Chapter Five: Newton, fluxions and forces

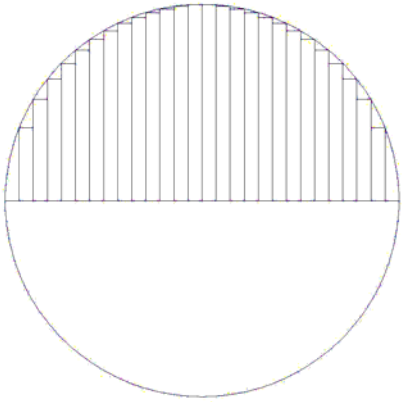
Newton was born one year after Galileo died, 1643. Some have suggested he was a reincarnation of Galileo! Newton's accomplishments were truly amazing and his work awed his contemporaries and the generations that followed him. On the one hand, he was one of the most powerful mathematicians of all time, inventing calculus and thus putting order to all the bits and pieces of reasoning dealing with areas and volumes and rates and ratios which goes back to Eudoxus. And on the other hand, he was an amazing physicist giving, among many things, a definitive answer to the age-old question of planetary motion, and in so doing, relating it to the falling of bodies (such as an apple!) here on earth.

Newton had a sad life: his father died when he was 3 months old and he was raised in a loveless family of grandparents, mother and stepfather. He never married, had a vicious and vindictive temper and suffered several nervous breakdowns. He kept his ideas to himself and avoided publication (and the possibility of criticism) as much as possible, until friends or enemies forced him to publish. The core of his great ideas all seem to have been discovered in a period of about 2 years when he was forced by the plague at age 22 to work all alone at his home in Lincolnshire.

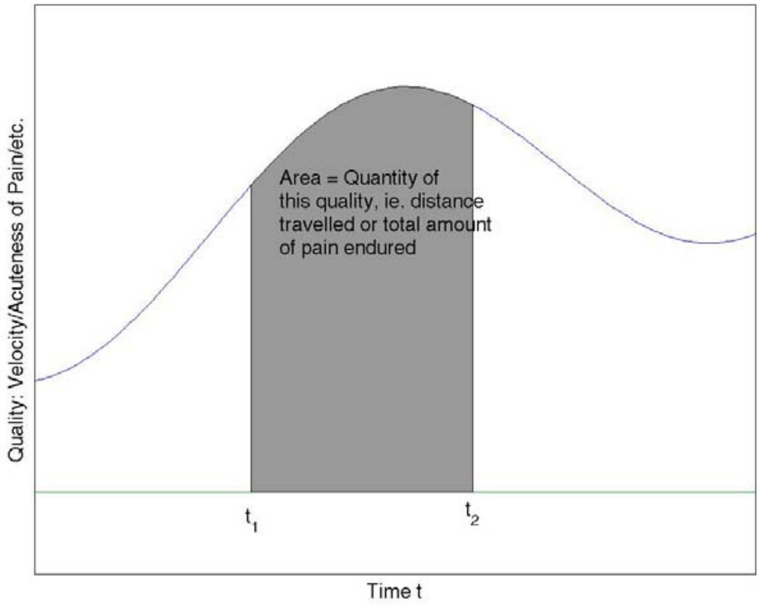
Newton's most famous book is entitled *Philosophiae Naturalis Principia Mathematica* (Mathematical Principles of Natural Philosophy), and was published in 1687, some 20 years after the key ideas had been worked out and then only because of the insistence of his friend, the astronomer Halley, and the counter-claims of his enemy Hooke. It is a monumental book, modeled after Euclid's *Elements* in that (a) it starts with Axioms and then is all Definitions, Propositions and Theorems and (b) the methods of reasoning are purely geometric. Not that he wasn't an expert in algebra too, but he kept this as well as his whole development of calculus to himself until later. He had written out his theory of calculus in 1671 in a manuscript entitled *De Methodis Serierum et Fluxionum* (On the Method of Series and Fluxions), but he no one saw this for 40 years until he brought out a modified version in 1711. He actually described what he had done in the 3rd person in course of his priority fight with Leibniz over the discovery of calculus:

> *By the help of the new analysis Mr. Newton found out most of the Propositions in his Principia Philosophiae: but because the Ancients for making things certain admitted nothing into Geometry before it was demonstrated synthetically, he demonstrated the Propositions synthetically, that the System of the Heavens might be founded upon good Geometry. And this makes it now difficult for unskillful men to see the Analysis by which those Propositions were found out.*

We don't have to follow Newton and stick with 'the good Geometry'. In fact, knowing the techniques of calculus make a whole lot of things much much easier, not only for Newton but for everyone after him. Let's first study what Newton did inventing the calculus (at the same time as Leibniz) before going on to his laws for the universe. Recapping what we have seen, Archimedes had the basic idea of estimating an area by dividing it up into a very large number *n* of very small pieces and then getting the exact value by letting *n* go to infinity. An example is shown below. Oresme, on the other hand, graphed many types of ‘qualities’ and interpreted the area of his graph as being the total quantity of that quality: in particular, if he graphed the velocity of an object, he realized that this area was the total distance traveled:



Integration in Archimedes



Integration in Oresme

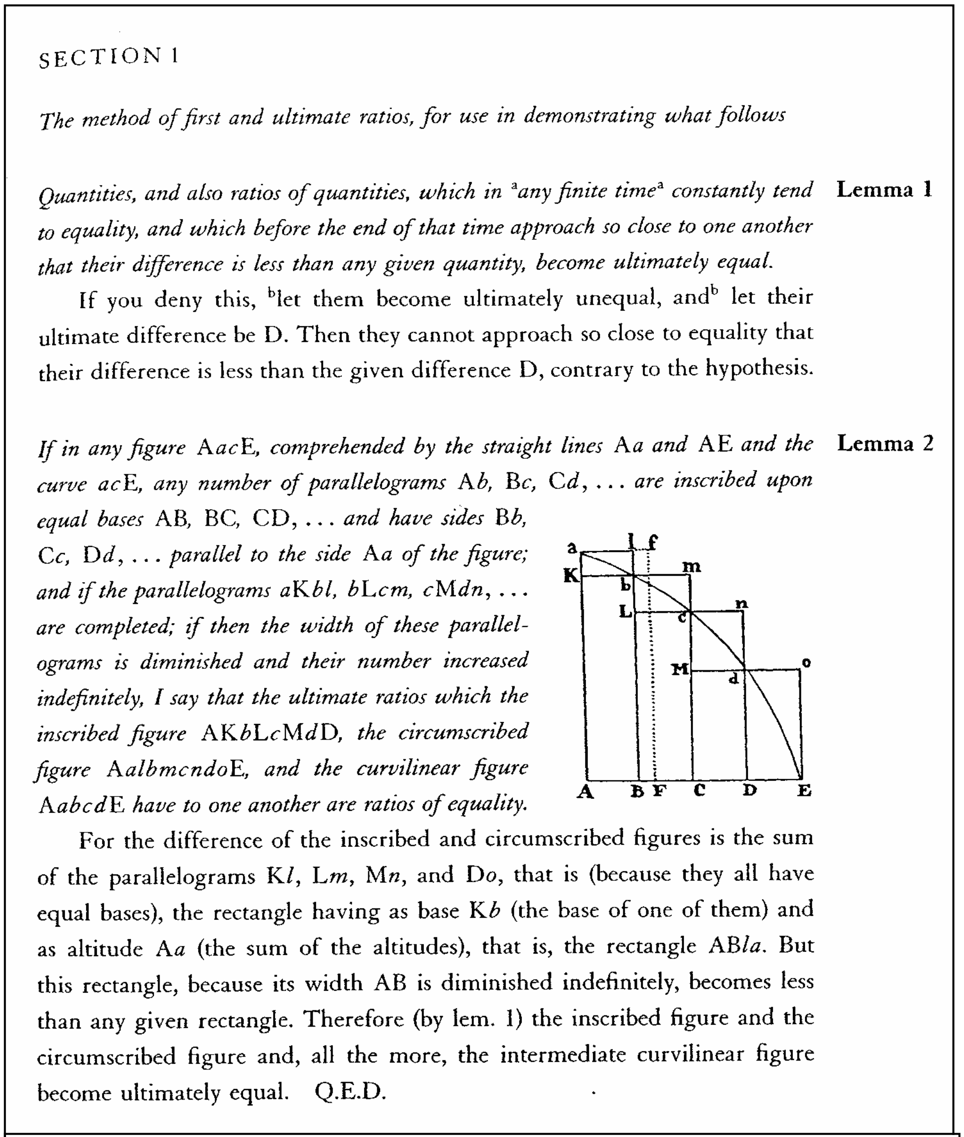
Going over to modern language and notation, let’s call the ‘quality’ being graphed a ‘function’ and write it $x(t)$. Newton, as we will see, called it a *fluent* (meaning a smoothly flowing measurable thing). If the function x was increasing or decreasing uniformly, its rate of increase or decrease is obviously given by $(x(t_2)-x(t_1))/(t_2-t_1)$. But an age-old problem was how to measure the rate of change of a function when this rate of change was not constant: a ‘difformly difform quality’ in Oresme’s language. To measure the *instantaneous* rate of change, you want to use a ratio $(x(t_2)-x(t_1))/(t_2-t_1)$ for t_1 and t_2 very close to each other. If we have excellent data for x at a discrete set of sample points t (this is called a ‘time series’) , then, as far as a computer is concerned, the best definition of its rate of change is going to be this ratio for two adjacent data points. But if we are thinking about an ideal world in which all values of x can be known as accurately as you want, then you can form this ratio for any t_1 and t_2, but it is still not the exact instantaneous rate of change if $t_1 \neq t_2$; but if $t_1 = t_2$, the formula gives us 0/0 which makes no sense! Catch 22! Just as in the argument about integrals, this led many people to introduce a new sort of thing, an *infinitesimal*: let $t_2 - t_1 = dt$ be an infinitesimal increase in t, and let dx be the corresponding infinitesimal increase in x. Then why shouldn’t dx/dt make sense? This is certainly the way the computer does it if, instead of dt, you take the smallest change in t which your data affords (often called Δt). In that case we can write our ratio $(x(t_2)-x(t_1))/(t_2-t_1)$ as $\Delta x/\Delta t$. Unfortunately, coming up with a clear logical definition of what ‘infinitesimal’ means is very hard (though several rather tricky but rigorous methods have been invented). Instead Newton did exactly what modern math books do, which formalizes what

Archimedes, had done: he took the *limit* of the ratios $(x(t_2)-x(t_1))/(t_2-t_1)$ as t_1 and t_2 approach a value t. He called this the *ultimate ratio*. On the right, from *Principia*, is how he defined *limit* (1st lemma) and how he formulated Archimedes' integration method (2nd lemma) in terms of ultimate ratios.

After the passage reproduced on the right, he goes on to talk about measuring area with unequal division of the base and about comparing such divisions of similar figures

In the second box he gets to tangent lines and chords to curves, which is the geometric way of considering the instantaneous rate of change vs. the rate of change over an interval that we have been discussing and says the limiting value of the slope of the chord equals that of the tangent line. (This is the edition of I. B. Cohen showing the 'proof' in edition 3 and, in the footnote, that of edition 1 – both being a bit 'hand-wavy' by modern standards.)

Going over to functions, suppose the arc *ACB* is the graph of *x(t).* Then $(x(t_2)-x(t_1))/(t_2-t_1)$ equals the tangent of the angle between the chord *ABb* and the horizontal, i.e. the slope of the chord. Thus he has identified the 'ultimate value'



SECTION 1

The method of first and ultimate ratios, for use in demonstrating what follows

Lemma 1

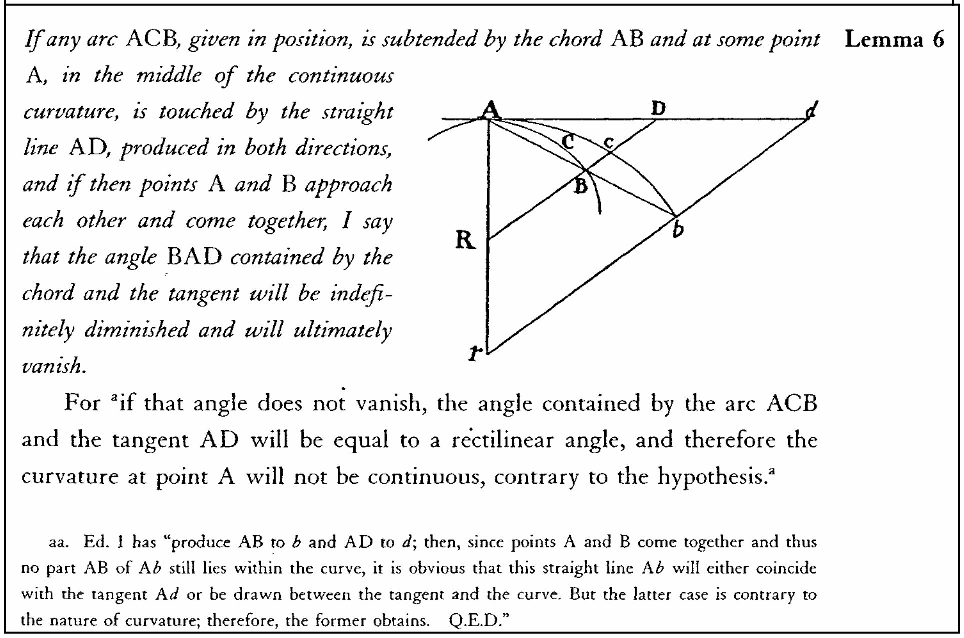
Quantities, and also ratios of quantities, which in [a]any finite time[a] constantly tend to equality, and which before the end of that time approach so close to one another that their difference is less than any given quantity, become ultimately equal.

If you deny this, [b]let them become ultimately unequal, and[b] let their ultimate difference be D. Then they cannot approach so close to equality that their difference is less than the given difference D, contrary to the hypothesis.

Lemma 2

If in any figure A*ac*E, *comprehended by the straight lines* A*a and* AE *and the curve ac*E, *any number of parallelograms* A*b*, B*c*, C*d*, . . . *are inscribed upon equal bases* AB, BC, CD, . . . *and have sides* B*b*, C*c*, D*d*, . . . *parallel to the side* A*a of the figure; and if the parallelograms a*K*bl*, *b*L*cm*, *c*M*dn*, . . . *are completed; if then the width of these parallelograms is diminished and their number increased indefinitely, I say that the ultimate ratios which the inscribed figure* AK*b*L*c*M*d*D, *the circumscribed figure* A*albmcndo*E, *and the curvilinear figure* A*abcd*E *have to one another are ratios of equality.*

For the difference of the inscribed and circumscribed figures is the sum of the parallelograms K*l*, L*m*, M*n*, and D*o*, that is (because they all have equal bases), the rectangle having as base K*b* (the base of one of them) and as altitude A*a* (the sum of the altitudes), that is, the rectangle AB*la*. But this rectangle, because its width AB is diminished indefinitely, becomes less than any given rectangle. Therefore (by lem. 1) the inscribed figure and the circumscribed figure and, all the more, the intermediate curvilinear figure become ultimately equal. Q.E.D.



Lemma 6

If any arc ACB, *given in position, is subtended by the chord* AB *and at some point* A, *in the middle of the continuous curvature, is touched by the straight line* AD, *produced in both directions, and if then points* A *and* B *approach each other and come together, I say that the angle* BAD *contained by the chord and the tangent will be indefinitely diminished and will ultimately vanish.*

For [a]if that angle does not vanish, the angle contained by the arc ACB and the tangent AD will be equal to a rectilinear angle, and therefore the curvature at point A will not be continuous, contrary to the hypothesis.[a]

[aa] Ed. 1 has "produce AB to *b* and AD to *d*; then, since points A and B come together and thus no part AB of A*b* still lies within the curve, it is obvious that this straight line A*b* will either coincide with the tangent A*d* or be drawn between the tangent and the curve. But the latter case is contrary to the nature of curvature; therefore, the former obtains. Q.E.D."

of the ratio $(x(t_2)-x(t_1))/(t_2-t_1)$ with the slope of the tangent line to the curve.

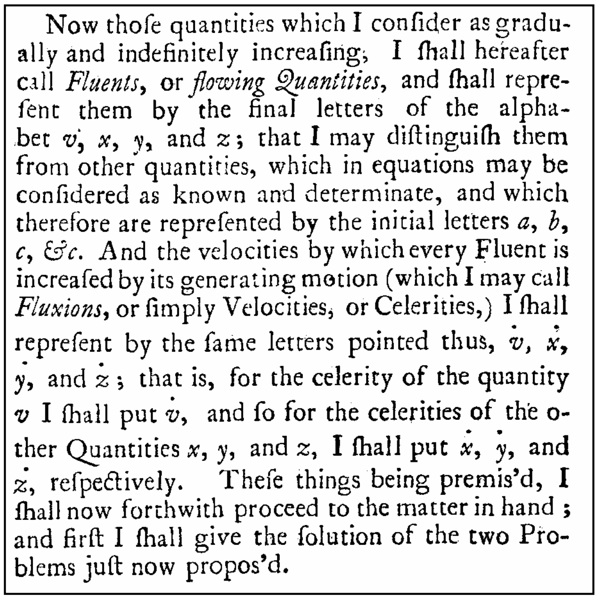
In what follows, he goes on and works out many 'ultimate ratios' such as the ratio of the length of the chord *AB* to the arc length *ACB*. But then he admits this business of using limits can get cumbersome and that it is much easier to work with infinitesimals (called here *indivisibles* but meaning the same thing):

> *In any case, I have presented these lemmas before the propositions in order to avoid the tedium of working out lengthy proofs by reductio ad absurdum in the manner of the ancient geometers. Indeed, proofs are rendered more concise by the method of indivisibles. But since the hypothesis of indivisibles is problematical and this method is accounted less geometrical, I have preferred to make the proofs of what follows depend on the ultimate sums and ratios of vanishing quantities and on the first sums and ratios of nascent quantities, that is, on the limits of such sums and ratios, and therefore to present proofs of those limits beforehand as briefly as I could. For the same result is obtained by these as by the method of indivisibles, and we shall be on safer ground using principles that have been proved. ....*
>
> *It may be objected that there is no such thing as an ultimate proportion of vanishing quantities, inasmuch as before vanishing the proportion is not ultimate, and after vanishing it does not exist at all. But the answer is easy: ... the ultimate ratio of vanishing quantities is to be understood not as the ratio of quantities before they vanish or after they have vanished but the ratio with which they vanish.*

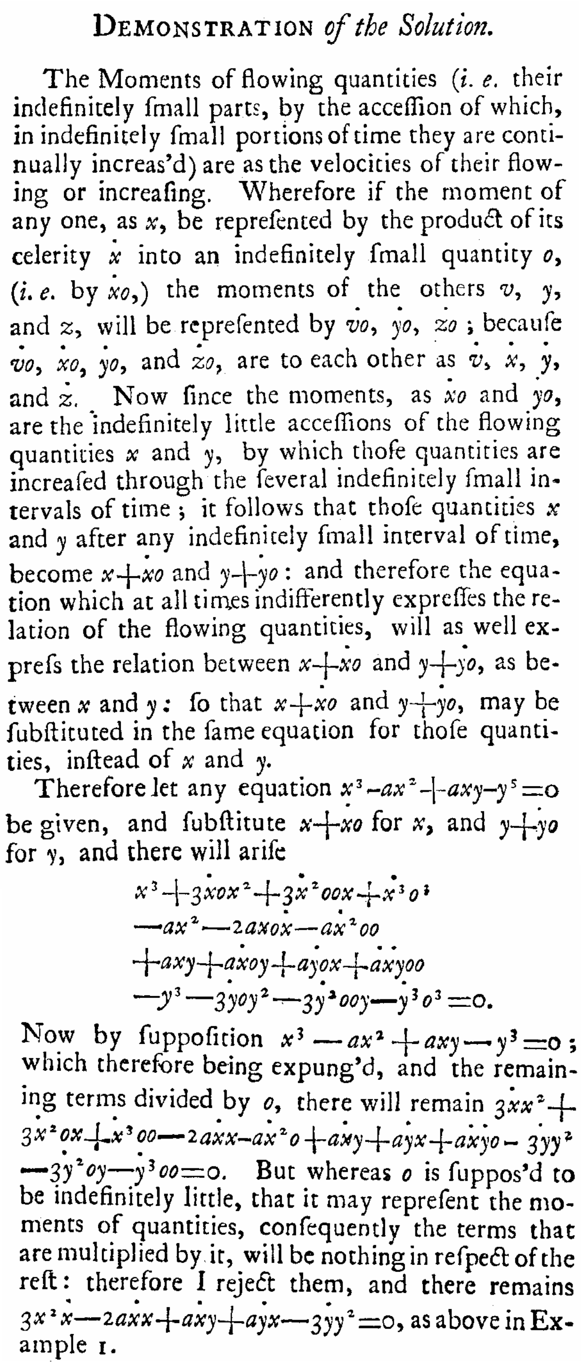
A no nonsense approach! In his book on fluxions, he adopts infinitesimals in an unabashed way. Here is the paragraph in which he sets up his notation (this is a reproduction of the English 1737 translation and, I'm sorry, uses the old 'f' for 's'). First he introduces a general class of situations in which a collection of quantities *v,x,y,z* are all varying with time. Elsewhere he says it doesn't have to be time, but could be any other variable, which we can regard as varying 'equably'. Then their rates of change are always to be denoted by putting a dot over them. We still use this notation, though *dv/dt, dx/dt,* etc is more common. You are probably more used to *dx/dt* so just remember:

$$\dot{x} \text{ and } \frac{dx}{dt}$$

are exactly the same thing!! He then states the two fundamental problems: calculate the velocities, given the quantities and inversely, given the velocities, calculate the original quantities. He proceeds to give the rule by which, if there is a polynomial relation between *x* and *y*, one finds a linear relation between $\dot{x}$ and $\dot{y}$. He gives 5 examples before justifying the rule as shown on the next page.



Now thoſe quantities which I conſider as gradually and indefinitely increaſing, I ſhall hereafter call *Fluents*, or *flowing Quantities*, and ſhall repreſent them by the final letters of the alphabet v, x, y, and z; that I may diſtinguiſh them from other quantities, which in equations may be conſidered as known and determinate, and which therefore are repreſented by the initial letters a, b, c, *&c.* And the velocities by which every Fluent is increaſed by its generating motion (which I may call *Fluxions*, or ſimply Velocities, or Celerities,) I ſhall repreſent by the ſame letters pointed thus, $\dot{v}$, $\dot{x}$, $\dot{y}$, and $\dot{z}$; that is, for the celerity of the quantity v I ſhall put $\dot{v}$, and ſo for the celerities of the other Quantities x, y, and z, I ſhall put $\dot{x}$, $\dot{y}$, and $\dot{z}$, reſpectively. Theſe things being premis'd, I ſhall now forthwith proceed to the matter in hand; and firſt I ſhall give the ſolution of the two Problems juſt now propos'd.



DEMONSTRATION *of the Solution.*

The Moments of flowing quantities (*i. e.* their indefinitely ſmall parts, by the acceſſion of which, in indefinitely ſmall portions of time they are continually increas'd) are as the velocities of their flowing or increaſing. Wherefore if the moment of any one, as x, be repreſented by the product of its celerity $\dot{x}$ into an indefinitely ſmall quantity o, (*i. e.* by $\dot{x}o$,) the moments of the others v, y, and z, will be repreſented by $\dot{v}o$, $\dot{y}o$, $\dot{z}o$; becauſe $\dot{v}o$, $\dot{x}o$, $\dot{y}o$, and $\dot{z}o$, are to each other as $\dot{v}$, $\dot{x}$, $\dot{y}$, and $\dot{z}$. Now ſince the moments, as $\dot{x}o$ and $\dot{y}o$, are the indefinitely little acceſſions of the flowing quantities x and y, by which thoſe quantities are increaſed through the ſeveral indefinitely ſmall intervals of time; it follows that thoſe quantities x and y after any indefinitely ſmall interval of time, become $x+\dot{x}o$ and $y+\dot{y}o$: and therefore the equation which at all times indifferently expreſſes the relation of the flowing quantities, will as well expreſs the relation between $x+\dot{x}o$ and $y+\dot{y}o$, as between x and y: ſo that $x+\dot{x}o$ and $y+\dot{y}o$, may be ſubſtituted in the ſame equation for thoſe quantities, inſtead of x and y.

Therefore let any equation $x^3-ax^2+axy-y^5=o$ be given, and ſubſtitute $x+\dot{x}o$ for x, and $y+\dot{y}o$ for y, and there will ariſe

$$\begin{aligned}&x^3+3\dot{x}ox^2+3\dot{x}^2oox+\dot{x}^3o^3\\&-ax^2-2a\dot{x}ox-a\dot{x}^2oo\\&+axy+a\dot{x}oy+a\dot{y}ox+a\dot{x}\dot{y}oo\\&-y^3-3\dot{y}oy^2-3\dot{y}^2ooy-\dot{y}^3o^3=o.\end{aligned}$$

Now by ſuppoſition $x^3-ax^2+axy-y^3=o$; which therefore being expung'd, and the remaining terms divided by o, there will remain $3\dot{x}x^2+3\dot{x}^2ox+\dot{x}^3oo-2a\dot{x}x-a\dot{x}^2o+a\dot{x}y+a\dot{y}x+a\dot{x}\dot{y}o-3\dot{y}y^2-3\dot{y}^2oy-\dot{y}^3oo=o$. But whereas o is ſuppos'd to be indefinitely little, that it may repreſent the moments of quantities, conſequently the terms that are multiplied by it, will be nothing in reſpect of the reſt: therefore I reject them, and there remains $3x^2\dot{x}-2a\dot{x}x+a\dot{x}y+a\dot{y}x-3\dot{y}y^2=o$, as above in Example 1.

Note that little 'oh' is the infinitesimal increase in time *dt*, and thus $\dot{v}o, \dot{x}o$, etc are what we call *dv,dx,…*

As we have said, for two millennia people have argued about whether philosophically it was ok to talk about infinitesimals *o*. But today we have computers and a large proportion of all integrations and differentiation are carried out numerically on a computer. For a computer, the problem of limits disappears! You always have only a finite amount of data, say $x(t_1), x(t_2),....,x(t_n)$ where $t_k = a + k\Delta t$ are equally spaced 'samples' of *x*. If you are integrating *x*(*t*) from *a* to t_n, for the computer, we might as well approximate:

$$\int_a^{t_n} x(t)dt \approx \sum_{k=1}^{k=n} x(t_k)\Delta t$$

Likewise, with the computer, you can always approximate the derivative

$$\dot{x}(t_k) \text{ or } \frac{dx}{dt}(t_k) \approx (x(t_k) - x(t_{k-1}))/\Delta t$$

What was the real impact of Newton's work on calculus? It was two fold. Firstly, he saw that Oresme's insight that the area under the graph of velocity was distance traveled was really a completely general fact. Namely, if for any function *x*(*t*), the two operations of (a) taking the area under the graph between *a* and *t*, and (b) taking its derivative were inverse to each other. This is the 'fundamental theorem of calculus'.

In its simplest form, imagine $x_1, x_2, x_3, \cdots$ is any sequence of numbers. Then we can do 2 things:

a) take "cumulative sums": $x_1, x_1 + x_2, x_1 + x_2 + x_3, \cdots$

or b) take differences: $x_2 - x_1, x_3 - x_2, x_4 - x_3, \cdots$

These processes are *inverse* to each other: start from the x's, call their cumulative sums y's and then take differences. You get back the x's:

$$\begin{array}{ll} \text{if } y_1 = x_1 & \text{then } x_1 = y_1 - 0 \\ y_2 = x_1 + x_2 & x_2 = y_2 - y_1 \\ y_3 = x_1 + x_2 + x_3 & x_3 = y_3 - y_2 \\ \text{etc.} & \text{etc.} \end{array}$$

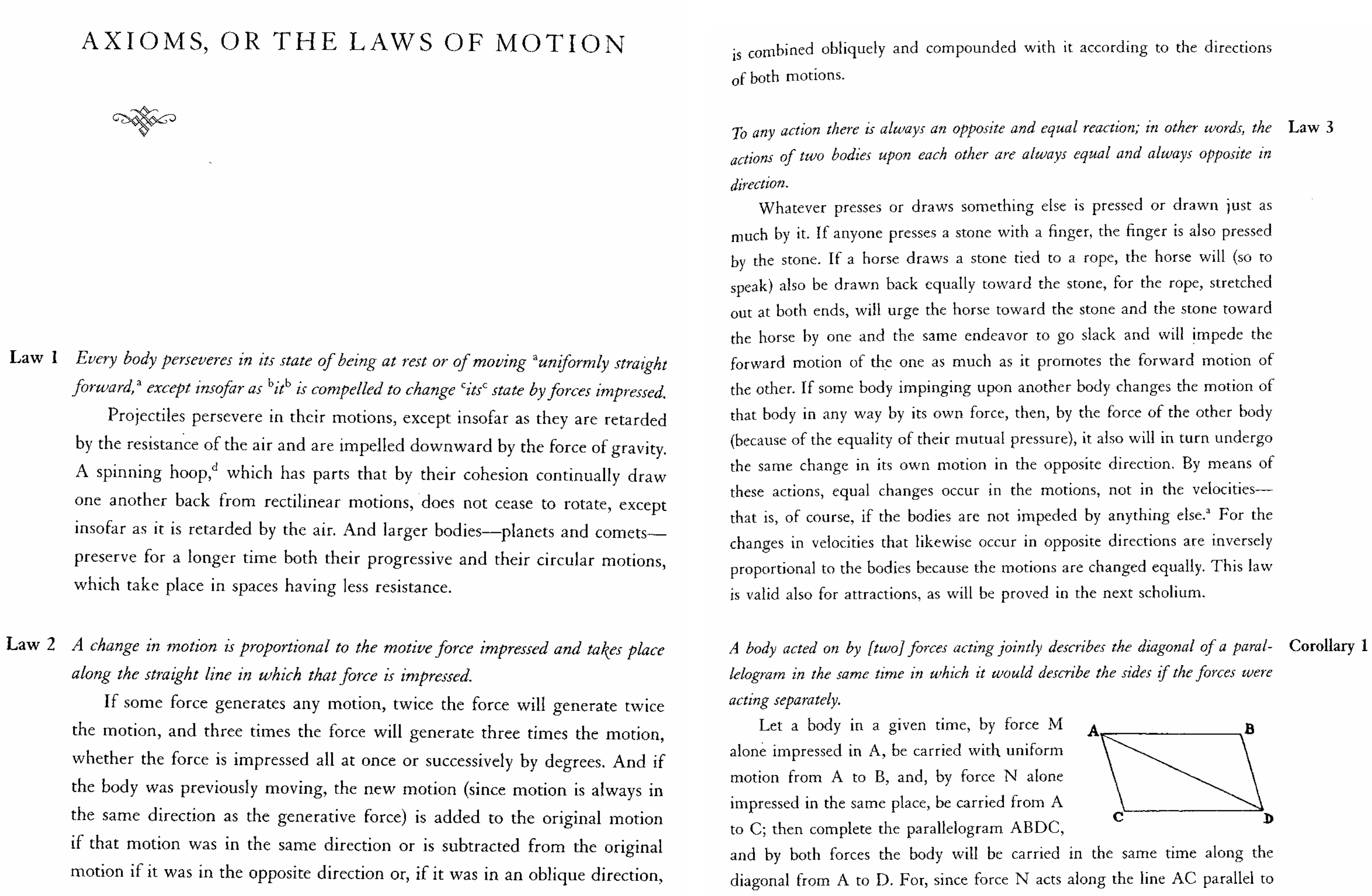
Or start from y's, call their differences x's and then take cumulative sums: you get back the y's. This is essentially the first column of the table below in which we have two *functions* $x(t)$, $y(t)$ and we describe 3 ways of constructing y from x or constructing x from y. In all cases one operation inverts the other:

$y(t_n) = (x(t_n) - x(t_{n-1}))/\Delta t$	$y(t)$ = slope of graph of $x(t)$	$y(t) = \dot{x}(t)$ or $\dfrac{dx}{dt}(t)$
$x(t_n) = \left(y(t_1) + y(t_2) + \cdots + y(t_n)\right)\Delta t + x(t_0)$	$x(t)$–$x(t_0)$ = area under graph of y between t and t_0	$x(t) = \int_{t_0}^{t} y(t)dt + x(t_0)$

On the left, we have the relationship between x and y shown in discrete terms for finite sequences of values, as in a computer, slightly modified from what we said before. It is immediate to verify that the top and bottom formulas in the middle invert each other. In the middle, we have the original geometric idea behind it. In Newton's language, if the area under a curve from a to t is considered a fluent, then its fluxion, the rate of change of the area as the left hand edge of the graph is moved, is just the height of the curve at t. On the right is the fundamental theorem as it is usually stated in calculus books. But simply stated the 2 operations of d/dt and $\int^t$ are inverse to each other. Starting from any function $x(t)$, we get a doubly infinite sequence of related functions $\cdots, \iiint x, \iint x, \int x, x, \dot{x}, \ddot{x}, \dddot{x}, \cdots$ (where we now use Newton's and the physicists notation of $\dot{x}$ for the derivative) or $\cdots, \iiint x, \iint x, \int x, x, \dfrac{dx}{dt}, \dfrac{d^2x}{dt^2}, \dfrac{d^3x}{dt^3}, \cdots$ (where we use Leibniz's notation *dx/dt*).

The second insight in Newton's work was that, whereas it is generally hard to evaluate areas and integrals, it is usually quite easy to evaluate derivatives. Thus if you work out the derivatives of a whole lexicon of functions, you can readily compile a table of integrals by seeking functions whose derivative is the function you want to integrate. Newton compiled such tables.

Enough of Newton's mathematics. Let's look at his physics. *Principia* begins with some definitions, in particular defining the phrase 'quantity of matter' to mean what we would call *mass* and 'quantity of motion' to mean what we would call *momentum*, i.e. the product of mass and velocity. The next part is where he states his three laws:



AXIOMS, OR THE LAWS OF MOTION

Law 1 *Every body perseveres in its state of being at rest or of moving* [a]*uniformly straight forward,*[a] *except insofar as* [b]*it*[b] *is compelled to change* [c]*its*[c] *state by forces impressed.*

Projectiles persevere in their motions, except insofar as they are retarded by the resistance of the air and are impelled downward by the force of gravity. A spinning hoop,[d] which has parts that by their cohesion continually draw one another back from rectilinear motions, does not cease to rotate, except insofar as it is retarded by the air. And larger bodies—planets and comets—preserve for a longer time both their progressive and their circular motions, which take place in spaces having less resistance.

Law 2 *A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed.*

If some force generates any motion, twice the force will generate twice the motion, and three times the force will generate three times the motion, whether the force is impressed all at once or successively by degrees. And if the body was previously moving, the new motion (since motion is always in the same direction as the generative force) is added to the original motion if that motion was in the same direction or is subtracted from the original motion if it was in the opposite direction or, if it was in an oblique direction, is combined obliquely and compounded with it according to the directions of both motions.

Law 3 *To any action there is always an opposite and equal reaction; in other words, the actions of two bodies upon each other are always equal and always opposite in direction.*

Whatever presses or draws something else is pressed or drawn just as much by it. If anyone presses a stone with a finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse will (so to speak) also be drawn back equally toward the stone, for the rope, stretched out at both ends, will urge the horse toward the stone and the stone toward the horse by one and the same endeavor to go slack and will impede the forward motion of the one as much as it promotes the forward motion of the other. If some body impinging upon another body changes the motion of that body in any way by its own force, then, by the force of the other body (because of the equality of their mutual pressure), it also will in turn undergo the same change in its own motion in the opposite direction. By means of these actions, equal changes occur in the motions, not in the velocities—that is, of course, if the bodies are not impeded by anything else.[a] For the changes in velocities that likewise occur in opposite directions are inversely proportional to the bodies because the motions are changed equally. This law is valid also for attractions, as will be proved in the next scholium.

Corollary 1 *A body acted on by [two] forces acting jointly describes the diagonal of a parallelogram in the same time in which it would describe the sides if the forces were acting separately.*

Let a body in a given time, by force M alone impressed in A, be carried with uniform motion from A to B, and, by force N alone impressed in the same place, be carried from A to C; then complete the parallelogram ABDC, and by both forces the body will be carried in the same time along the diagonal from A to D. For, since force N acts along the line AC parallel to

A mechanical system is here conceived as a set of bodies or parts of bodies each of which, by itself, would move at a constant velocity and in a constant direction, but which interact with each other through forces. Moreover, these forces have the effect of changing the motion (by which he means the mass times the velocity) of each body by an amount proportional to this force. Newton did not write this as a differential equation: as we saw, he felt at least in writing *Principia* that he must express his ideas as geometrically as possible to attain the same level of rigor as that of the ancient Greeks. In his own notes, as published in his book on fluxions, we do find differential equations. But oddly, nowhere do you find the second law above expressed in calculus terms. It is easy to do this, however. Suppose we have a system of *n* bodies, with positions (x_i, y_i, z_i) and mass m_i. Then their momentum or quantity of motion is just $(m_i\dot{x}_i, m_i\dot{y}_i, m_i\dot{z}_i)$ and the rate of which this changes is the force on each body: call this force (F_i, G_i, H_i). We can assume this force is some function depending on the positions and velocities of the other bodies and possibly on the time as well. Then we come up with some set of *n* equations of the form:

$$\boxed{\begin{aligned} m_i\ddot{x}_i &= F_i(x_1,...,z_n,\dot{x}_1,...,\dot{z}_n,t), \\ m_i\ddot{y}_i &= G_i(x_1,...,z_n,\dot{x}_1,...,\dot{z}_n,t), \\ m_i\ddot{z}_i &= H_i(x_1,...,z_n,\dot{x}_1,...,\dot{z}_n,t) \end{aligned}}$$

which say that the second derivative of the position of each body, i.e. its acceleration, is some function of the configuration of whole system, which determines the force. The third law constrains the forces and we won't worry about that here. These equations are what is usually meant by stating Newton's law as *F=ma*, force = mass x acceleration. In the next several Chapters, we will look at many examples of such laws and see what they do. In *Principia*, Newton did not restrict himself to systems of rigid bodies either, but applied these ideas to fluids and air, to infinite sets of variables as well as finite sets. In fact, very similar laws also apply to electricity and magnetism, hence to light. It is really not unreasonable to say that this framework and its natural generalizations are the universal framework for the laws of the physical universe.

The simplest case of these laws is when the force is a constant. Galileo had already understood this case. For a single body, if (*F,G,H*) is the vector of force acting on this body, its motion must satisfy:

$$\boxed{\begin{aligned} \ddot{x} &= F/m \\ \ddot{y} &= G/m \\ \ddot{z} &= H/m \end{aligned}}$$

As Galileo saw, if the acceleration is constant, the velocities must increase (or decrease) at a constant rate. Thus:

$$\begin{aligned} \dot{x}(t) &= (F/m)t + \dot{x}(0) \\ \dot{y}(t) &= (G/m)t + \dot{y}(0) \\ \dot{z}(t) &= (H/m)t + \dot{z}(0) \end{aligned}$$

and then the positions themselves must change quadratically:

$$\begin{aligned} x(t) &= (F/m)\frac{t^2}{2} + \dot{x}(0)t + x(0) \\ y(t) &= (G/m)\frac{t^2}{2} + \dot{y}(0)t + y(0) \\ z(t) &= (H/m)\frac{t^2}{2} + \dot{z}(0)t + z(0) \end{aligned}$$

Note that if there is no force, *F=G=H=0*, then the body moves in a straight line with constant speed. This was Newton's first law.

And if *z* is the vertical coordinate in 3-space, *F=G*=0 and *H* is a negative number representing gravity, then we are in Galileo's case of a projectile:

$$\begin{aligned} x(t) &= \dot{x}(0)t + x(0) \\ y(t) &= \dot{y}(0)t + y(0) \\ z(t) &= (H/m)\frac{t^2}{2} + \dot{z}(0)t + z(0) \end{aligned}$$

As before, we see that the horizontal position (*x,y*) moves in a straight line, but the altitude of the projectile reaches a maximum and then decreases, making its path into a parabola. But to agree with Galileo's central observation that objects of different masses fall at the same rate, note that the force *H* created by gravity must be proportional to the mass *m* of the projectile, *H*=*gm*, where *g* is the acceleration caused by gravity. Newton's amazing idea, which we will discuss in more detail when we look at planetary motion, was that any two bodies, with masses m_1 and m_2 attract each other with a force proportional to the product of their masses divided by the square of their distance *r* apart:

$$F = \frac{Gm_1m_2}{r^2}$$

where *G* is now a universal constant, the strength of all gravitational forces. In particular, *g* must be *G* times the mass of the earth divided by the square of the radius of the earth.

I want to end this Chapter by writing these in discrete computer friendly form, which show in exactly what sense they are a recipe for predicting the future. Imagine we sample time at discrete but very small intervals Δ*t*. To make our notation less of a mess, let's write x_i for all the variables (x_i, y_i, z_i) (so now there are 3*n* components to *x*, not just *n*) and F_i for all the forces. (In effect, we consider the 2nd and 3rd equation as cases of the 1st.) Let's write $x_i(k.\Delta t) = x_{i,k}$: in other words, we have a two-dimensional array of numbers *x* representing the position of the various parts of the mechanical system at various times. With discrete time, what happens to acceleration? Well

$$\text{velocity between } (k+1)\Delta t \text{ and } k\Delta t \approx \frac{x_{i,(k+1)} - x_{i,k}}{\Delta t},$$

$$\text{velocity between } k\Delta t \text{ and } (k-1)\Delta t \approx \frac{x_{i,k} - x_{i,(k-1)}}{\Delta t}, \text{ so}$$

$$\text{acceleration near } k\Delta t \approx \left(\frac{x_{i,(k+1)} - x_{i,k}}{\Delta t} - \frac{x_{i,k} - x_{i,(k-1)}}{\Delta t}\right)\Big/\Delta t \approx \frac{x_{i,(k+1)} - 2x_{i,k} + x_{i,(k-1)}}{\Delta t^2}$$

Substituting this into Newton's laws above, we get:

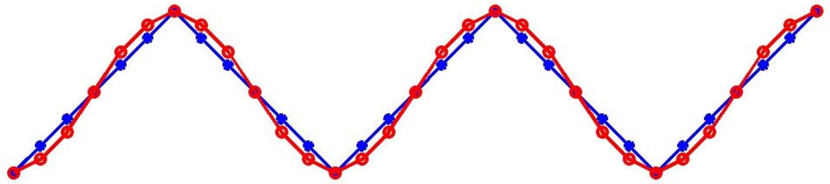
$$\boxed{\frac{x_{i,(k+1)} - 2x_{i,k} + x_{i,(k-1)}}{\Delta t^2} = F_i\left(x_{1,k}, \ldots, x_{n,k}, \frac{x_{1,k} - x_{1,k-1}}{\Delta t}, \ldots, \frac{x_{n,k} - x_{n,k-1}}{\Delta t}, k\right)}$$

which are explicit recipes for computing the whole vector of numbers $(x_{1,k+1}, \ldots, x_{n,k+1})$ representing the first step into the future, in terms of the present, $(x_{1,k}, \ldots, x_{n,k})$ and the immediate past $(x_{1,k-1}, \ldots, x_{n,k-1})$. This is powerful magic.

Chapter Six: Simple Harmonic Motion

Oscillations are a ubiquitous phenomena in the physical world. The tides rise and fall roughly twice daily, the weight at the end of a pendulum swings back and forth, sounds transmitted by the air turn out to oscillations in air pressure, you can feel the vibration of a guitar string, the water surface rises and falls with each passing wave, the seasons present a slow oscillation of mean temperature, people bounce on pogo sticks and trampolines. Producing controlled oscillations opened up the technology of clock building, as we have seen above.

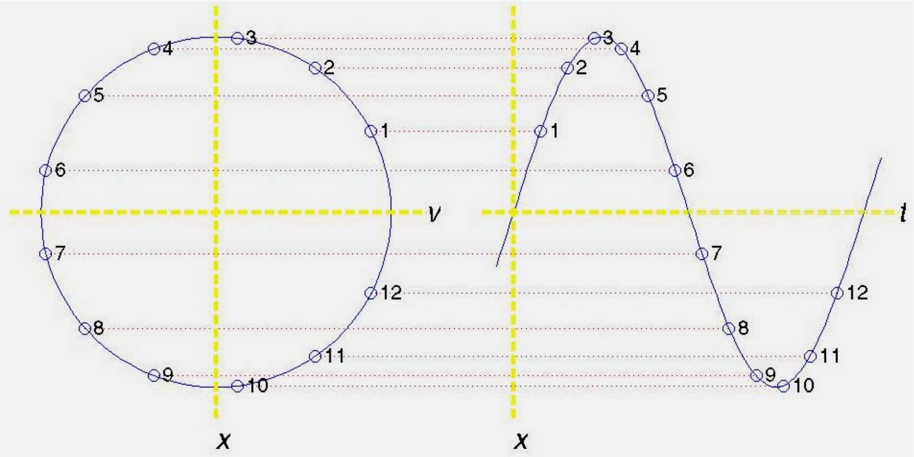
But how are oscillations to be modeled mathematically? The Babylonians encountered many astronomical situations in which some aspect of solar, lunar and planetary positions and velocities appeared to increase and decrease periodically. They simply used *zigzag* functions to model these oscillations: functions which increased at a constant rate, then instantly reversed themselves and decreased at a constant rate. But this was wrong. For example, there is a rule of thumb for the raise and fall of the tides which is remembered as "1,2,3,3,2,1". This means, if the tide raises a certain amount in the first hour after low tide, it will rise twice as much in the second hour, 3 times as much in the third, 3 times as much in the fourth, twice as much in the fifth and the same amount the last hour before high tide (there are about 6 hours between low and high tide). We can make a little graph contrasting these two. The blue with stars is the Babylonian zigzag and the red with circles is the more realistic tidal model.



It is curious that Oresme never considered oscillating qualities as a special case of difformly difform qualities – perhaps because he was always thinking of positive qualities and it is most natural to suppose that an oscillating function is equally often positive and negative. Galileo was fascinated by oscllations as we have seen: he studied the pendulum and the tides. But he never formulated any law for their motion.

It is a probably a much older idea but the first reference I can find to the correct mathematical model occurs in 1585. in the work of the obscure Venetian mathematician Giovanni Benedetti. He wrote that suppose you start with a body moving at a constant rate around a circle. Then you gradually tilt the circle backwards until you are viewing it sideways: now the visible motion of the body is simply an oscillation, first left then right etc. His main point was to illustrate how unphysical was Aristotle's idea that a body could not reverse the direction of its motion without going through an intermediate stage of rest. Any point in the circular motion would become the point of rest if it was viewed from the right angle! If he had only pursued this, he would have been led to describe simple harmonic motion.

Below is a figure which describes Benedetti's idea:



On the left is uniform circular motion shown by sample points 1 through 12, with angles $2\pi k/13$. On the right, we take only their vertical coordinate (here shown as x) but now plot it against time t. We get a fine looking oscillation. Once we have Benedetti's idea, we immediately get a formula. To go around a circle of radius r at uniform speed, we let the angle be a linear function of time. The most general formula would be:

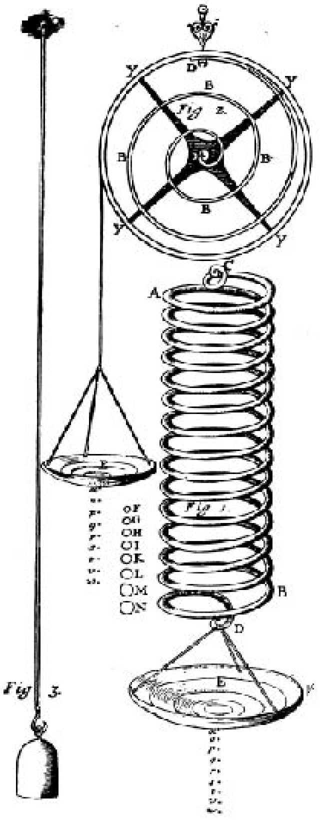
$$\left(r.\cos(a.t+b),\ r.\sin(a.t+c)\right)$$

Then taking one coordinate, we get:

$$x = r.\cos(a.t+b)$$

In other words, the trigonometric functions sine and cosine, which had already been tabulated by Ptolemy for astronomical calculations, now emerge as the *fluents*, the functions which most naturally describe oscillations.

The last ingredient in this historical narrative was the analysis by Hooke of the physics of a spring. Hooke was a generation older than Newton, a great experimentalist but quite weak as a mathematician. He somehow became one of Newton's enemies and Newton refused to join the Royal Society until Hooke died, at which point, Newton accepted the Presidency and remained President for the rest of his life.



Hooke's illustration of various oscillating systems: a pendulum, coil spring and ordinary spring

Hooke seems to have been the first to hit upon a truly simple oscillating system that could be readily analyzed: an oscillating spring. He realized that a very simple thing happens with a spring: the farther you stretch it, the stronger it pulls back and more you compress it, the more it pushes back. You can check this easily by hanging heavier and heavier weights on the spring and noting that it stretches linearly with the weight (up to some limit). He was so pleased with this, he published it as an *anagram*!

<u>`ceiiinosssttuv`</u>

What he had in mind was the Latin phrase:

<u>`Ut tensio sic vis`</u> (As the stretch, so the force)

But his mathematical theory of the spring was totally wrong. The right way to put his Latin into mathematics comes from using Newton's laws. If *x*=0 is the resting position of the spring, then we may say that the spring exerts a force proportional to *x* with a negative coefficient. Using Newton's laws, we should have the differential equation:

$$\boxed{m\ddot{x} = F = -cx, \text{ some } c > 0}$$

Thus if *x>0*, then the force seeks to decrease *x*, while if *x<0*, then the force seeks to increase *x* – this is called a restoring force.

To make the link with the circular model of Benedetti, we need to check that all solutions of this are given by *sines* and *cosines*. What we need are the two rules:

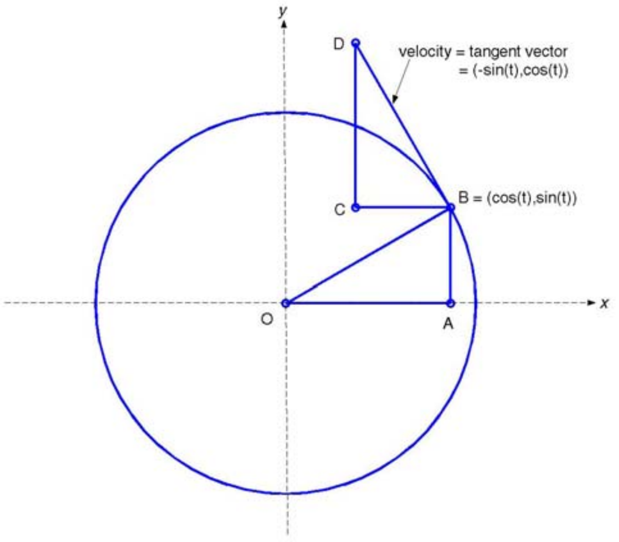
$$\boxed{\begin{array}{l} \dfrac{d}{dt}\sin(t) = \cos(t) \\ \dfrac{d}{dt}\cos(t) = -\sin(t) \end{array}}$$

You probably know these rules, but they are so important that I want to show where they come from again – in 2 ways, one algebraic and one geometric.

First, here is a geometric proof. In the figure, imagine a point moving at constant speed counter-clockwise around the unit circle. Its position is given by $x = \cos(t)$, $y = \sin(t)$ as is seen in the figure. When *t*=0, it is on the positive *x*-axis, then it moves up and gradually to the left and at *t*=π/2 (in radians!), it hits the *y*-axis. Its velocity vector is the tangent line to the circle, as shown in the figure. Now let's do Euclidean style geometry: the triangles *OAB* and *DCB* are congruent. In fact, if we rotate *OAB* through 90 degrees around *B*, it becomes *DCB*. Thus

$$\overline{OA} = \cos(t) = \overline{CD} \text{ and}$$

$$\overline{\mathrm{AB}} = \sin(t) = \overline{CB}$$



Note that as a vector, *BC* is pointing backwards. This proves that the tangent vector to the unit circle is just (–sin(*t*),cos(*t*)), hence we get the derivative rules in the box above.

Second, here is an algebraic proof. You need to recall from trig the basic addition formulas for the sine and cosine of the sum of two angles:

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b)$$

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$$

Then we can apply the limit definition of derivative:

$$\begin{aligned}\frac{d}{dt}\sin(t) &= \lim_{\Delta t\to 0}\frac{\sin(t+\Delta t)-\sin(t)}{\Delta t}\\ &= \lim_{\Delta t\to 0}\frac{\sin(t)\cos(\Delta t)+\cos(t)\sin(\Delta t)-\sin(t)}{\Delta t}\\ &= \sin(t)\cdot\lim_{\Delta t\to 0}\frac{\cos(\Delta t)-1}{\Delta t}+\cos(t)\cdot\lim_{\Delta t\to 0}\frac{\sin(\Delta t)}{\Delta t}\end{aligned}$$

But by looking at little slivers of triangles in the Chapter on Archimedes, we saw that the first limit is 0 and the second is 1. So the derivative of sin is cos! The formula for the derivative of cos comes out the same way using the addition formula for cos.

Now we know the solutions of the differential equation: $\ddot{x} = -x$. *x*=cos(*t*) solves the equation, as does *x*=sin(*t*). So actually any function

$$x = A.\cos(t) + B.\sin(t)$$

solves the equation too. Because the equation has second derivatives, it's a standard fact that there can be at most two unknowns in the most general solution, so all solutions have this form. In general there are constants *m* and *c*. Let $c/m=a^2$, so our equation is $\ddot{x} = -a^2x$. This means the acceleration is scaled up or down and this means we have the same form of solution only time must run faster or slower. In fact, if we change the rate of time by the factor *a* and look at what happens if we set:

$$x = A.\cos(a.t) + B.\sin(a.t)$$

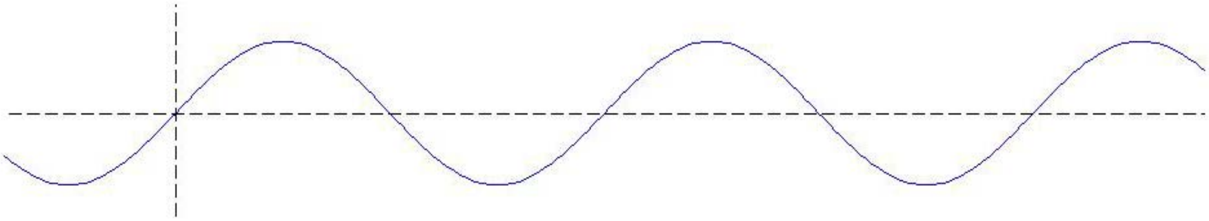
we get the solution again: each derivative of a scaled up function *f*(*a.t*) is *a* times the derivative of *f*(*t*), so taking the derivative twice, we get the required factor a^2.

Trig functions can be manipulated in various ways and it's important to pause a bit and rewrite this sum of a sine and a cosine in a way that one can see what its graph looks like. We use a little trick: write *A* and *B* in polar coordinates!, i.e.: *A*=*C*.sin(*b*), *B*=*C*.cos(*b*). Then:

$$x = C.\sin(b).\cos(a.t) + C.\cos(b).\sin(a.t) = C.\sin(at+b)$$

using the formula for the sine of a sum of two angles we mentioned above. (We could have written it with cosine instead.) Now here *C* is clearly the maximum value of *x* and –*C* is the minimum. *C* is called the *amplitude* of the oscillation. The constant *b* just shifts the oscillation in time: it is called the *phase*.

What do the all these solutions look like? They are called *sinusoidal oscillations* and they all look like this (up to shifts and stretches in the horizontal *t*-axis and stretches in the vertical *x*-axis):



These are the universal graphs for the simplest oscillations, those governed by Hooke's law and which are known as *simple harmonic motion*.

Two numbers are very important in describing simple harmonic motion: the *period* and the *frequency*. They are inverse to each other. The period *p* is simply the length of time needed for the system to return to its starting point. This is 2π/*a* because:

$$C \cdot \sin(a(t + 2\pi / a) + b) = C \cdot \sin(at + b + 2\pi) \equiv C \cdot \sin(at + b)$$

(Recall that 2π radians is 360 degrees, which is when sine and cosine repeat.) The frequency *f* is the number of times (or fractions thereof) in which the motion repeats its cycle in each unit of time. If time *t* is measured in seconds, the frequency is measured in repeats per second, which are called *hertz*. The frequency is inverse to the period:

$$f = 1 / p = a / 2\pi$$

Because the frequency *f* is more important than the constant *e*, we often write harmonic motion as:

$$x(t) = C \cdot \sin(2\pi f \cdot t + D)$$

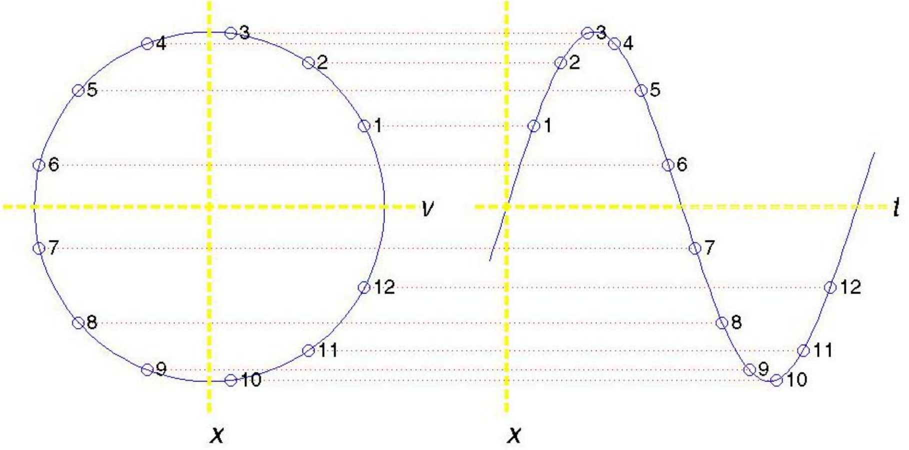
A historical note: the first place where I have found this curve described is in the 1634 book *Traite des Indivisibles* by the fairly obscure French mathematician Roberval, where it is called the *companion of the cycloid*. Here it is only a step to the construction of a much less important curve, the cycloid, the curve traced by a point on the circumference of a wheel rolling on a plane, and its real importance is missed.

This curve is very close to the odd rule of thumb curve given by the increments (1,2,3,3,2,1). In fact, divide 180 degrees into 6 equal steps, starting at –90 degrees (where sin = –1) and going to +90 (where sin=+1) in increments of 30 degrees. Then the sine function has values:

$$\begin{aligned}
&\sin(-\pi / 2) = -1 = \quad -6/6,\\
&\sin(-\pi / 3) = -\sqrt{3}/2 \approx -5/6,\\
&\sin(-\pi / 6) = -1/2 = -3/6,\\
&\sin(0) = 0 = \qquad\quad 0/6,\\
&\sin(+\pi / 6) = +1/2 = +3/6.\\
&\sin(+\pi / 3) = +\sqrt{3}/2 \approx +5/6,\\
&\sin(+\pi / 2) = +1 = \quad +6/6
\end{aligned}$$

which advance very nearly by (1,2,3,3,2,1)/6.

Benedetti passed from uniform circular motion to simple harmonic motion. But it's just as simple to find uniform circular motion implicit in simple harmonic motion. Suppose x satisfies $\ddot{x} = -x$. Then you consider both the position x and the velocity $v = \dot{x}$ as two functions of time, two fluents in Newton's language and make a *two-dimensional plot* of the points (x,v). As time progresses, this point moves around in the plane and, lo and behold, it moves uniformly around a circle! As x increases to its maximum, v decreases to 0 and as x now goes back to 0, v becomes negative and moves to a negative minimum. This is seen in the same figure we used before, where the (x,v) plot is on the left (n.b. the vertical axis is x and the horizontal axis is v) and the graph of x alone is on the right.



Newton, as you would expect, worked out simple harmonic motion on his way to deriving his theory of planetary motion. He was concerned with all laws whereby a body in the center of the plane (or space) attracts another movable body, but with a force which varies depending on far apart they are. He was thinking of the sun and the earth and the 'inverse square law' of gravity, but a much simpler case was when the attraction went up linearly with the distance, becoming greater when the bodies were further apart, which is just Hooke's law that we have been studying.

Problem: We can use simple harmonic motion as an opportunity to explore how well computers can approximate the exact solutions to differential equations. In the computer, derivatives like *dx/dt* are replaced by the approximation *Δx/Δt* and the best we can hope for is that we get values for $x(t)$ close to the correct ones. Suppose we use the recipe:

$$\boxed{\frac{x_{i,(k+1)} - 2x_{i,k} + x_{i,(k-1)}}{\Delta t^2} = F_i\left(x_{1,k}, \ldots, x_{n,k}, \frac{x_{1,k} - x_{1,k-1}}{\Delta t}, \ldots, \frac{x_{n,k} - x_{n,k-1}}{\Delta t}, k\right)}$$

There's only one x_i, which we call x, and this is to be sampled at a sequence of discrete times, i.e. $x_k = x(k\Delta t)$. The above equation, then, reduces to:

$$\boxed{x_{k+1} = 2x_k - x_{k-1} - \Delta t^2 (g/L) x_k}$$

Take g/L=1 for simplicity and start with x_0 = 0, x_1 = Δt (approximating the assumption that the initial velocity $\dot{x}(0)=1$) and solve this with various Δt's (say .5, .25, .1, .05). You want at least the solution for $0 \le t \le 4\pi$, so you need all k's until $k\Delta t \ge 4\pi$. Plot the sequence of values x_k against k and also plot x_k against the estimated velocity $(x_{k+1} - x_k)/\Delta t$. Find the exact solution as a special case of $x(t) = A.\sin(t) + B.\cos(t)$ and superimpose on your plots the plot for the exact solution say with dashed lines. Compare them: what are you finding? Find the maximum of the absolute value of the error for each Δt.

Now let's do it another way. Introduce explicitly a second set of variables for the estimated velocities v_k. Now make our update rule:

$$\boxed{\begin{aligned} x_{k+1} &= x_k + \Delta t \cdot v_{k,} \\ v_{k+1} &= v_k - \Delta t \cdot x_k \end{aligned}}$$

Here we are updating the x's using the estimated velocity and updating the velocities using the acceleration calculated from the estimated position. Carry out the same calculations and plots. Take Δt = .25 and plot your calculation up to k = 250. Something quite different has happened! Doing things approximately can be tricky.