CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Warwick Research Archives Portal Repository

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): David Tall
Article Title: Understanding the calculus
Year of publication: 1985
Link to published version:
Publisher statement: None

Understanding the calculus

David Tall

Mathematics Education Research Centre
University of Warwick
COVENTRY CV4 7AL

A number of significant changes have have occurred recently that give us a golden opportunity to review the teaching of calculus. The most obvious is the arrival of the microcomputer in the mathematics classroom, allowing graphic demonstrations and individual investigations into the mathematical ideas. But equally potent are new insights into mathematics and mathematics education that suggest new ways of approaching the subject.

In this article I shall consider some of the difficulties encountered studying the calculus and outline a viable alternative approach suitable for specialist and non-specialist mathematics students alike.

Students' views of the calculus

The first encounter with the calculus is usually through considering two points A,B on a graph and seeing how the chord AB tends to the tangent at A as B approaches A. (Figure 1.)

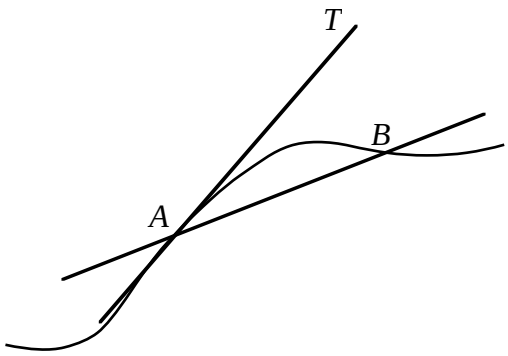

Figure 1: what happens as $B \to A$?

The informal language used at this stage can cause unforseen difficulties. There are problems interpretating in phrases such as 'tends to a limit' or 'as close as we please' which pervade the initial stages of the calculus[5]. Bernard Cornu[2] has noted corresponding difficulties in the French language. One of his most engaging observations is that many students say that the sequence 0.9, 0.99, 0.999, ... has *limit* 1 but *tends to* 'nought point nine recurring'. The explanation is that 1 is a limit in the sense of a 'speed limit': it is something that cannot be be passed. The expression 'nought point nine recurring' is a dynamic process, going on for ever and this is what the sequence 'tends' to do: it never reaches 1 in the process. Thus 'nought point nine recurring' is

considered to be just less than one, though in a strict mathematical sense the two expressions are equal.

Students observing a chord approaching a tangential position are likely to have similar feelings: the chord nearly gets there, but not quite. As one sixth-former explained patiently to me: ‘The chord line doesn’t get close to the tangent because the line is infinite; even if the lines look close, far away at infinity they are still a long way apart.’ This deeply perceptive remark shows that we mathematicians must be careful how we phrase our explanations. It is quite possible for students to have intelligent interpretations of mathematics which differ fundamentally from ours.

Sometimes a student gives an explanation that appears utterly strange to a mathematician and yet is a perfectly logical deduction in his own terms. Based on the idea of a limit as something that cannot be passed, one student noted the chord lay on both sides of the tangent and insisted: ‘The tangent can’t be a limit because you can’t get past a limit and part of the chord is already past it .’

A very common difficulty occurs because the term ‘chord’ is used in circle geometry to mean the line segment between two points on the curve. Students see the chord as a finite line segment which tends to zero length as the points get close together, so ‘the chord tends to the tangent’ because the vanishing chord gets closer to the tangent line. The gradient of chord and tangent are irrelevant and the formal explanation of the limiting process is regarded as true for quite the wrong reason.

Some students faced with figure 1 see a static picture with no movement, others (quite sensibly) see *B move along the chord* to *A*.

With this variety of interpretations, it is no wonder that the statement:

> as $B \to A$, the line through AB tends to the tangent at A

is regarded as false by a significant minority (around 30% in samples taken). A more precise, but slightly more complex statement:

> as $B \to A$ the limit of the gradient of the chord AB is the gradient of the tangent AT

produces more confusion. In some classes less than 50% regard this statement as being true before studying the calculus and a sizeable minority still consider it false after their first dose of theory.

A significant number of students are therefore bemused by the initial explanations of the derivative as a limiting process. They are shown a few simple examples, such as the derivative of x^2, developed from ‘first principles’ and a number of others are studied using the ‘delta notation’. Here the Cornu phenomenon arises again, as I have detailed elsewhere [8]:

students looking at the expression dy/dx tending to dy/dx often imagine dx tending to dx and dy tending to dy. A significant number think of dx and dy as extremely small numbers that are not zero. But they are told that dy/dx is a single symbol, not a ratio.

No wonder there is an air of relief when they move from 'first principles' to the formal algorithm for differentiating polynomials. As one bright student said: 'It's typical of teachers to show us a lot of difficult methods before getting on to the easy way to do it.'

As they move through the calculus further conflicts and contradictions arise. In the 'function of a function rule':

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

they will be told that they cannot cancel the dx, it has no separate meaning, but in

$$\int \mathrm{f}(x)\ \mathrm{d}x$$

dx now means 'with respect to x'. Later a complete volte-face occurs with a differential equation such as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{y}.$$

Now the indivisible symbol dy/dx is separated but dy, dx are not quantities, they are 'with respect to something-or-other' concoctions in the equation

$$\int y\ \mathrm{d}y = \int -x\ \mathrm{d}x.$$

Faced with such a bewildering variety of ideas is it any wonder that students are sometimes confused?

A cognitive approach to the calculus

In his paper 'Why calculus cannot be made easy' Rolph Schwarzenberger [4] argued that the mathematical difficulties of classical analysis do not admit a simple explanation. According to this thesis, any intuitive formulation of these ideas will implicitly contain the underlying difficulties to haunt the students. However, this argument demonstrates the *mathematical* difficulties of *analysis*, not the *cognitive* difficulties of the *calculus*. It shows that any 'talking down' of high-powered mathematics will contain inherent difficulties for the learner. It does not address itself to the alternative possibility of seeing the matter from the students' viewpoint and attempting to help them build up to the ideas from their current position.

To see the difference between the two possible approaches, consider the logical steps we take in explaining the derivative $f'(x)$ of a function $f(x)$. The derivative is itself a function, but we don't attack it immediately at this level: we 'simplify' the theory by first concentrating on what happens at a point. We fix x and consider the gradient from x to $x+h$ for variable h, then we take the limit as h tends to 0 (with all the hidden difficulties this implies) and when this is done we vary x again to get the derivative function. What a palaver!

In presenting the ideas in this logical order the mathematician feels that he is making the approach as 'intuitive' as possible. But he is not. A mathematician uses the term 'intuitive' as the antithesis of 'rigorous'. A psychologist uses the term 'intuitive' to signify an immediate response to a situation. There is a fundamental mismatch between the two uses of the word.

If the idea of calculating the gradient of a tangent by a limiting process were intuitive in a psychological sense then students faced with the problem would respond immediately with a limiting argument. This patently does not happen. In the current research several hundred students have answered questionnaires on the calculus. One question was:

> On the graph $y=x^2$, the point A is (1,1), the point B is (k,k^2) and T is a point on the tangent to the graph at A.
>
> (i) Write down the gradient of the straight line through A,B…
>
> (ii) Write down the gradient of AT...
>
> Explain how you might find the gradient of AT from first principles.

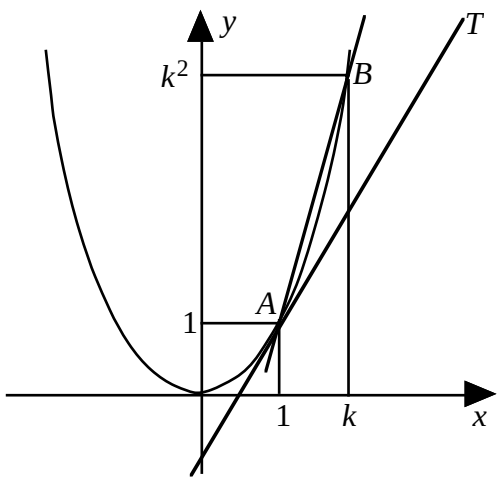

Figure 2: finding the gradient of the tangent

To date no-one who has not studied the calculus has offered a ‘limiting’ argument for the gradient of the tangent. This could be attributed to the use of the technical term ‘from first principles’ which they may regard with suspicion or not understand. But there are those with no knowledge of the calculus who are able to express the gradient of the chord for $k \neq 1$ as

$$\frac{k^2-1}{k-1} = k+1$$

and calculate the gradient of the tangent by eye as 2. When shown that as k tends to 1, so $k+1$ tends to 2, the idea appears to them as a revelation.

My conclusion is that the limiting process may be ‘intuitive’ in a mathematical sense but not in a cognitive sense. To distinguish between the two meanings of the term I suggest that the word ‘cognitive’ be used to describe mathematical ideas that appeal to the intuition. The Latin root of the word is the verb *cognoscere*, to know. The aim of the mathematics educator should be to provide a range of experiences that develop the ideas of the calculus in a cognitive manner, so that the learner both knows and understands. The desired goal is relational understanding in the sense of Skemp [6], with the concepts fitting together in a coherent, mutually supportive manner. The non-specialist may benefit from this relational understanding as an end in itself, whilst the specialist may use it as a basis for a subsequent logical understanding of the formal arguments in the sense of Skemp [7]. An appropriate cognitive development should eventually lead to proofs that are both intuitive (in a psychological sense) and rigorous. There need be no dichotomy between intuition and rigour. But how is this to be done?

The major difference between the mature mathematician and the learner is that the mathematician already has a global picture of the concept, so that when he breaks it down into a number of stages, he can see each stage as part of the whole. The learner, on the other hand sees only the part in the context of his limited understanding. He colours this part with his own meanings and places obstacles in the way of his understanding the total concept.

If somehow the learner were to get an intimation of the whole concept first, then he would be in a better position to organise his thinking processes to cope with it. Ausubel called such an intimation an ‘advance organiser’ [1]. Other schools of psychology refer to the ‘gestalt’ or wholeness of a concept being greater than the sum of its parts. Is there an approach to the calculus which provides the student with an advance

organiser for the concept of derivative to give him a gestalt to guide the development of ideas?

The micro computer

The micro offers a resource in the mathematical classroom undreamt of not long ago. It enables a cognitive approach to the calculus without the pre-requisites of limiting processes, chords approaching tangents and so on, based on the simple fact that the derivative is not just the gradient of the tangent, *it is the gradient of the graph itself.*

Give a group of students a computer program capable of magnifying graphs over tiny ranges and ask them what they can find. A typical response will be that when a graph is magnified it looks ‘less curved’. (Figure 3.) Small portions of most graphs that students magnify tend to approximate to straight lines. Thus the notion of the gradient of such a curved graph becomes reasonable: highly magnify the graph and take the gradient of the resulting (approximately) straight line segment.

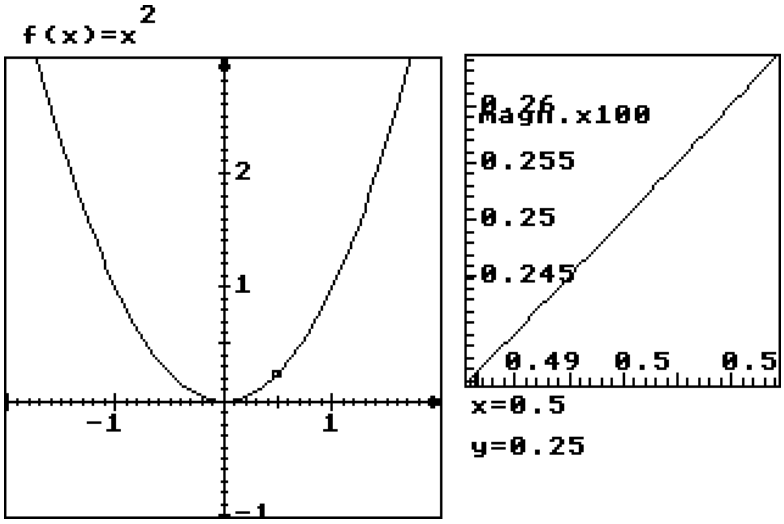

Figure 3: magnifying a small part of a graph

In this way it is possible to ‘see’ the gradient of a curved graph without any use of tangents, chords or any other superstructure. But chords prove a helpful way of representing the gradient on a computer screen by taking two points a, b close together and drawing the extended chord through $(a,\mathrm{f}(a))$, $(b,\mathrm{f}(b))$. (Figure 4.)

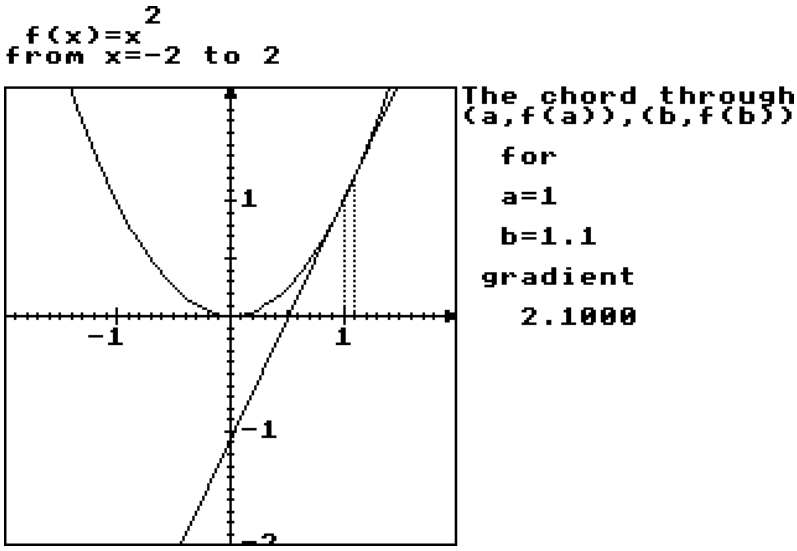

Figure 4: the extended chord through two close points

To obtain a global understanding of the derivative, draw the chord through (x,f(x)), (x+c,f(x+c)), keep the value of c a small constant and plot the chord gradient as x moves from left to right. Figure 5 shows a 'freeze-frame' of the gradient function of f(x)=x^2 being built up. (It is far more easy to see when the picture is growing dynamically!)

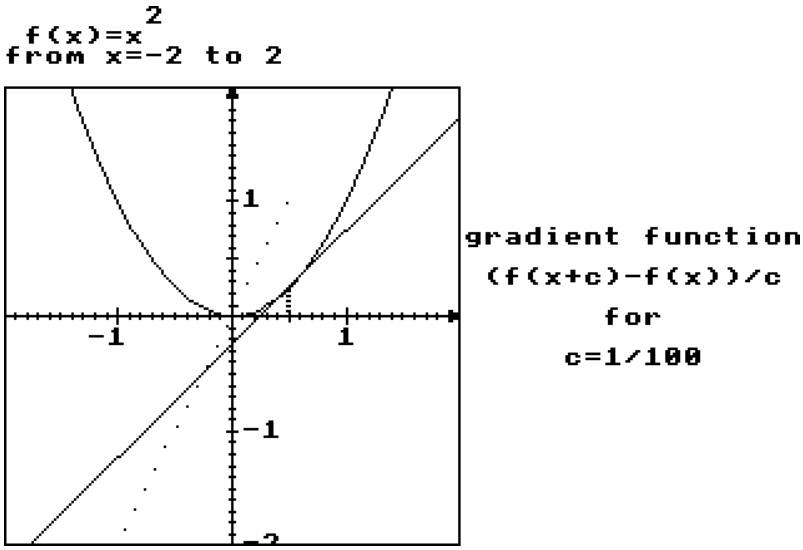

Figure 5: Building up the gradient function

This kind of approach allows students to investigate idea of gradient of a curved graph and gain a cognitive understanding of concepts that are mathematically difficult. For instance, they could look at functions which are not differentiable at certain points: the graph of f(x)=$|x^2-x|$ magnified around x=1 gives two line segments meeting at an angle. At x=1 it has a clear gradient to the left and a different gradient to the right: a simple example of a function which is not differentiable at some point. (Figure 6.)

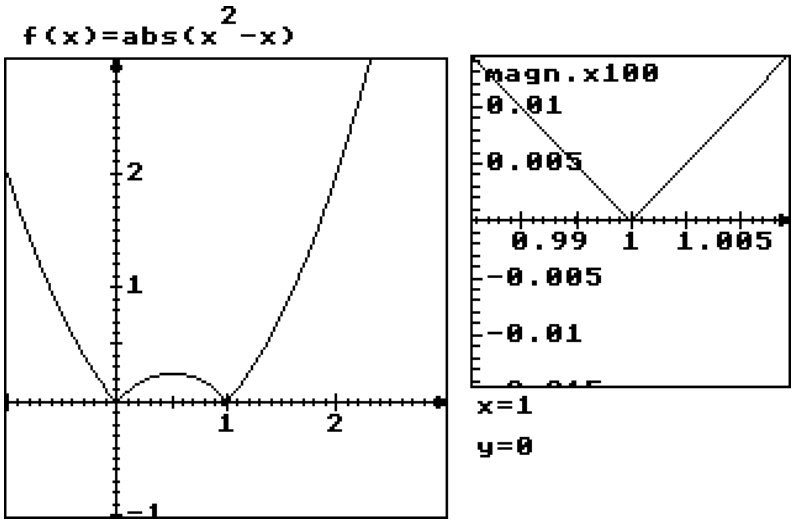

Figure 6: a graph not locally straight at certain points

In my software *Graphic Calculus I* [9] there is a program to draw of a function which is *nowhere* differentiable: it is so wrinkled that it nowhere magnifies to give a straight line. The theory necessary to prove this result is mathematically difficult but cognitively simple: if the cognitive ideas are followed through in a sensible way, they eventually lead to a simple, but rigorous, mathematical proof!

Students of all abilities can use appropriate software to investigate the gradient functions for x, x^2, x^3 to see if there is a pattern and conjecture the formula for the derivative of x^n. Having suggested a formula it may be tested for other values of n, say n=4, n=5 or n=32. Negative values such as n=–1, –2 or fractional values such as n=1/2, n=3/4 or n=–5/6 may be investigated, giving a good feeling for the ideas before formal proof is possible. It will not be long before students see the limitations of guessing the derivative from a global picture and respond positively to a formula derived from a limiting argument. But now they have the advance organiser of the notion of derivative to guide their thinking.

As the theory progresses, further investigations will be possible which allow students to take an active part in the learning instead of passively receiving the theory. A good example is the definition of the mathematical constant e and the derivative of e^x. A formal proof of the derivative of e^x requires an explanation of how e is defined together with the calculation of the limit

$$\frac{e^{x+h}-e^x}{h}$$

as h tends to zero. Instead, start with the function f(x)=2^x and plot the approximate gradient function. (Figure 7.) It is clearly the same shape as that of 2^x but a little lower. A similar investigation with y=3^x gives a graph whose gradient is the same shape, but this time a little higher. Is

there a graph $y=k^x$ with k between 2 and 3 whose gradient is identical to the original? It is easy to narrow the search down a little more to give k between 2.5 and 3. The number e is defined as the value such that $y=e^x$ has the derivative $dy/dx=e^x$.

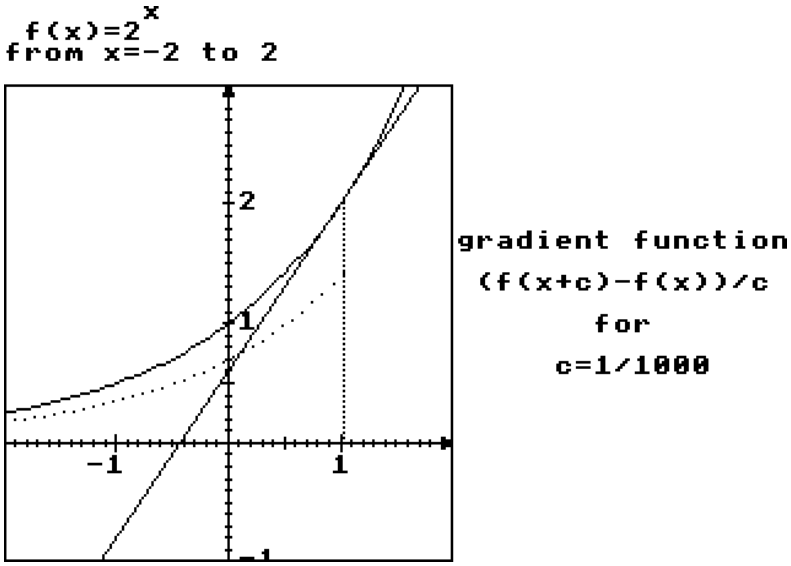

Figure 7: A graph with a similar shaped gradient

These investigations offer precisely the kind of cognitive development I mentioned earlier, helping students to get a coherent relational understanding of the ideas as they gain experience of the subject. The vast majority of students do not go on to study formal mathematical analysis and will surely benefit from a coherent feeling for geometric ideas of rate of change. Those who become mathematics specialists will benefit from more in-depth investigations that allow them to obtain insight into some of the subtler problems of the mathematical analysis.

Tony Orton [3] has noted that students find great difficulty with the concept of the integral

$$\int_a^b f(x)\,dx$$

when $f(x)$ is negative or b is less than a. Interpreted geometrically a profound cognitive demonstration of the ideas can be given.

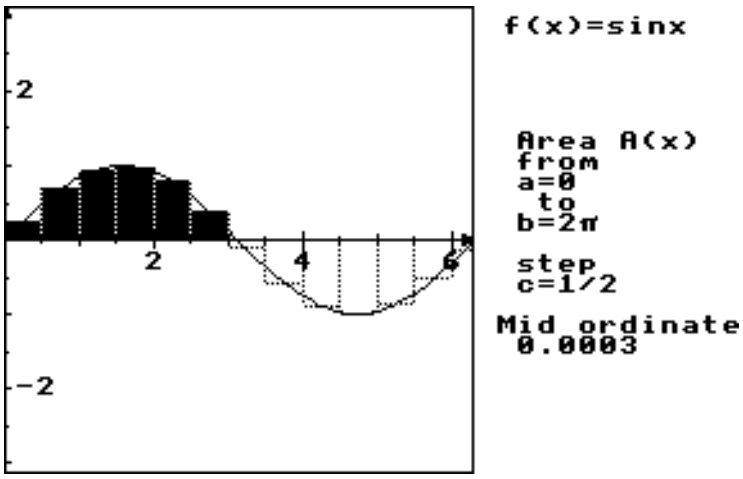

Figure 8: positive and negative area calculations

Figure 8 shows the area between the *x*-axis and the graph of sin*x* being calculated from 0 to 2π using the midordinate rule and step width 1/2. The area of each strip is calculated as the step-size times the ordinate and the result is negative where the step is positive and the ordinate negative. It is shown in the picture by shading the rectangles differently. Taking a large number of strips, it is clear that the sum gets close to the area under the graph (figure 9) and the areas are given the appropriate signs. The case where *b* is to the left of *a*, and the step-length is negative, also gives a logical choice of sign.

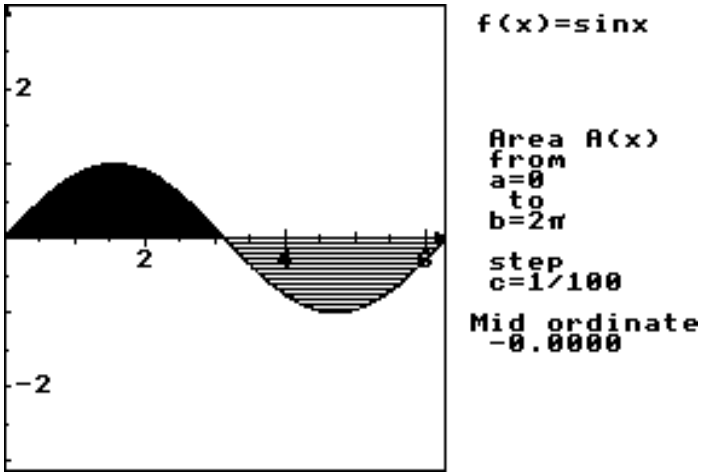

Figure 9: area calculation with tiny steps

A geometric approach using a computer should not be thought of as a lower form of mathematical experience than a formal logical development. On the contrary, it can prove to have direct benefits in understanding mathematical theory that is inadequately handled in the current A-level.

As an example, consider the solution of the differential equation

$$\frac{dy}{dx} = \frac{1}{x^2}$$

The analytic method is to try to think of a function $y=\mathrm{f}(x)$ so that $\mathrm{f}'(x)=1/x^2$. This is fairly straightforward in this case and is usually given as

$$y=-1/x+c.$$

But this is not the most general solution. Geometric insight shows that two solution curves have the same gradients and therefore a solution curve may be shifted up or down by a constant amount to give another solution. But if the domain of the function is in separate parts, then different shifts are possible on each part, so a more general solution of the given problem is

$$y=\begin{cases}-1/x+k\ (\text{for}\quad x<0)\\ -1/x+c\ (\text{for}\quad x>0)\end{cases}$$

where c and k may be different.

A second fundamental flaw in the attempt to solve differential equations by spotting formulae for the solution is that it only leads to a number of isolated techniques for a small number of particular equations. There are equations such as

$$\frac{\mathrm{d}y}{\mathrm{d}x}=x^2+y^2$$

which have no solutions as formulae in the standard functions. But from a geometric viewpoint a solution curve $y=\mathrm{s}(x)$ through a point (a,b) in the plane must have gradient $\mathrm{s}'(a)=a^2+b^2$. By drawing short line segments with the appropriate gradient at each point, it is possible to visualize the direction taken by solution curves. These are drawn in figure 10 and the computer has been used to follow along one of them. Using the computer, accurate numerical approximations to solution curves may be easily found.

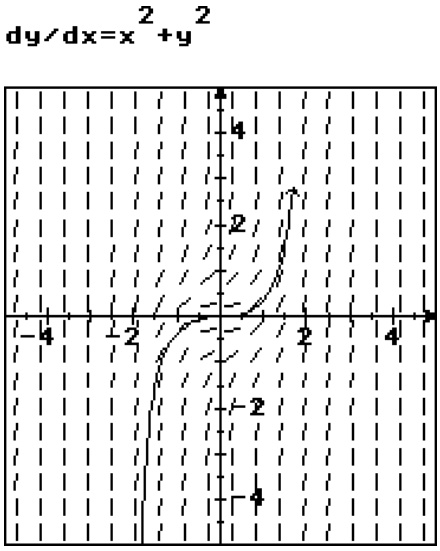

Figure 10: The numerical solution of a differential equation

In the real world the theory of calculus is used to solve differential equations by numerical methods in weather-forecasting, aerodynamic design and many other areas where solutions using formulae are totally inappropriate. It is time for us to make our mathematics more relevant to its applications.

In subsequent articles I will look at this approach to the calculus in greater detail, tracing the development from a geometric idea of differentiation through integration and on to differential equations. At an elementary level geometric insight can prove helpful for students of all ranges of ability. For more able students we will see the ideas give insight into more difficult theorems of mathematical analysis.

References

1. D P Ausubel, J D Novak & H Hanesian: *Educational Psychology: A Cognitive View*, Holt, Rinehart & Winston,1968.
2. B Cornu B: *Apprentissage de la notion de limite: Conceptions et Obstacles*, Doctoral thesis, Grenoble,1983.
3. A Orton: *A cross-sectional study of the understanding of elementary calculus in adolescents and young adults*, Ph.D. thesis, Leeds University,1980.
4. R L E Schwarzenberger: 'Why calculus cannot be made easy', *Mathematical Gazette*, 64, 158–166, 1980.
5. R L E Schwarzenberger & D O Tall: 'Conflicts in the learning of real numbers and limits', *Mathematics Teaching*, 82, 44–49,1978.
6. R R Skemp: 'Relational understanding and instrumental understanding', *Mathematics Teaching*, 81, 20–26, 1976.
7. R R Skemp: 'Goals of learning and qualities of understanding', *Mathematics Teaching*, 88, 44–49,1979.
8. D O Tall & S Vinner: 'Concept image and concept definition in mathematics', with particular reference to limits and continuity', *Educational Studies in Mathematics*, 12, 151–169,1981
9. D O Tall: *Graphic Calculus I, II, III*:, Glentop Publishing,1986.