Chords, Tangents and the Leibniz Notation

David Tall

Mathematics Education Research Centre
Warwick University
COVENTRY CV4 7AL

In this article I continue my quest for "understanding the calculus" [1,2] by looking at a practical approach to the notion of a tangent and linking it to the Leibniz notation dy/dx in a meaningful way. The latter is a *bête noire* for students: it *looks* like a quotient, it *acts* like a quotient, yet the seeds of a classic psychological conflict are sown in their minds when they are told *it must not be thought of as a quotient*. I shall discuss how this conflict may be resolved so that the chain law allows cancellation.

Practical tangents

A number of computer programs for drawing graphs claim to draw tangents to curves, including the *MEI programs for Mathematical Computing* [3] and my own program *SuperZoom* in the Supergraph package [4]. None of the programs do anything of the sort. When requested to draw the tangent to y=f(x) at the point x=a, they actually draw the straight line through two close points on the curve, (a,f(a)) and (a+s,f(a+s)) where s is small. *SuperZoom* uses s=0.0001. This 'practical tangent' is really nothing more than a secant (or extended chord) through two close points on the graph.

The arithmetic accuracy on computers usually allows the practical tangent to be calculated with sufficient precision to satisfy the limited requirements of the visual display. It certainly works satisfactorily for most standard functions met in the sixth form, as the picture of the tangent to y=e^x at x=1 shows. (Figure 1.)

Published in *Mathematics Teaching*, (1985), **112** 48–52.

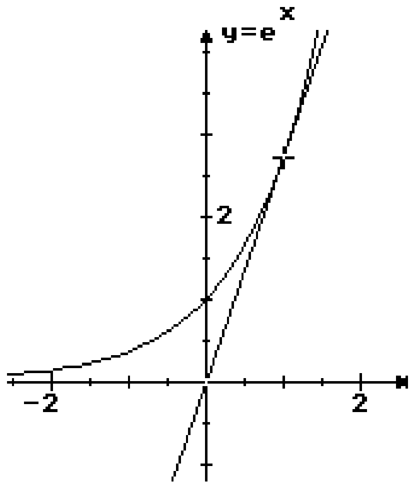

Figure 1 : A practical tangent to $y=e^x$ at $x=1$

However, it may fail dismally when the curve has tiny wrinkles or corners. The function f(x)=x+abs(1-x^2) has ‘corners’ at x=–1 and x=1. The ‘practical tangent’ drawn at x=1 plots the line through (1,f(1)), (1.0001,f(1.0001)), giving a line that seems to touch the graph only to the right of the point concerned. (Figure 2.)

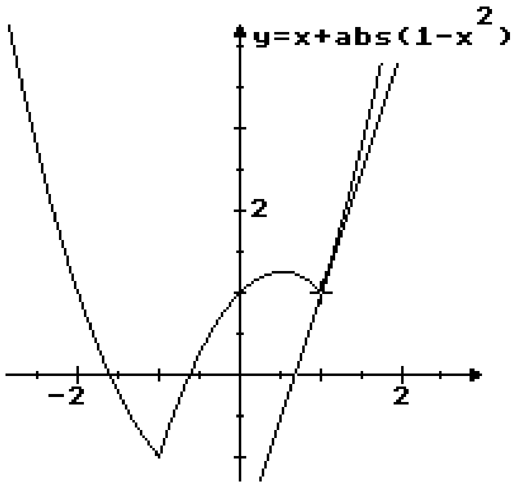

Figure 2: A ‘practical tangent’ at a corner

SuperZoom has a very flexible line-plotting routine that allows the line between two specified points to be drawn. By taking points on the graph with x=1 and x=0.99999, the line drawn gives a "practical tangent" that seems to touch the curve to the left. (Figure 3.)

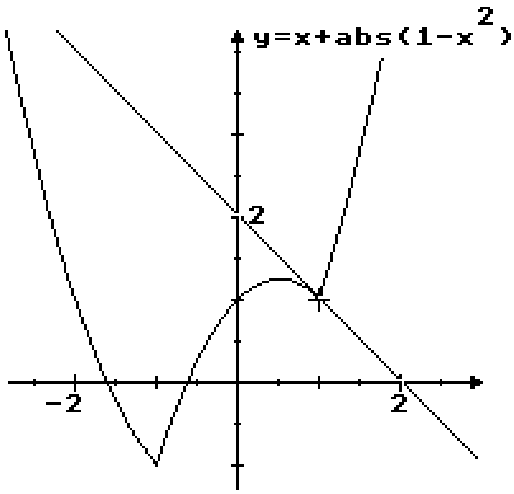

Figure 3: Another ‘practical tangent’ from x=0.9999 to x=1

These are not the only possibilities, one may investigate what happens in this case when the extended chord is drawn through the points with x-coordinates $1–h$ and $1+h$. For instance the line through the points on the graph with x=1-1/10000 and x=1+1/10000 looks as if it ‘balances’ on the corner. It even seductively passes through the other corner on the graph, supporting its candidacy as a genuine tangent. (Figure 4.)

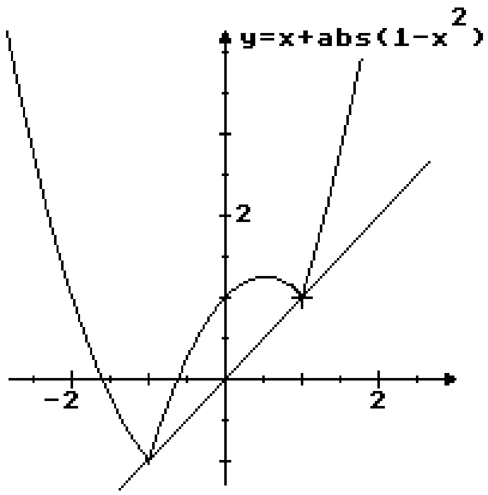

Figure 4: A ‘balance tangent’ from x=1–1/10000 to x=1+1/10000

Lines that ‘touch’ a curve

Some textbooks do not give a definition of the tangent, preferring to use the intuitive idea that a tangent is a line that ‘touches’ the curve. If a graph has a ‘corner’, students may believe that it has an infinite number of tangents there. Thus the graph $y=x+\text{abs}(1–x^2)$ may be thought to have an infinite number of tangents at x=–1 and

x=1. But here the graph magnifies up to look like two half-lines meeting at an angle with different left and right gradients.

The situation at a point where the graph magnifies to look straight is quite different. Using *SuperZoom* to draw the curve $f(x)=x^2$ magnified through the point x=1/2, y=1 and superimposing the tangent $y=x+1/4$ reveals the graph and tangent (almost) indistinguishable within the error of drawing (figure 5).

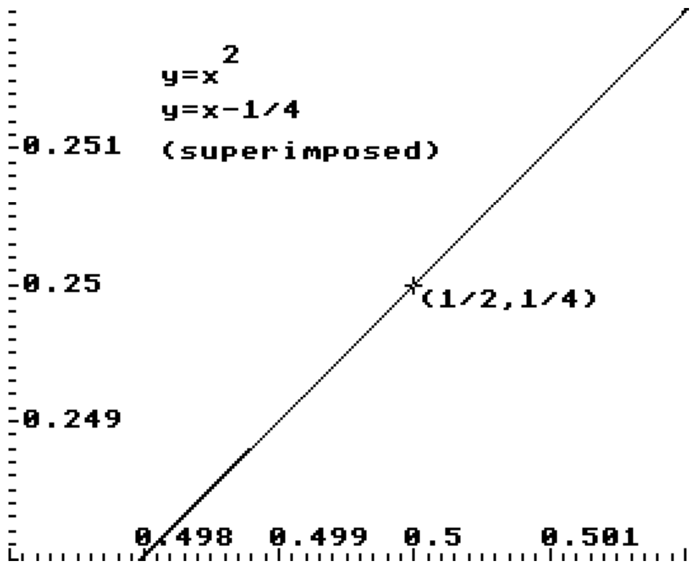

Figure 5 : High magnification of a graph and the tangent at a point

If a graph has a tangent, under high magnification a small part of the graph and the tangent are practically indistinguishable.

It is my experience that students need guidance over this point. In early trials with Graphic Calculus, we found that students easily appreciated the fact that a curve had a gradient at those points where it magnified to look straight. But without explicit discussion over the links with the tangent, it was still possible for them to believe that a graph with a corner had no gradient there, yet it had many tangents.

It is appropriate to see the three ideas gradient, tangent and derivative operating in parallel, so that a graph has a gradient at a point if and only if it has a tangent at that point, in which case the derivative equals the gradient.

The Leibniz Notation

The notation used by Leibniz in his original paper[6] for the gradient of a curve is dy/dx. The symbols δx, δy for increments in x,y respectively came into prominent use over a century later in a textbook by the English mathematician Woodhouse[7]. We have absorbed these into our modern culture by using dx to represent any increment in x and denoting the corresponding increment in y=f(x) as

$$\delta y=\mathrm{f}(x+\mathrm{d}x)-\mathrm{f}(x).$$

(Figure 6.)

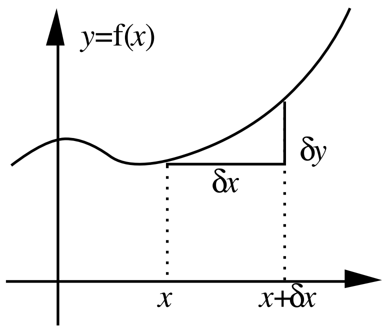

Figure 6: The Woodhouse notation for increments

The gradient of the chord from (x,y) to (x+δx,y+δy) is then δy/δx. As δx gets small, if the gradient δy/δx tends to a fixed limit, we denote the latter by dy/dx and say:

$$\text{as } \delta \mathrm{x} \text{ tends to 0, so } \frac{\delta y}{\delta x} \text{ tends to } \frac{\mathrm{d}y}{\mathrm{d}x}\ ,$$

or

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x}\ .$$

But dy/dx is no longer interpreted in its historical sense. For instance, Geoffrey Matthews writes[9] (page 9):

> dy/dx is simply a notation, signifying the gradient of the curve in question. It is not to be considered here as a ratio, as δy/δx is, but just a handy way of expressing 'the limit as δx→0 of δy/δx'.

This is firmly supported by Hilary Shuard and Hugh Neill in their excellent book on *Teaching the Calculus* [10] (page 13):

> The student ... has to learn that, in spite of all the evidence to the contrary, which seems to him to build up from statements such as
>
> $$\frac{\mathrm{d}y}{\mathrm{d}x} \times \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}t}$$
>
> dy/dx is not a symbol for a fraction, but for the limit of the gradient of a chord.

It is expressed even more forcefully in one of the earlier versions of SMP Advanced Mathematics [11] (page 221):

> 'dy/dx' must, at least for some considerable time, be regarded as an inseparable whole, just as 'δx' is. It does not in any simple or straight-forward way mean anything like 'dy divided by dx', and a statement such as '$dy/dx \times dx/dt = dy/dt$, by cancelling dx' is just so much gibberish.

The reader may very well agree with the substance of these expressed views. Yet they must severely tax the patience of students when dy/dx patently seems to work as a quotient and is later used as a quotient to solve differential equations by separation of the variables.

Tony Orton suggests [12]:

> Elaborate symbolism might only serve to confuse the issue. Perhaps examination syllabuses have made the mistake of demanding the use of dy/dx too soon.

Removing the dy/dx notation from the beginning of the course is a helpful move, but it serves no long-term purpose if the difficulties remain unresolved at a later stage. The resolution is found by looking carefully at the dy/dx notation to find out why it works in the way it does, and to see if it can be given a meaningful interpretation as a quotient.

The original definition of Leibniz

Leibniz is often misquoted as introducing the notation dy/dx as a quotient of infinitesimals. It is not true. The expression dy/dx is initially considered a quotient of finite quantities. In the first publication on the calculus [7] in 1684 he referred to a diagram which I have simplified in this article by referring only to the standard variables x,y. (Figure 7.)

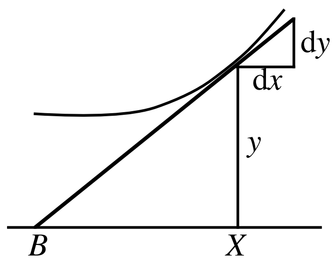

Figure 7 : The Leibniz definition of dx and dy

The curve represents a variable y depending on x, and B is the point where the tangent to the curve meets the x-axis.

Condensing what Leibniz said to concentrate on the variables x,y we get the statement:

> Jam recta aliqua pro arbitrio assumta vocetur dx, & recta quae sit ad dx ut y est ad XB vocetur dy.

which translates to

> Now some straight line selected arbitrarily is called dx, and the line which is to dx as y is to XB is called dy.

Thus the length dx is arbitrary and the length dy is the corresponding increment in y such that the quotient dy/dx equals y/XB. Disentangling the definition, we see that dx is any increment and dy is the corresponding increment to the tangent. (Figure 8.)

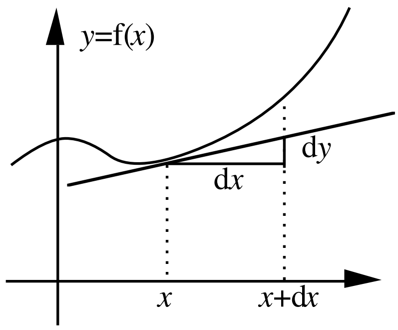

Figure 8 : The differentials of Leibniz as increments to the tangent

There is no mention of infinitesimals: they came later in the paper when Leibniz had to develop a method of calculating the direction of the tangent. Today we (usually) calculate the tangent direction by a limiting process, but there is no reason why we should not use the Leibniz notation in its original meaning.

Suppose the derivative f'(x) is known and dx is any real number, then we may follow the standard practice in many modern texts [13] to define

$$\mathrm{d}y=\mathrm{f}'(x)\mathrm{d}x$$

and (for dx≠0) obtain

$$\mathrm{d}y/\mathrm{d}x=\mathrm{f}'(x)$$

as a quotient of lengths.

Thus δx and δy are increments in x,y to the graph, whilst dx and dy are increments to the tangent. Both $\delta y/\delta x$ and dy/dx are quotients in exactly the same way.

What is interesting is what happens when we look at tiny increments under a microscope. As the tangent is then practically indistinguishable from the curve, taking dx=δx, we then find dy≈δy (figure 9).

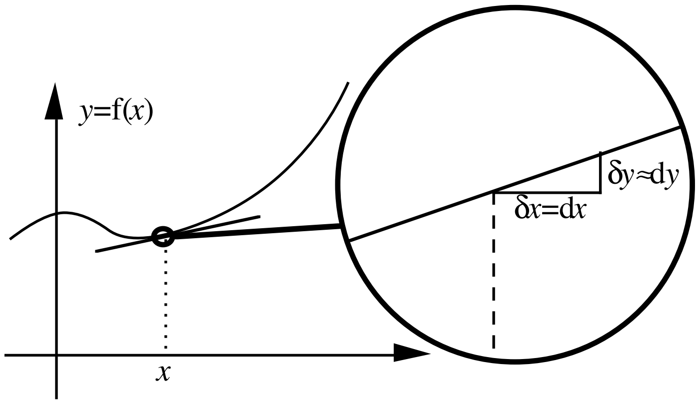

Figure 9: Magnifying a tiny locally straight part of a graph

Because dy=f'(x)dx, this gives

$$\delta y \approx f'(x)\delta x,$$

which is the usual formula for approximations interpreted visually.

The tangent vector

Using the given values dx,dy, a point (x+r,y+s) on the tangent must satisfy

$$s/r = \mathrm{d}y/\mathrm{d}x.$$

(Figure 10.)

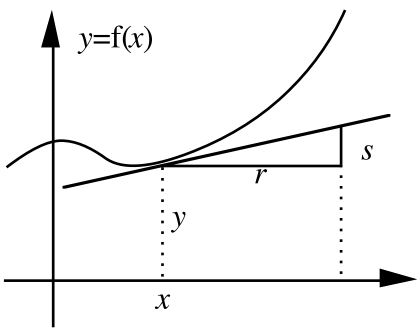

Figure 10: The tangent vector

If we take r=k.dx, then s=k.dy, so that the point on the tangent is

$(x+k.\mathrm{d}x,\ y+k.\mathrm{d}y)$.

Writing this in the form

$(x+k.\mathrm{d}x,\ y+k.\mathrm{d}y) = (x,y) + k(\mathrm{d}x,\mathrm{d}y)$

we see that every point on the tangent is at a vector displacement $k(\mathrm{d}x,\mathrm{d}y)$ from the point (x,y) on the curve. The tangent vector is therefore in the direction $(\mathrm{d}x,\mathrm{d}y)$.

Vertical tangents

Certain curves, such as $y=x^{1/3}$ at the origin, have vertical tangents. (To get the computer to calculate the cube root for negative x, it may be necessary to type in the cube root of the positive value abs$x$, then multiply the result by the sign, sgn$x$.) (Figure 11.)

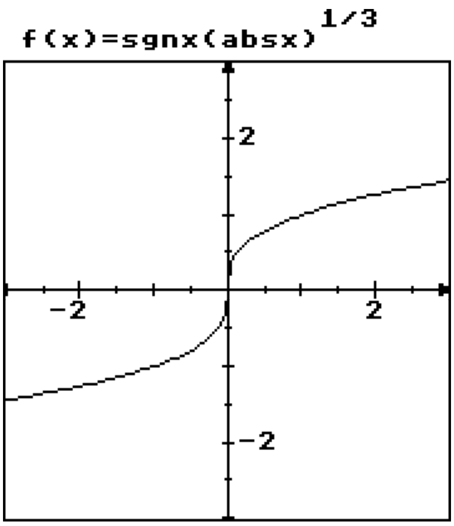

Figure 11 : The vertical tangent to $y=\sqrt[3]{x}$ at the origin

In such a case it is quite legitimate to take $\mathrm{d}x=0$, $\mathrm{d}y=1$ to get a tangent direction (0,1) along the y-axis. A point on the tangent at (0,0) is then of the form

$(0,0)+k(\mathrm{d}x,\mathrm{d}y)$

$= (k.0,k.1)$

$= (0,k)$,

which is simply a point on the y-axis, as expected.

If we refer to the tangent as a vector, we may include the anomalous case where the tangent is vertical. I am not too bothered whether we say that the gradient $\mathrm{d}y/\mathrm{d}x$ ‘does not exist' or that it is 'infinite', though the latter has the advantage that every tangent then has a corresponding gradient. The case is worth discussing because it aligns vertical tangents with others where a tiny portion of the graph looks straight under magnification. Whichever convention we adopt, students may meet either in other contexts; it is as well for them to know that some things in mathematics are a matter of individual opinion. If you don't believe this, draw the graph of y=sqr(abs$x$)) (figure 12). Do you think this has a tangent at the origin? Some mathematicians think so, but it doesn't magnify to look like a straight line ...

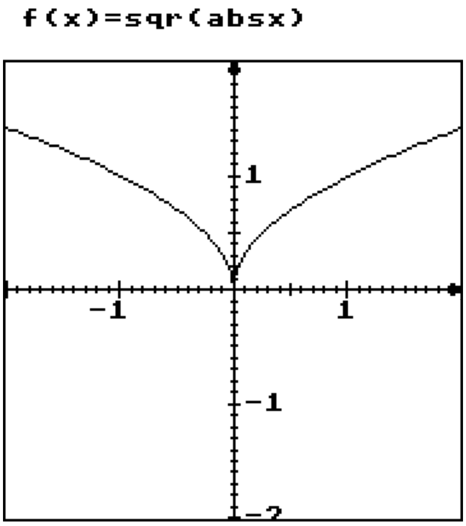

Figure 12: The graph of $y=\sqrt{|x|}$ – does it have a tangent at the origin?

Three dimensions

The picture in three dimensions is not fundamentally different from two. Given two functions x=f(t), y=g(t), then, as x varies, the point (t,f(t),g(t)) describes a curve in three dimensional space. The projection onto the first two coordinates (t,f(t)) gives the graph of x=f(t), with a similar picture for y=g(t) on the (t,y) plane. If the curve has a tangent at a point P=(t,f(t),g(t)) on the curve may be considered as the diagonal of a rectangular box (figure 13).

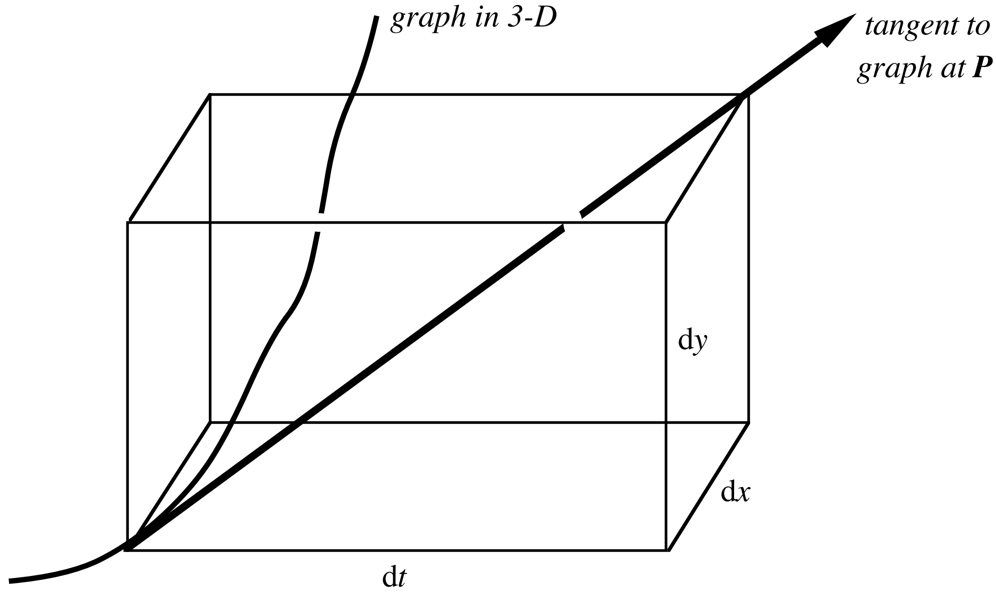

Figure 13: A tangent to a curve in (t,x,y) space

Here the sides are denoted by (dt,dx,dy) because the projection down onto a coordinate plane gives the picture of the tangent to the curve in two dimensions, using the Leibniz notation (figure 14).

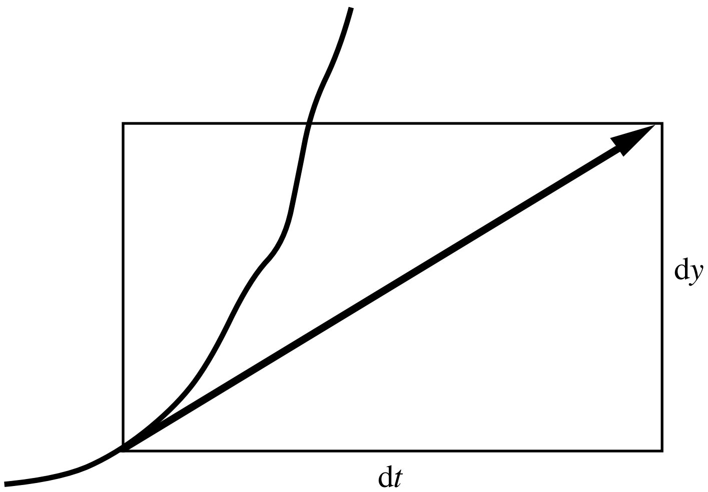

Figure 14: The projection onto the t-x plane

The chain rule

If x=f(t) is a function of t and y=h(x) is a function of x, then writing g(t)=h(f(t)) expresses y=g(t) as a function of t. In three dimensional (t,x,y) space the graph (t,f(t),g(t)) is a curve and the components of the tangent vector are dt,dx,dy. Provided that dt and dx are not zero, in each coordinate plane the gradient of the tangent is given by a quotient: $\mathrm{d}x/\mathrm{d}t$ in the x-t plane, $\mathrm{d}y/\mathrm{d}t$ in the y-t plane, and $\mathrm{d}y/\mathrm{d}x$ in the x-y plane. Thus the equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} \times \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}t}$$

is true as quotients of lengths. The lengths are just the components of the tangent vector in three-dimensional space.

The one place where this argument breaks down is if $\mathrm{d}t$ or $\mathrm{d}x$ were zero. Now we can choose $\mathrm{d}x$ to be anything we like, so we can take $\mathrm{d}x \neq 0$. But then $\mathrm{d}t$ is determined by the equation

$$\mathrm{d}x=\mathrm{f}'(t)\mathrm{d}t.$$

If $\mathrm{f}'(t)=0$ then we must have $\mathrm{d}x=0$. But now

$$\mathrm{d}y=\mathrm{g}'(t)\mathrm{d}t,$$

so we must have $\mathrm{d}y=0$ also, whence

$$\mathrm{d}y/\mathrm{d}x=0$$

and the chain rule is true because both sides are zero.

Implicit curves

The functions $x=\mathrm{f}(t)$, $y=\mathrm{g}(t)$ give a curve $(\mathrm{f}(t),\mathrm{g}(t))$ in the x-y plane. If the three-dimensional curve $(t,\mathrm{f}(t),\mathrm{g}(t))$ has a tangent vector, then, for an increment $\mathrm{d}t$ in t, we obtain the other components $\mathrm{d}x=\mathrm{f}'(t)\mathrm{d}t$ and $\mathrm{d}y=\mathrm{g}'(y)\mathrm{d}t$ of the three-dimensional tangent vector and the tangent to the projection in the x-y plane is in the direction $(\mathrm{d}x,\mathrm{d}y)$. The curve in the x-y plane need not simplify to give y as a function of x. For instance, when

$$x=\sin t,\ y=\cos t$$

then the relationship between x and y is the implicit relation:

$$x^2+y^2=1.$$

(Figure 15.)

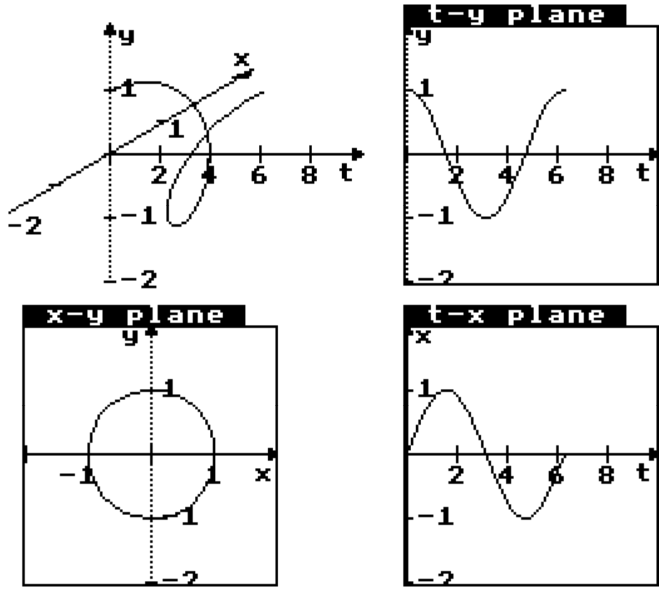

Figure 15 : $x = \sin t$, $y = \cos t$ drawn in 3D and projected onto the three coordinate planes

In the computer drawing I have given a 3-dimensional view with the *t*-*x* plane horizontal and the other planes vertical. As the curve is drawn in three-dimensional space, the tangent at the current point projects down to give the tangent to the curves in each of the coordinate planes. Both x and y are functions of t, so that dx and dy may be calculated by the formulae

$$dx = f'(t)dt = \cos(t)dt,\ dy = g'(t)dt = -\sin(t)dt.$$

The direction of the tangent to the implicit curve in the *x*-*y* plane is

$$(dx,dy) = (\cos(t)dt, -\sin(t)dt)$$

$$= (\cos(t), -\sin(t))dt,$$

which is in the direction $(\cos(t),-\sin(t))$. As t increases from 0 to 2π, the tangent moves smoothly round the unit circle, passing twice through the vertical when $t=\pi/2$ and $3\pi/2$. Sticking relentlessly to the derivative concept dy/dx, there are two alternatives. One is to break the circle up in parts so that in each part one of the pair x,y is a function of the other, say

$y=\sqrt{(1-x^2)}$ for $y>0$
$y=-\sqrt{(1-x^2)}$ for $y<0$
$x=\sqrt{(1-y^2)}$ for $x>0$
$x=-\sqrt{(1-y^2)}$ for $x<0$

so that one may speak of the derivative as a function in each region. The other approach is to enter into discussions about what happens when the gradient becomes 'infinite'.

The value of seeing the central concept as the tangent vector, instead of the quotient dy/dx now becomes clear. In dealing with implicit functions it is so much simpler. It also generalises more readily to higher dimensions. The tangent vector is the best linear approximation to the curve; the tangent plane may be described as the best linear approximation to a surface, and so on.

This combination of simplicity and power proves to have a unifying influence on the calculus. In the next article I shall look at the process of antidifferentiation, which seeks the solution curve $y=\mathrm{I}(x)$, given the derivative $\mathrm{d}y/\mathrm{d}x=\mathrm{f}(x)$. The theory taught in schools for many years is an instrumental reversal of differentiation: look at the list of known derivatives to find a function $\mathrm{I}(x)$ such that $\mathrm{I}'(x)=\mathrm{f}(x)$. This presentation has a fatal flaw that is exposed by a geometric approach. Pictorial insight also generalises to give a unified approach to differential equations that is currently absent from the A-level syllabus.

References

1. D O Tall: Understanding the Calculus, *Mathematics Teaching* 110, 49-53 1985.

2. D O Tall: The Gradient of a graph, *Mathematics Teaching* 11, 48-52 1985.

3. P Couzens & D Butler: *M.E.I. Programs for Mathematical Computing*, Oundle 1983.

4. D O Tall: *Supergraph*, Glentop Publishing, Barnet 1985.

5. D O Tall: Graphic Calculus I: *Differentiation*, Glentop Publishing, Barnet 1986.

6. S M P : *Additional Mathematics* Part 2.

7. G W Leibniz: Nova methodus pro maximis et minimis, itemque tangentibus, qua nec fractas, nec irrationales quantitates moratur, & sinulare pro illis calculi genus, *Acta Eruditorum* 467-473, October 1684.

8. R Woodhouse: *The Principles of Analytical Calculation*, Cambridge 1803.

9. G Matthews: *Calculus*, John Murray 1964.

10. H Neill & H Shuard: *Teaching Calculus*, Blackie, 1982.

11. S M P : *Advanced Mathematics Book 1*, C.U.P. 1967.

12. A Orton: When should we teach the calculus? *Mathematics in School*, 14,2 11-15 1985.

13. D A Quadling: *Mathematical Analysis*, O.U.P. 1955.