An imaginary number is any real number times i, and since the reals include the other number sets, we can have imaginary integers, imaginary rationals, even imaginary naturals!

The final set of numbers is just a combination of the reals with the imaginaries, which are called complex numbers. The set of all complex numbers is denoted **C**, and each member of the set has the form

$$x + iy \qquad x, y \in \mathbf{R}$$

where x is called the "real part", and iy is called the "imaginary part." Complex numbers may be added, subtracted, multiplied and divided according to the usual rules of algebra.

If c_1 and c_2 (read "c-sub-one and c-sub-2") are two complex numbers, with $c_1 = x_1 + iy_1$ and $c_2 = x_2 + iy_2$, then the rules of complex arithmetic are as follows:

Rule C1 (complex addition): To add two complex numbers, add the real and imaginary parts independently, i.e.,

$$c_1 + c_2 = (x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2)$$

Rule C2 (complex subtraction) (similar to addition):

$$c_1 - c_2 = (x_1 + iy_1) - (x_2 + iy_2) = (x_1 - x_2) + i(y_1 - y_2)$$

Rule C3 (complex multiplication): The product is formed by the ordinary rules of algebra:

$$c_1 c_2 = (x_1 + iy_1)(x_2 + iy_2) = x_1 x_2 + iy_1 x_2 + ix_1 y_2 + i^2 y_1 y_2$$
$$= (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + y_1 x_2)$$

(Remember that by definition, $i^2 = -1$)

Rule C4 (complex division): Again, ordinary algebra is used to define the quotient:

$$\frac{c_1}{c_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{x_1 x_2 + y_1 y_2 + i(y_1 x_2 - x_1 y_2)}{x_2^2 + y_2^2}$$

obtained by multiplying by

$$\frac{x_2 - iy_2}{x_2 - iy_2} \quad \text{which is equivalent to 1.}$$

While a function is most generally stated in algebraic form, it is often enlightening to draw graphs in order to get a clear idea of how a function varies as its argument changes. The conventional graph uses a horizontal line to represent the independent variable, and a vertical scale to represent values of the function. Thus, in order to find the value of a function for some value of the independent variable, say, $x = 3$, we slide one finger along the horizontal axis until we point at 3, then move straight up (or down) to find the value $f(x = 3)$ (read: "the function f at $x = 3$").

A glance at Figure 1 tells us several things about the function $f(x) = .5x + 1$. First, the graph is a straight line, sloping upwards to the right; second, it crosses the vertical axis at the value +1; third, it crosses the horizontal axis at the value −2. In fact, any function which has the form

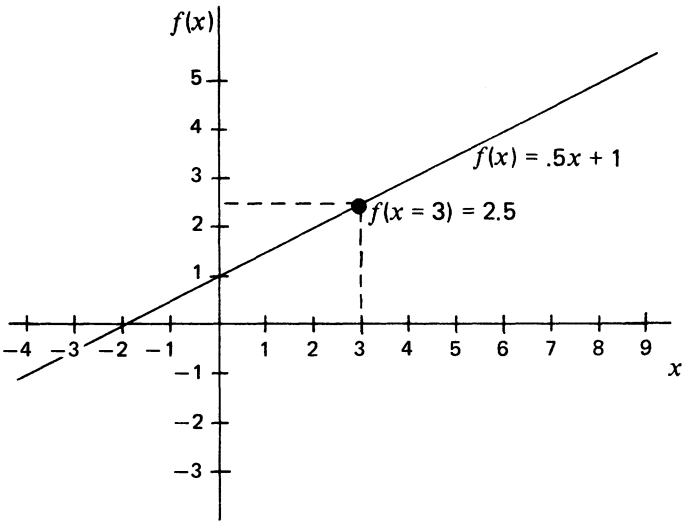


Figure 1. A graph of $f(x) = .5x + 1$

$$f(x) = mx + b \qquad m, b \text{ constants}$$

is the graph of a straight line. m is called the slope of the function since it is the amount by which the function changes for a unit change in x. Setting m to .5, as in Figure 1, means that every time x increases by one, $f(x)$ will increase by .5, hence a positive slope is associated with lines sloping upward to the right. $f(x)$ will always cross the vertical axis when $x = 0$, and since $f(x = 0) = b$, b is called the *vertical axis intercept* of f. The horizontal axis will be crossed, of course, when $f(x)$ equals zero, which occurs in this example at:

$$f(x) = .5x + 1 = 0$$
$$x = -2$$

Actually, Figure 1 is not a graph of $f(x) = .5x + 1$, but more precisely a graph of this function for the values of x between −4 and +9, or in most proper notation:

$$f(x) = .5x + 1 \qquad -4 \leqslant x \leqslant +9$$

The original function could extend for *all* x, that is $-\infty < x < +\infty$, but graphing the entirety of such a function would require a very big piece of paper indeed. Graphs are useful to get the general picture of a function, but they can serve other purposes as well. For example, it is often useful to add graphs directly, especially when it is difficult to do the addition algebraically, or when the algebraic sum of two functions is difficult to interpret. Graphical addition of two functions consists of *carefully* drawing both functions on the same graph, and then *carefully* adding up the vertical distances for all (or many) values of the independent variable, to obtain a graph of the sum function (see Figure 2). Such graphical techniques are, of course, only approximate, but often sufficient to gain considerable insight into the shape of composite functions.

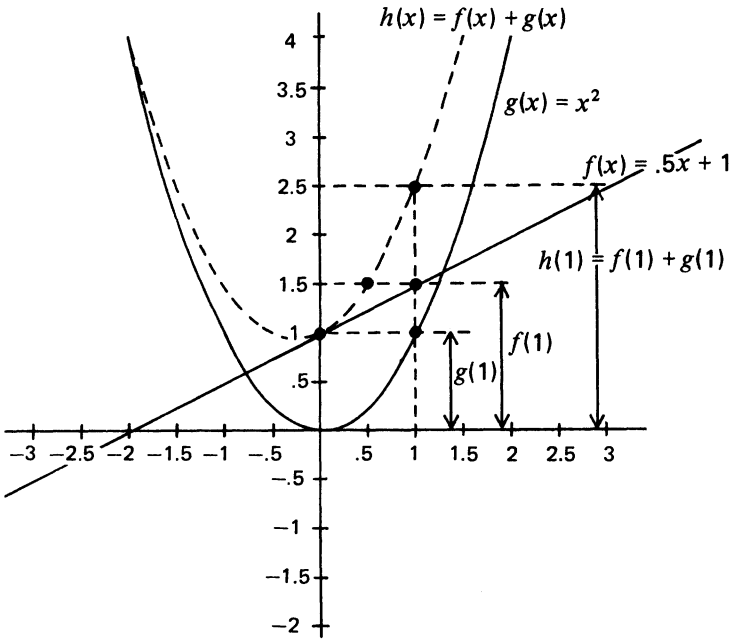


Figure 2. Graphical addition of $f(x) = .5x + 1$ and $g(x) = x^2$ to get graph of $h(x) = f(x) + g(x) = x^2 + .5x + 1$

Polynomials and Roots

A *polynomial* is an algebraic expression which has the form:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n$$

The a's are constants (numbers) called *coefficients*, and the highest power of x which occurs in any given polynomial (n) is called the *degree* of the polynomial. Thus $f(x)$ in Figure 2 is a first-degree polynomial, since the greatest power of x in $.5x + 1$ is one. Both $g(x)$ and $h(x)$ from the same figure are second degree, or *quadratic*, polynomials. Third degree polynomials are called *cubic*, fourth degree *quartic*, and so on, though after that one rarely hears of, say, "quintic polynomials" instead of "fifth-degree polynomials." A polynomial is "solved" by setting it to zero, and finding which values of the independent variable make the equation true. For example, to find the roots of the quadratic equation $x^2 + x - 6 = 0$, we can do any of at least three things:

1. *try every value* of x and see when the formula is true,
2. try to *factor* the polynomial, or
3. use the *quadratic formula*, which will give the roots for *any* quadratic polynomial.

Method 1 may sound a bit absurd, but sometimes it is the best we can do. Method 2 means trying to write the polynomial in the form $(x - z_1)(x - z_2) = 0$. z_1 and z_2, are called the "zeroes" of the function, since if x is equal to z_1, the first factor, and hence the product, will be zero; and similarly for $x = z_2$. Method 3 requires remembering the general solution for any second-degree polynomial (or looking it up), called *the* quadratic formula:

if the equation has the form

$$ax^2 + bx + c = 0$$

then

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The method 2 solution yields:

$$x^2 + x - 6 = 0$$
$$(x + 3)(x - 2) = 0$$
$$x = -3 \text{ or } +2$$

The method 3 solution, with $a = 1$, $b = 1$, and $c = -6$ also yields

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-1 \pm \sqrt{1^2 - 4 \cdot 1 \cdot (-6)}}{2 \cdot 1}$$

$$= \frac{-1 \pm \sqrt{25}}{2} = \frac{-1 \pm 5}{2} = -3 \text{ or } 2$$

What about such formulas as $x^2 + 1 = 0$? The quadratic formula works just as well on those:

$$a = 1, b = 0, c = 1, \text{so}$$

$$x = \frac{0 \pm \sqrt{-4}}{2} = \frac{\pm 2i}{2} = +i \text{ or } -i$$

which says that again there are 2 roots, and that they are both imaginary. In factorial form, we could have written

$$x^2 + 1 = (x - i)(x + i) = 0$$

The Fundamental Theorem of Algebra states that *any* n^{th}-degree polynomial *always* has exactly n roots, that they may in general be complex (having both real and imaginary parts), and that all the roots may not be different from each other (distinct). Also, we might have guessed that if $+i$ is a solution to $x^2 + 1 = 0$, then $-i$ is also, since complex roots always appear in conjugate pairs if the coefficients of the polynomial are real numbers. (If $c = x + iy$ is a complex number, then its conjugate, written c^*, is $x - iy$.)

If the general formula method works so well, why would we ever use factoring, or trial and error? The answer is both simple and unfortunate: General formulas exist only for polynomials with degree less than 5, and in fact the French mathematician Galois proved that no such formulas can exist for degree 5 or more. Even the general quartic formula is very complicated; it is often easier to factor than to use it! And finally, trial and error solutions are often implemented with computers, using special guessing algorithms such as Newton's Method, which work remarkably well.

Exponents, Logarithms, and the Number *e*

If we say that addition and subtraction are easy, that multiplication and division are harder, and that taking a number to a power is most difficult, then the rules of expo-

nents show us how many problems in mathematics may be made one level easier! It is important to remember which *kinds* of numbers these rules apply to, so in the following list, we will use p and q to stand for any real numbers (that is, $p, q \in \mathbf{R}$), a and b are positive reals ($a, b \in \mathbf{R}^+$), and m and n are positive integers ($m, n \in \mathbf{N}$).

Rule E1: $a^p \cdot a^q = a^{p+q}$

Rule E2: $\frac{a^p}{a^q} = a^{p-q}$

Rule E3: $(a^p)^q = a^{pq}$

Rule E4: $\sqrt[n]{a^m} = a^{m/n}$

Rule E5: $a^{-p} = \frac{1}{a^p}$

Rule E6: $a^0 = 1$ (if $a \neq 0$)

Rule E7: $\sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}$

Rule E8: $(ab)^p = a^p b^p$

Using these rules, we can deduce such things as $4^{.5} = 2$ (Rule E4, since .5 = ½), $x^2/x^5 = x^{-3} = 1/x^3$ (Rules E2 and E5), and $\sqrt{6}/\sqrt{2} = \sqrt{3}$ (Rule E7). In fact, the first 3 rules are so useful in doing calculations, that the entire system of logarithms has been devised to make them universally applicable to the more "difficult" problems of multiplication, division, and exponentiation.

If $a^p = x$, where a is not 0 or 1, then p is called the *logarithm to the base a of x*, written $\log_a x = p$. Thus, $\log_2 8 = 3$, since $2^3 = 8$, and $\log_{10} 10000 = 4$, since $10000 = 10^4$. The rules for logarithms are derived from E1, E2 and E 3, above:

Rule L1: $\log_a xy = \log_a x + \log_a y$

Rule L2: $\log_a \frac{x}{y} = \log_a x - \log_a y$

Rule L3: $\log_a x^y = y \log_a x$

where $x, y \in \mathbf{R}$.

Also, if $\log_a x = p$, then x is called the antilogarithm of p to the base a, written $x = \text{antilog}_a p$, since by definition $a^p = x$. Any number except 0 or 1 may be used for the base, but in fact only three numbers are used very often: 10, 2, and $e = 2.71828\ldots$. Logarithms to the base 10 are used because we commonly use a decimal (base 10) number system for everything else! Logarithms to the base 2 are very often encountered in the relatively new fields of computer science and information theory, since computers typically operate using binary arithmetic (internally), and both computers and information theory define the unit of information as a *bit* (short for *binary digit*). Logarithms to the base e are called "natural" logarithms, and are the most used in mathematics.

It is hard for us today to appreciate what a boon logarithms were to mathematicians before the advent of computers and pocket calculators. Logarithms were so useful that two 16th century mathematicians literally devoted most of their lives to calculating "log tables" in order to relieve their colleagues of the drudgery of multiplication and division: Briggs calculated the so-called common, Briggsian, or base 10 logarithms, and Napier the "natural", Naperian, base e logarithms. Base 2 logarithms are not found in mathematical handbooks, and they probably never will be, since their computation today is largely a matter of button-pushing. Also, if the log of a number is available in any one base, it is easy to change it to another base using the following relationships:

$$\log_a x = K \log_b x$$

where

$$K = \frac{1}{\log_b a}$$

K is given in the following table for base changes among 10, 2, and e:

		b = 10	b = 2	b = e
a	10	1	0.30103...	0.43429...
	2	3.32193...	1	1.44270...
	e	2.30259...	0.69315...	1

Thus

$$\log_{10} x = .30103 \log_2 x = .43429 \ln x$$

and so on, where ln stands for "natural logarithm" (i.e., $\ln x = \log_e x$). Logarithms are defined only for positive numbers.

Where does the number e come from? Unfortunately, its true origins are buried deep within calculus, which is not a part of our subject matter, but some of its properties, as we shall see, turn out to be remarkable. e is an irrational number like π, which means that its decimal expansion is both infinite and that it never repeats itself:

$$e = 2.71828\ 18284\ 59045\ 23536\ 0287\ldots$$

If you would like to calculate it to more accuracy than this, the following formula may be used:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots$$

where $n!$ means *n factorial*, which is the product of all the integers from one to n (3! = 6, 4! = 24, 5! = 120, etc.).

A more useful form of this infinite expression yields the value of e raised to any power x:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

Another way to write the same thing is with sum notation:

$$e^x = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}$$

which "says" exactly the same thing. The capital sigma (Σ) is

used to denote that we should add up all values of $x^n/n!$ starting with $n = 1$, then $n = 2$, etc. (It is read: "the sum over n from one to infinity of x to the n divided by n factorial".)

Sums and Series

Such formulas as the one above for e^x are called infinite series, or infinite sequences, since there are infinitely many terms in the sum, even though we know what any one of them would be. Such sums need not be infinite, of course. For example, the following formula illustrates a finite sum:

$$1 + 2 + 3 + \cdots + n = \sum_{k=1}^{n} k$$

which is just the sum of the first n integers. It is both interesting and useful that many such sums have a general, or "closed-form" formula, making it unnecessary to carry out the lengthy addition sequence. For example:

$$\sum_{k=1}^{n} k = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

The closed form is clearly more useful if n is greater than 3 or 4 or so. Other sum formulas often crop up in digital signal processing. For example

$$\sum_{k=0}^{\infty} ar^k = a + ar + ar^2 + \cdots = \frac{a}{1-r}$$

This sum exists only if $r < 1$, since otherwise the sum will be infinite. a is the first term in the sequence, and r is called the ratio, since it is multiplied by any term to get the next term in the sequence. Thus, we see that

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \sum_{k=0}^{\infty} 2^{-k} = \frac{1}{1-\frac{1}{2}} = 2$$

If there is not an infinite number of terms, we can remove the restriction that r be less than 1:

$$\sum_{k=0}^{n-1} ar^k = \frac{a(1-r^n)}{1-r} \qquad r \neq 1$$

If $r \neq 1$, and the last term $l = ar^{n-1}$, then this sum is also equal to

$$\frac{a - rl}{1 - r}$$

Trigonometry

[It has been said that a tribe called the Trigonometric Indians once roamed the earth, that they spoke in sine language, and never used wrong angles. The secret name of their beautiful princess was known only to initiates, for it conveyed all of their secrets at once. The name of their princess was Sohcahtoa.]

If we label a right triangle (one which contains a right angle) with respect to an angle θ (see Figure 3), side O is *o*pposite the angle, side A is *a*djacent to angle θ, and side H is, of course, the *h*ypotenuse of the triangle.

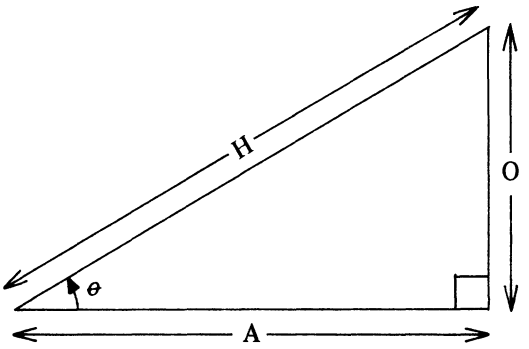


Figure 3. A right triangle with inscribed angle θ and sides O, A, and H

The 3 basic trigonometric functions are defined as follows:

$$\text{sine of } \theta = \sin\theta = \frac{O}{H}$$

$$\text{cosine of } \theta = \cos\theta = \frac{A}{H}$$

$$\text{tangent of } \theta = \tan\theta = \frac{O}{A}$$

Clearly, the size of the right triangle doesn't matter, since, for a given angle θ, if we double the length of one of the sides, the others will double as well. Only the *ratios* of their lengths are needed to define the trigonometric functions.

The 3 remaining trigonometric functions are defined in terms of the first 3:

$$\text{cosecant of } \theta = \csc\theta = \frac{1}{\sin\theta} = \frac{H}{O}$$

$$\text{secant of } \theta = \sec\theta = \frac{1}{\cos\theta} = \frac{H}{A}$$

$$\text{cotangent of } \theta = \cot\theta = \frac{1}{\tan\theta} = \frac{A}{O}$$

Radians, Degrees, and Grads

As almost everyone knows, if you slice a pie into 360 equal wedges, you have not only very small slices to eat, but the angle at the tip of each slice will be one degree (1°). If you are very hungry, however, and slice the pie into 4 equal pieces, the angle at the tip of each slice will be 90°, which is exactly right.

Another measure is to divide the circular pie into 400 equal pieces, or 400 grads. But by far the most common measure of angles used in mathematics is the *radian*. Since the ratio of the circumference of a circle to its diameter is $\pi = 3.14159\ 26535\ldots$, and since the radius of a circle

is exactly one half its diameter, the circumference of a circle is exactly 2π times the length of its radius, and we say there are 2π radians in a circle. A right angle is then any of 90°, 100 grads, or $\pi/2$ radians, depending on which measure we are using.

If we choose a circle with radius equal to one unit, and we inscribe our right triangles inside the circle (see Figure 4),

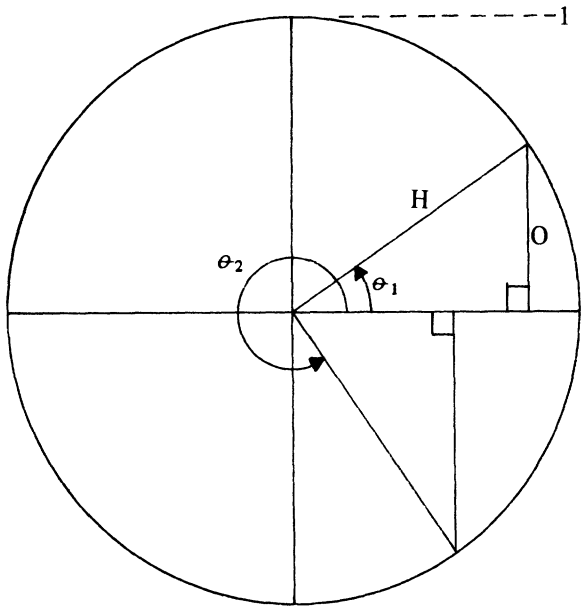


Figure 4. A unit circle with inscribed right triangles

we can "solve" the triangles conveniently with the Pythagorean theorem: $O^2 + A^2 = H^2$, or $O^2 + A^2 = 1$, $O = \sqrt{1 - A^2}$, and $A = \sqrt{1 - O^2}$. Angles are conventionally measured counter-clockwise from the right hand horizontal axis (see θ_1, and θ_2 in the figure). Angles measured in a clockwise direction are considered negative.

We can treat the angle θ as an independent variable and graph the basic functions as shown in Figure 5.

The *inverse* trigonometric functions are defined in a similar way to the antilogarithm: if $\sin \theta = x$, then the arcsine of $x = \sin^{-1} x = \theta$, and so on, for each of the six trigonometric functions.

We can see from the graph of $\sin \theta$ that the function is *periodic*, that is, it repeats itself over and over again as θ gets larger or smaller by 2π, which is called the *period* of $\sin \theta$. Furthermore, $\sin \theta$ always has a value between +1 and −1 inclusive, so we say that the domain of the sine function is the set of all real numbers between +1 and −1, or in more mathematical form:

$$\sin \theta \in \mathbf{R} \quad , \quad -1 \leqslant \sin \theta \leqslant +1$$

Because of this restricted domain, it is meaningless to write $\sin^{-1} 2 = \theta$, since no angle θ has a sine equal to 2. But what about $\sin^{-1} 1 = \theta$? From the graph, it is clear that $\sin \pi/2 = 1$, so, $\theta = \pi/2$ is one solution to this equation. But $\sin 5\pi/2$ is also equal to one, as is $\sin -3\pi/2$. In fact $\sin^{-1} 1 = \theta$ has infinitely many solutions, all of the form $\theta = \pi/2 + k2\pi$, where k is any integer. The *principle values* of the inverse trigonometric functions are chosen to be close to $\theta = 0$, and these are used to resolve the problem of which answer to choose. Thus:

$$-\frac{\pi}{2} \leqslant \sin^{-1} x \leqslant \frac{\pi}{2} \ ,$$

$$0 \leqslant \cos^{-1} x \leqslant \pi \ , \text{ and}$$

$$-\frac{\pi}{2} \leqslant \tan^{-1} x \leqslant \frac{\pi}{2} \ .$$

Inspection of Figure 5 also shows that the sine and cosine functions are also identical to each other, except for their starting place at $\theta = 0$, i.e., they differ only in *phase*:

$$\sin\left(\frac{\pi}{2} + \theta\right) = \cos \theta \ , \text{ and}$$

$$\cos\left(\theta - \frac{\pi}{2}\right) = \sin \theta \ .$$

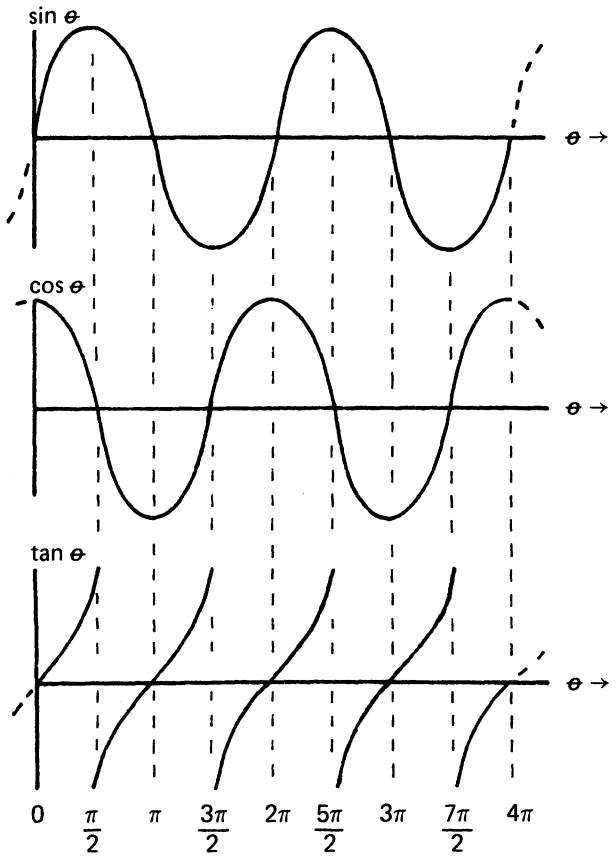


Figure 5. Graphs of $\sin \theta$, $\cos \theta$, and $\tan \theta$ as functions of θ, θ in radians.

Trigonometric Identities

Many formulas may be derived from the basic definitions of the trigonometric functions which are often useful in the manipulation of equations involving trigonometric functions. They are called identities since, like all equations, the expressions on either side of the equal sign "say" exactly the same thing, but in a useful way. In the following identities, A and B are any angles:

(T1): $\sin 2A = 2 \sin A \cos A$

(T2): $\cos 2A = \cos^2 A - \sin^2 A$

(T3): $\sin^2 A = \frac{1}{2} - \frac{1}{2} \cos 2A$

(T4): $\cos^2 A = \frac{1}{2} + \frac{1}{2} \cos 2A$

(T5): $\sin A + \sin B = 2 \sin \frac{1}{2}(A + B) \cos \frac{1}{2}(A - B)$

(T6): $\sin A - \sin B = 2 \cos \frac{1}{2}(A + B) \sin \frac{1}{2}(A - B)$

(T7): $\cos A + \cos B = 2 \cos \frac{1}{2}(A + B) \cos \frac{1}{2}(A - B)$

(T8): $\cos A - \cos B = 2 \sin \frac{1}{2}(A + B) \sin \frac{1}{2}(B - A)$

(T9): $\sin A \sin B = \frac{1}{2}[\cos(A - B) - \cos(A + B)]$

(T10): $\cos A \cos B = \frac{1}{2}[\cos(A - B) + \cos(A + B)]$

(T11): $\sin A \cos B = \frac{1}{2}[\sin(A - B) + \sin(A + B)]$

(T12): $\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$

(T13): $\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$

These identities are fairly easy to derive from each other, and, of course, many more exist.

Like e^x, the sine and cosine functions may be represented as summation series:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

where x is an angle measured *in radians.*

Using Trigonometric Functions to Represent Musical Sounds

One of the great pleasures of mathematics is that it can be used to understand portions of the "real world." If some phenomenon naturally behaves in a way which can be described mathematically, mathematics provides a wealth of intellectual "tools" which allow that phenomenon to be analyzed (i.e., understood), perhaps modified in a predictable and desirable way, and possibly synthesized (created in a new and flexible way). Such phenomena are the sounds of music and speech.

Sounds are vibrations in the air to which our ears are sensitive. Acoustical studies have shown that the quality of a sound as we perceive it is related to certain characteristics of the "shape" of the vibrations, i.e., we draw a graph of the air pressure fluctuations as a function of time and observe its graphical shape. If the waveshape is fairly regular and repetitive (i.e., roughly periodic) it will sound like a tone with a steady pitch, such as a violin note or a fog horn. If the waveform is irregular and aperiodic, the sound will have little or no pitch, but instead sound like a noise such as steam rushing or a cymbal crash. In speech, periodic waveforms are associated with voiced sounds, such as vowels and voiced consonants. Aperiodic waveforms are associated with unvoiced consonants, such as s and f. The period of a periodic waveform is closely related to what pitch it will have. Period and frequency are two names for two ways of describing the same thing: how *often* does the waveform regularly repeat itself. If the frequency of repetition is between about 20 to 20, 000 times per second, then the vibration will be heard as a sound. In other words, pitched sounds have periods ranging from about 1/20 to 1/20, 000 of a second. The amplitude, or strength, of the vibration is a measure of how far the pressure deviates from the atmospheric mean. One could measure the peak deviation from the mean, or possibly the average deviation, but the word amplitude generally refers to the peak deviation, unless stated otherwise, and is related to our perception of the loudness of a sound. Finally, the general shape of the waveform determines its tone quality, or timbre. All of these factors interact perceptually. For instance, the pitch can be affected by the amplitude and the shape as well as the period of a waveform. Hence it is important to distinguish between frequency, which is a measure of the repetition rate of a periodic waveform, and pitch, which is our perception of something like the "tonal height" of a sound.

An important mathematical tool which will be described in Part II of this tutorial is *Fourier's theorem*, which states that *any* periodic waveform can be described as the sum of a number, possibly an infinite number, of sinusoidal variations, each with a particular frequency, amplitude, and phase. Futhermore, there is a method for determining exactly what these frequencies, amplitudes, and phases must be in order to re-construct the waveform by adding together sine waves, which are seen to be the basic "building blocks" of periodic waveforms. Actually there are a few other requirements as well as periodicity; suffice it to be said that any waveform which could exist in the physical world will obey these other conditions (called the Dirichlet conditions).

Stated mathematically, the waveform must obey the condition $f(t) = f(t + T)$, where f is the periodic waveform, t is time, and T is the period of the waveform. Then

$$f(t) = \sum_{k=0}^{\infty} A_k \sin(k\omega t + \phi_k)$$

where:

- A_k is the amplitude of the k^{th} sinusoidal component of $f(t)$,
- ω is the *fundamental frequency* ($= 1/T$) of the waveform times 2π, and
- ϕ_k is the phase of the k^{th} sinusoidal component of $f(t)$.

Another way which is more commonly used of stating the same thing is

$$f(t) = \sum_{k=0}^{\infty} (a_k \cos k\omega t + b_k \sin k\omega t)$$

where both the amplitudes and phases of the previous expression are imbedded in the a's and b's of the second expression. To see that this is so, we can use trigonometric identity T12 (we omit the subscripts for the moment):

$$\begin{aligned} A \sin(k\omega t + \phi) &= A(\sin k\omega t \cos\phi + \cos k\omega t \sin\phi) \\ &= A \sin\phi \cos k\omega t + A \cos\phi \sin k\omega t \\ &= a \cos k\omega t + b \sin k\omega t \end{aligned}$$

where

$$a = A \sin\phi \quad \text{and} \quad b = A \cos\phi$$

Similarly, we can show from these expressions for a and b that:

$$a^2 + b^2 = (A \sin\phi)^2 + (A \cos\phi)^2$$
$$= A^2 \sin^2\phi + A^2 \cos^2\phi \text{ (by Rule E8)}$$
$$= A^2 (\sin^2\phi + \cos^2\phi)$$
$$= A^2 \text{ (since } \sin^2\phi + \cos^2\phi = 1 \text{ by Pythagorean theorem)}$$

Therefore,

$$A = \sqrt{a^2 + b^2}$$

Also,

$$\frac{a}{b} = \frac{A \sin\phi}{A \cos\phi} = \tan\phi \text{ (by the basic definition of sin, cos, and tan)}$$

Therefore

$$\phi = \tan^{-1}\frac{a}{b}$$

What we have done is not only to show that the two formulas for $f(t)$ above are the same, but also how to derive one form from the other.

The Most Beautiful Formula in Mathematics

In the 19th century, the German mathematician Euler proved the following remarkable identity:

$$e^{ix} = \cos x + i \sin x$$

thereby relating algebraic exponentials to the trigonometric functions. This key formula is the basis for much of the mathematics used in signal processing, for it allows some very powerful manipulations to be made using sinusoidal functions that would otherwise prove very tedious. For example, by using rules E3 and E8 regarding exponents, it is easy to see that

$$(re^{i\theta})^p = r^p e^{ip\theta} \text{ (De Moivre's theorem)}$$

By using Euler's relation we can see that this innocent-looking equation "says" the same thing as

$$[r(\cos\theta + i \sin\theta)]^p = r^p(\cos p\theta + i \sin p\theta)$$

This form of De Moivre's theorem may be used to demonstrate many of the trigonometric identities in a very economical way. For example, if we let $r = 1$ and $p = 2$,

$$\cos 2\theta + i \sin 2\theta = (\cos\theta + i \sin\theta)^2$$
$$= \cos^2\theta - \sin^2\theta + i\, 2 \sin\theta \cos\theta$$

Since two complex numbers are equal if an only if both their real parts are equal and their imaginary parts are equal, this simple procedure has just shown that

$$\cos 2\theta = \cos^2\theta - \sin^2\theta \quad \textit{and}$$
$$\sin 2\theta = 2 \sin\theta \cos\theta$$

This demonstrates the validity of both identity T1 *and* identity T2. In other words, by using the complex exponential in Euler's relation, we can, in effect, solve two equations at once!

But Euler's relationship tells us something else, something which is at the same time profound, elegant, and simple. It tells us of a relationship among *all* of the known fundamental constants of mathematics in a way that mathematicians, and perhaps by now the reader, can only consider *beautiful*. It is easy to see from Figure 5 that the following relationships are true:

$$\cos\pi = -1$$
$$\sin\pi = 0$$

If we substitute π for x in Euler's relationship, we are unerringly led to what has been rightly called "the most beautiful formula in mathematics:"

$$e^{i\pi} = \cos\pi + i \sin\pi$$
$$= -1 + 0$$

Therefore:

$$e^{i\pi} + 1 = 0$$

Conclusion of Part I

Mathematicians create mathematics, the rest of us merely use, and sometimes appreciate, what the mathematicians have created. Computers have at the same time reduced the need for human calculation and increased many fold the utility of human mathematics, especially to non-mathematicians who can now apply these powerful tools to the study of virtually anything. We now have to discover the models which state the *correspondence* between phenomena and mathematics. Once we know that a vibration is periodic, for instance, we know that we can use Fourier's techniques to find the elemental building blocks of the vibration. We also know that if we add up the same building blocks ourselves that we can reproduce the phenomenon at will. Or perhaps we might improve on the original a bit, once we're sure that the original is understood correctly.

Thus we can make machines that talk and sing, we can study the waves in the ocean, and the vibrations in an earthquake. Fourier himself was studying the transfer of heat at the time he devised his theorem about the way waves are shaped, which is all the more remarkable because it *doesn't* matter! Mathematics deals with the relationships, not with the things *per se*, and if a theorem correctly states that "A" has relation "R" to "B", and we note that the height of a mountain could be thought of as thing "A", then we know that something else will correspond to "B", and "R" will tell us where to look for it.

For the reader interested in using mathematics, a good mathematical handbook is heartily recommended, such as the excellent and inexpensive *Mathematical Handbook of Formulas and Tables* by Murray R. Spiegel, available as a Schaum Outline Series paperback (McGraw-Hill). For the reader interested in understanding mathematics in greater detail, it is recommended that this be treated in the same way as a desire to learn to play a piano: a good teacher and regular practice will suffice in a way that nothing else can. Reading books helps,* and there are certainly plenty of books to read on mathematics at every conceivable level, but not much more than it helps to read a book about playing a piano.

* An excellent book to read is *The Foundations of Mathematics* by Stewart and Tall (Oxford University Press).

Some Problems

1. Solve these equations for x:

   a) $x^2 - 1 = 0$ (Two solutions)
   b) $x^4 - 1 = 0$ (Four solutions)
   c) $\log_{10} x = .43429...$ (Hint: $\log_a a = 1$)
   d) $\text{antilog}_2 5 = x$
   e) $2^x = 20$ (Hint: $20 = 10 \cdot 2$)
   f) $ax^2 + bx + c = 0$

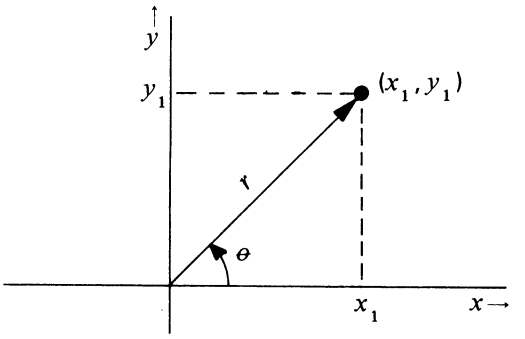
2. The sequence $(i, i^2, i^3, i^4, i^5, \ldots)$ is *periodic*, since it eventually repeats itself every n numbers. Find n.

3. Find the sum of all the integers between 100 and 1000, inclusive.

4. Rewrite the following sequences using summation notation (Σ) and find their solutions:

   a) $1,000,000 + 100,000 + 10,000 + \ldots$ (infinitely many terms)
   b) $100 + 200 + 400 + 800 + \ldots$ to 10 terms
   c) $10^6 + 2.5 \times 10^5 + 6.25 \times 10^4 + \ldots$ to n terms

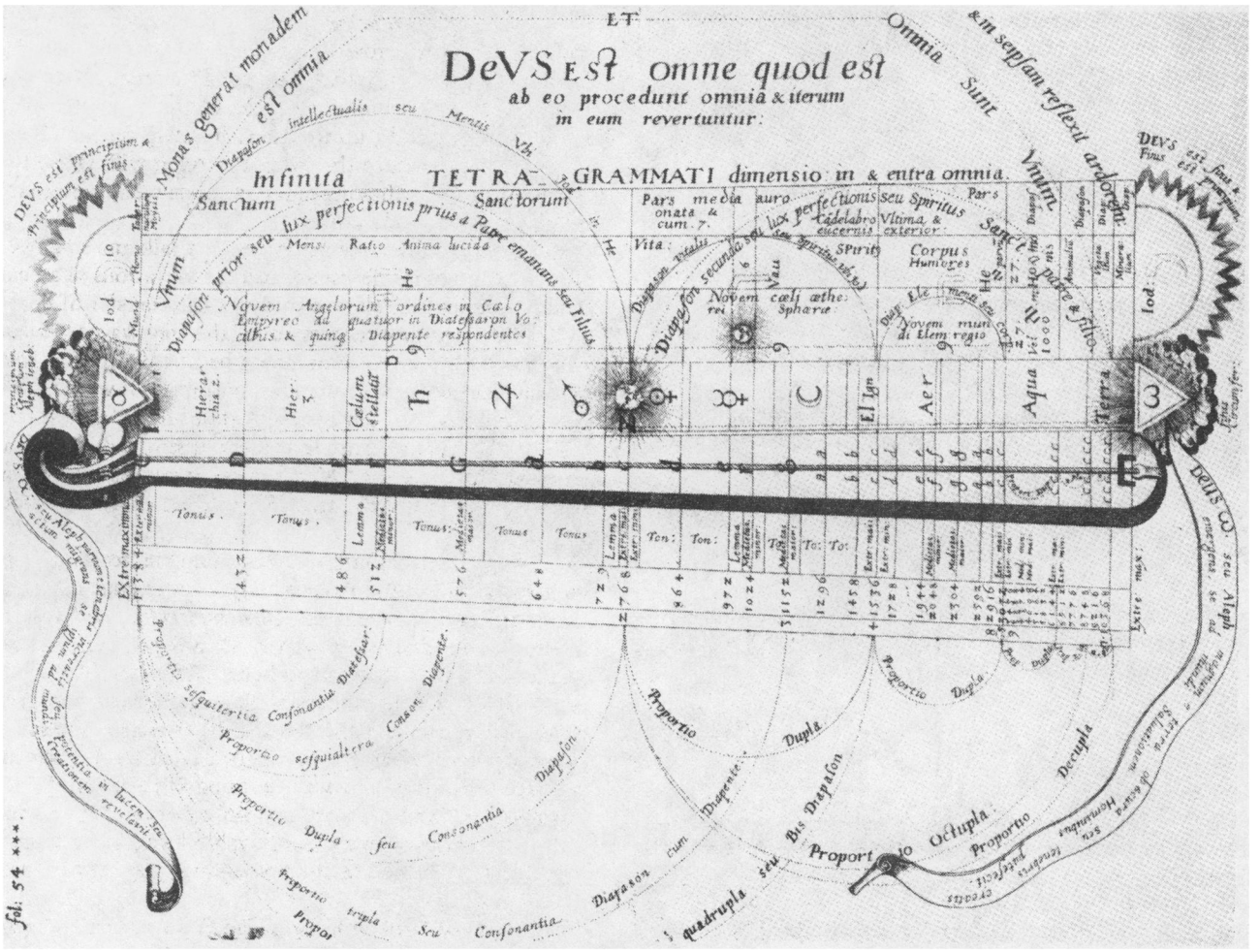
5. If we graph a complex number $c = x_1 + iy_1$ on a plane, we can use x_1 and y_1 for the horizontal and vertical coordinates of a point on that plane



   If we draw a straight line from the origin (point (0, 0)) to point (x_1, y_1), we could also use the length r and angle θ of that line to define the locations of the point (x_1, y_1). Find r and θ in terms of x_1 and y_1.
   (Hint: Pythagoras' theorem states that the square of the length of the hypotenuse of a right triangle is equal to the sum of the squares of the other sides).

6. Show that $\sin^2\theta + \cos^2\theta = 1$.

7. Show that $e^{i\theta} = \cos\theta + i \sin\theta$
   (Hint: Use the summation formula, also called the *power series expansion*, for e^x).



Monochord illustrating universal relationships (from Robert Fludd's *Monochordum mundi*, Frankfort, 1622.)