



© 1997–2009, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

January 2004

Features



Making the grade: Part II

by Chris Sangwin



In Making the grade: Part I in Issue 27 of Plus we considered what the gradient of a curve might mean, and how to find it by appealing directly to the definition. In particular, we used direct arguments – which were really quite involved – to calculate the gradients of the curves x^2 and $\sin(x)$. To perform this kind of calculation every time we need to calculate such a gradient would be a nightmare – especially if we had a complicated function. In this article we think about the process of manipulating the algebraic expressions with which we usually describe functions in order to perform this calculation. This is differentiation as we know and love it!



The other kind of gradient.

Image [DHD Photo Gallery](#)

The whole point of having a set of formal rules is to allow us to *temporarily* forget the exact meaning and to concentrate on calculation. After all, we can only concentrate on a few things at a time. Of course, it is vital to keep the meaning in the back of our minds, as a check that the answer is sensible. Furthermore, by having a set of rules disjoint from a particular context, we can apply the rules in many different settings.

The rules allow us to differentiate just about any algebraic expression we care to write down. Of course we have to decide *which* formal rules to apply in a given situation, and *in what order*. Sometimes it is not clear which rule we should apply – there are a number of things we could do correctly. What then to do? How should we decide? As we shall see, the answer to these questions is surprising and illustrates the intimate way in which calculus and algebra interact.

Calculating gradients using the calculus

In the previous article we calculated the gradient by considering the change in f divided by the change in x , that is,

$$\frac{f(x+h) - f(x)}{h} \quad \text{for all } h \neq 0. \quad (1)$$

When we carried out this calculation for the function $f(x) = x^2$ we obtained

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x + h.$$

Taking the limit as h tends to zero, either positive or negative, gave $2x$ as the derived function, or derivative, of x^2 .

Let's generalize this and consider $f(x) = x^n$ where n is any natural number, that is, $n = 1, 2, 3, \dots$. Of course, we have already considered the cases when $n = 1$ (the straight line) and when $n = 2$ (the quadratic).

Making the grade: Part II

In order to calculate (1) when $f(x) = x^n$ we need to consider

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^n - x^n}{h}.$$

To simplify this we need to expand out the term

$$(x+h)^n.$$

When we do this for small values of n we get

$$\begin{aligned}(x+h)^0 &= 1, \\(x+h)^1 &= x+h, \\(x+h)^2 &= x^2 + 2xh + h^2, \\(x+h)^3 &= x^3 + 3x^2h + 3xh^2 + h^3, \\(x+h)^4 &= x^4 + 4x^3h + 6x^2h^2 + 4xh^3 + h^4, \\(x+h)^5 &= x^5 + 5x^4h + 10x^3h^2 + 10x^2h^3 + 5xh^4 + h^5.\end{aligned}$$

We could do this by hand for other values of n by multiplying out the brackets, but this is tricky, time-consuming and it is all too easy to slip up. In fact, there is a very regular pattern to the coefficients of the terms in the expansions above. If we ignore the x 's and h 's we obtain the pattern known as *Pascal's Triangle*, part of which is shown below.

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & 1 & 1 & & \\ & & 1 & 2 & 1 & & \\ & 1 & 3 & 3 & 1 & & \\ 1 & 4 & 6 & 4 & 1 & & \\ 1 & 5 & 10 & 10 & 5 & 1 & \end{array}$$

This pattern, which may be continued forever, is obtained by adding together two adjacent numbers in one row to generate the number below. For example, the blue 4 is generated by adding the 3 and 1 above it. Similarly, the red 10 is obtained by adding the 4 and 6 above it.

In general, the number in the position $k+1$ in from the left on the n th row is given by the formula

$${}^nC_k = \frac{n!}{(n-k)!k!}.$$

These numbers are known as the *binomial coefficients*, because nC_k is the coefficient of $x^{n-k} h^k$ when we expand $(x+h)^n$. This result is known as the *binomial theorem* and it allows us to exploit this pattern to write $(x+h)^n$ as

$$(x+h)^n = x^n + {}^nC_1 x^{n-1} h + {}^nC_2 x^{n-2} h^2 + {}^nC_3 x^{n-3} h^3 + \dots + {}^nC_{n-2} x^2 h^{n-2} + {}^nC_{n-1} x h^{n-1} + h^n.$$

We are currently interested in calculating the quantity (1) when $f(x) = x^n$. To do this we note that, for all values of n ,

$${}^nC_1 = \frac{n!}{(n-1)!} = \frac{n \times (n-1) \times (n-2) \times \dots \times 2 \times 1}{(n-1) \times (n-2) \times \dots \times 2 \times 1} = n.$$

Using this we have

$$\frac{(x+h)^n - x^n}{h} = \frac{(x^n + {}^nC_1 x^{n-1} h + {}^nC_2 x^{n-2} h^2 + \dots + h^n) - x^n}{h} = nx^{n-1} + h \times (\text{stuff}).$$

We don't really need to know exactly what the "stuff" is in the above expression in this case. This is because it is multiplied by h , and letting h tend to zero wipes out all these terms. What we are left with is the function nx^{n-1} . So we express this result as

$$\frac{d}{dx}x^n = nx^{n-1} \quad (2)$$

whenever n is a natural number (ie, $n = 1, 2, 3, \dots$). If $n = 1$ in this argument we have

$$\frac{d}{dx}x = 1,$$

which confirms that the gradient of the straight line $f(x) = x$ is constant.

This result may be expanded so that (2) holds whenever n is a *real* number, although this takes a little more work.

To take another example, let $n = 0$ (remember that $x^0 = 1$ for all x). So we have the formula

$$f(x) = x^0 = 1.$$

This is a horizontal straight line, which has gradient zero. Does our formula (2) agree?

$$\frac{d}{dx}x^0 = 0 \cdot x^{-1} = 0.$$

So the mechanical formula (2) again agrees with a simple case we can easily imagine. This is a useful check.

Constructing functions

Functions are fundamental to modern mathematics and you simply can't avoid using them. The idea of a function is to take two sets of objects known as the *inputs* and *outputs*. To every input the function assigns a unique output:

$$\text{input} \rightarrow \boxed{\text{Function}} \rightarrow \text{output}.$$

Most often the inputs and outputs are sets of numbers, such as the real line. The function is also most often described using a formula, in the form of an algebraic expression. This is exactly the idea of a function we have considered so far, although we haven't been explicit about it! This is also the way we will continue to think about functions.

Making the grade: Part II

The reason we pause now to think of functions in a more abstract way is simply to acknowledge that a function is much more general than a formula. In fact, the function $|x|$ introduced in the previous article was built from two formulae bolted together. Recall these were

$$|x| := \begin{cases} x, & x \geq 0 \\ -x, & x < 0. \end{cases}$$

The trigonometric functions $\sin(x)$, $\cos(x)$, etc. are constructed with reference to a geometrical shape – in this case a circle of radius 1. Other ways of building functions involve an infinite series (that is a sum) or a sequence of formulae. We won't consider these in this article, but just concentrate on how we can build up functions from simple operations.

Let us assume our input is a number x . The simplest operations we could perform on our variable are the arithmetic ones. That is addition, multiplication and the two inverse operations of subtraction and division. Because we can manipulate the formulae using algebra we can often write one formula in different ways.

For example, consider

$$f(x) = x^2(3 - x^2) = 3x^2 - x^4, \quad (3)$$

which is shown in Figure 1 below.

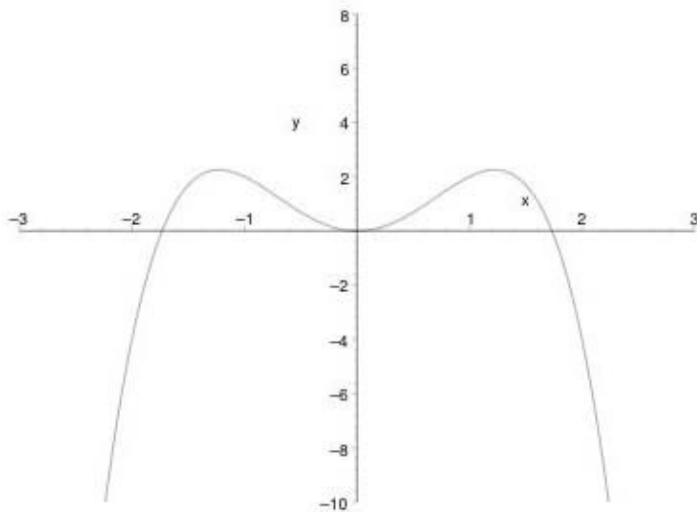


Figure 1: The function (3)

We can think of $f(x)$ as x^2 multiplied by $3 - x^2$. Both these functions are shown in Figure 2. Try to imagine what happens when you multiply the values on each graph together.

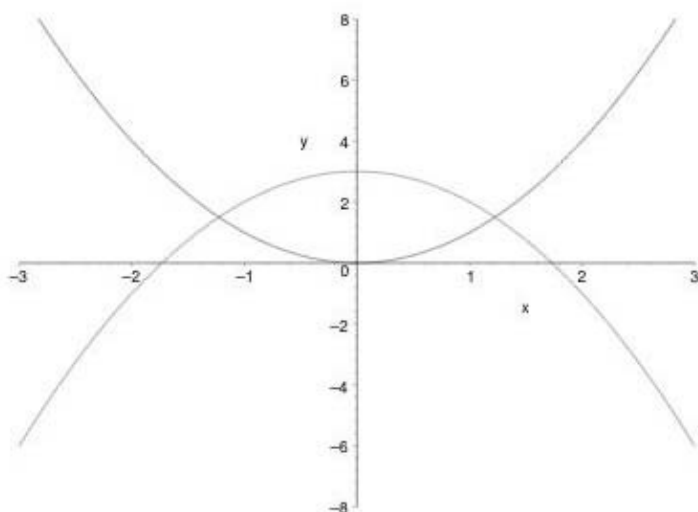


Figure 2: The functions x^2 and $3-x^2$

Alternatively, to calculate $f(x)$ we might subtract x^4 from $3x^2$. The graphs of these functions are shown in Figure 3. Since $f(x) = 3x^2 - x^4$ we can recreate Figure 1 by subtracting one from the other. No doubt there are other ways of constructing the same function f .

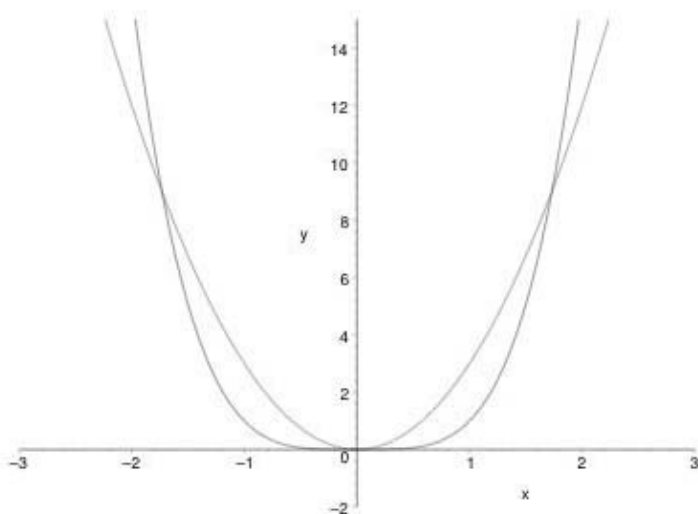


Figure 3: The functions x^4 and $3x^2$

Functions can also be applied in order, one after the other, as in

input $\rightarrow$ Function 1 $\rightarrow$ Function 2 $\rightarrow$ output.

So, our function (3) can be thought of as applying the function

$$g(u) = u(3 - u)$$

to the result of the function

$$u(x) = x^2.$$

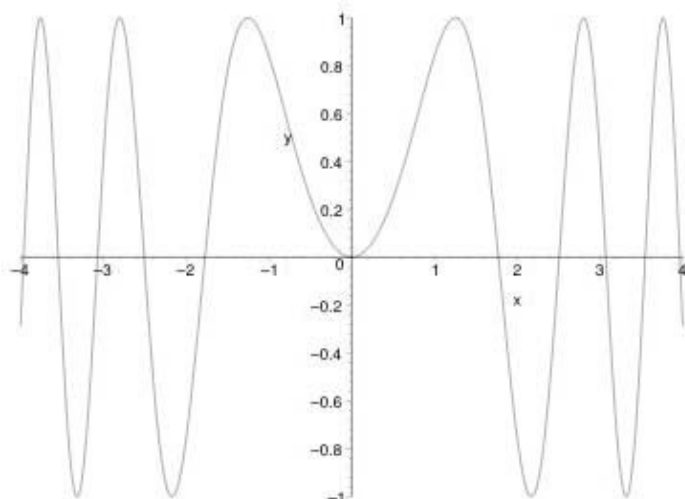


Figure 4: The function $\sin(x^2)$

Of course, we have to ensure that any output from Function 1 is a legitimate input for Function 2. In this case we say the two functions have been *composed*. For example, a function such as $\sin(x^2)$ can be thought of as the function that maps x to $\sin(x)$ applied to the *result* of the function that maps x to x^2 . Note that the order really does matter here and $\sin(x^2)$ and $\sin(x)^2$ are very different functions: see Figures 4 and 5.

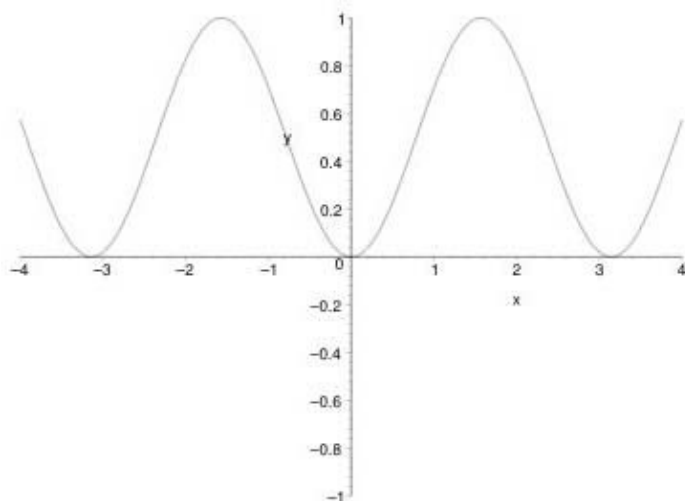


Figure 5: The function $\sin(x)^2$

Given the numerous ways we could express a function such as (3), how should we go about differentiating it? This is the question we address in the rest of this article.

Linearity of the differential calculus

You don't always get the same result if you do things in a different order!

The first general rule allows us to calculate the derivative of two functions which have been added together. If we want to find the gradient of $f(x)+g(x)$ we simply find the gradients of $f(x)$ and $g(x)$ separately and then add the results. In a more condensed (and easier to read) form this may be expressed as:

$$\frac{d}{dx}(f(x) + g(x)) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x). \quad (4)$$

Similarly, if $f(x)$ is multiplied by a constant a then

$$\frac{d}{dx}af(x) = a\frac{d}{dx}f(x). \quad (5)$$

Together the two rules above are known as *linearity*, and they allow us to easily calculate the derivative of *any* polynomial $p(x)$ by breaking it down into constant multiples of x^n for various n , and then applying (2). This is powerful indeed.

Example

For example, to calculate the gradient of the function f defined in (3), we write this as the unfactored form $3x^2 - x^4$ and can then apply the rules as follows:

$$\begin{aligned} \frac{d}{dx}(3x^2 - x^4) &= \frac{d}{dx}3x^2 + \frac{d}{dx}-x^4 \quad \text{using (4),} \\ &= 3\frac{d}{dx}x^2 - \frac{d}{dx}x^4 \quad \text{using (5),} \\ &= 6x - 4x^3 \quad \text{using (2) twice.} \end{aligned}$$

Any book on calculus will contain many similar examples and exercises for you to practice.

Before we go any further, we need a word of warning about notation. In particular, there are many ways of writing the derivative of a function f at the point x . Different authors have different preferences. So far we

have used the notation

$$\frac{d}{dx}f(x),$$

which was promoted by Leibnitz. Another notation, used by Newton, has two forms:

$$f'(x) \text{ or } \dot{f}(x).$$

Although neater in some circumstances, it is very easy to misread a dot or apostrophe and so care is needed. We will use both kinds of notation.

General rules

Linearity, which is expressed in the formulae (4) and (5), together with our result (2) allows us to calculate the derivative of any polynomial by breaking it into separate parts. In fact (4) and (5) involve two *general functions*. What would be really useful would be two rules which allow us to calculate gradients when general functions are multiplied or composed together, that is to say, rules which allow us to find

$$\frac{d}{dx}(f(x) \times g(x)) \tag{6}$$

and

$$\frac{d}{dx}f(g(x)), \tag{7}$$

where f and g are any differentiable functions.

We make a huge assumption in believing that such general rules really exist. However, *if they do* then the rules applied to x^n in various different ways must respect the result (2). For example, x^4 may be written as $x^2 \times x^2$, or as $x \times x^3$. The rule for (6), if it exists, must give $4x^3$ when applied to each of these ways of writing x^4 . Otherwise we could obtain different answers for the derivative. So, we look at different ways of writing x^n as a product, and try to find a rule which is consistent, at least for these.

Let's start by defining $F(x) = x^n$ and split this up into

$$f(x) = x^m \quad \text{and} \quad g(x) = x^{n-m},$$

so that

$$F(x) = f(x) \times g(x).$$

We know using (2) that

$$F'(x) = nx^{n-1}, \quad f'(x) = mx^{m-1}, \quad \text{and} \quad g'(x) = (n-m)x^{n-m-1}.$$

Our task is to write $F'(x)$ in terms of $f'(x)$ and $g'(x)$ as an attempt to gain some insight into what the general rule (6) might be. That is, we write

Making the grade: Part II

$$nx^{n-1} = \underbrace{mx^{m-1}}_{f'(x)} \times A(x) + \underbrace{(n-m)x^{n-m-1}}_{g'(x)} \times B(x),$$

where A and B are unknown functions of x . Now, using algebra we can confirm that

$$\underbrace{nx^{n-1}}_{F'(x)} = \underbrace{mx^{m-1}}_{f'(x)} \times \underbrace{x^{n-m}}_{g(x)} + \underbrace{(n-m)x^{n-m-1}}_{g'(x)} \times \underbrace{x^m}_{f(x)}.$$

Thus if we take $A(x) = x^{n-m} = g(x)$ and $B(x) = x^m = f(x)$ we have a correct general rule whenever we split x^n into (8). This rule may be written as

$$F'(x) = f'(x)g(x) + f(x)g'(x). \quad (8)$$

Immediately, by linearity, it follows that (8) holds for any polynomial.

Can we find a rule for general functions, like $\sin(x)$, which are not polynomials? Certainly, if a general rule exists, when we apply this to $x^n = x^m \times x^{n-m}$, the rule should reduce to (8). In fact, the rule for general functions turns out to be *precisely* (8), and is better known as the *Product Rule*.

Similarly, if we write

$$F(x) = x^{nm} = (x^n)^m$$

and

$$f(x) = x^n, \quad \text{and} \quad g(x) = x^m,$$

we have $F(x) = g(f(x))$. Again, we know using (2) that

$$F'(x) = nm x^{nm-1}, \quad f'(x) = nx^{n-1}, \quad \text{and} \quad g'(x) = mx^{m-1}.$$

Our task is to write $F'(x)$ in terms of $f'(x)$ and $g'(x)$ as an attempt to gain some insight into what the general rule (7) might be. But

$$nm x^{nm-1} = \underbrace{nx^{n-1}}_{f'(x)} \underbrace{m \overbrace{(x^n)^{m-1}}^{f(x)}}_{g'(f(x))}$$

which suggests a general rule

$$F'(x) = f'(x)g'(f(x)). \quad (9)$$



There's a rule for everything – even chains!

Image [DHD Photo Gallery](#)

In fact, the rule for general functions turns out to be *precisely* (9) again, and is better known as the *Chain Rule*.

These rules are really very comprehensive, and proofs may be found in any text book on advanced calculus. A huge range of functions can be build up, and conversely decomposed, using these rules. When trying to differentiate a complicated function, the method is to decompose it into simpler components, and work with these separately. The derivatives of these simple parts can be recombined using the general rules to find the derivative of the original function. For completeness we state these rules again below, in two common forms of notation and for each give a worked example.

The chain rule

The following rule allows us to differentiate functions built up by compositions. Let's assume that we have some function which we can choose to write as a composition $g(f(x))$. Let $u = f(x)$, then

$$\frac{d}{dx}(g(f(x))) = \left(\frac{d}{dx}f(x)\right) \left(\frac{d}{du}g(u)\right), \quad (10)$$

or, using alternative notation,

$$(g(f(x)))' = f'(x)g'(f(x)).$$

Example:

Let us return and consider the function $\sin(x^2)$. Writing $g(u) = \sin(u)$ and $u = x^2$, we already know that

$$\frac{d}{du}\sin(u) = \cos(u) \quad \text{and} \quad \frac{d}{dx}x^2 = 2x.$$

Substituting these into (10) gives

$$\frac{d}{dx} \sin(x^2) = 2x \cos(x^2).$$

Note: we have $\cos(x^2)$ here, not $\cos(x)$, as we have $u = x^2$.

The product rule

Functions can be built up by components which are multiplied together. For example, $f(x) \times g(x)$. To differentiate this we use the formula

$$\frac{d}{dx}(f(x) \times g(x)) = \left(\frac{d}{dx}f(x)\right) \times g(x) + f(x) \times \left(\frac{d}{dx}g(x)\right), \quad (11)$$

or, using alternative notation,

$$(f(x) \times g(x))' = f'(x)g(x) + f(x)g'(x).$$

Example:

This time consider the function $x^2 \sin(x)$. This can be decomposed into the two functions $x^2 \times \sin(x)$, each of which we know how to differentiate. Therefore we can use the rule (11) immediately to write

$$\frac{d}{dx} x^2 \sin(x) = 2x \sin(x) + x^2 \cos(x).$$

The quotient rule

Functions can be built up by components which are divided one by the other. Actually, since dividing by a is identical to multiplying by $\frac{1}{a}$ to work out the derivative of $\frac{f(x)}{g(x)}$ we could just apply the product rule to $f(x) \times \frac{1}{g(x)}$, and the chain rule to $g(x)$ composed with $\frac{1}{x} = x^{-1}$. But it is convenient to have a separate rule for this, which is

$$\frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = \frac{g(x) \left[\frac{d}{dx} f(x) \right] - \left[\frac{d}{dx} g(x) \right] f(x)}{g(x)^2}, \quad (12)$$

or, using alternative notation,

$$\left[\frac{f(x)}{g(x)} \right]' = \frac{g(x)f'(x) - g'(x)f(x)}{g(x)^2}.$$

Example:

This time we differentiate $\frac{\sin(x^2)}{x^2}$. We know how to differentiate $\sin(x^2)$, even though it is itself a composition. We applied (10) to this and it has derivative $2x \cos(x^2)$. Using (12) gives

$$\begin{aligned}\frac{d}{dx} \left[\frac{\sin(x^2)}{x^2} \right] &= \frac{x^2 \left[\frac{d}{dx} \sin(x^2) \right] - \left[\frac{d}{dx} x^2 \right] \sin(x^2)}{(x^2)^2} \\ &= \frac{x^2 \times 2x \cos(x^2) - 2x \sin(x^2)}{x^4} \\ &= \frac{2x^2 \cos(x^2) - 2 \sin(x^2)}{x^3}.\end{aligned}$$

Applying the rules

To finish, instead of giving lots of different examples (which can be found in any calculus text), we take the reverse approach and think about one example in more detail. In particular, we return to the function x^4 . We know from the previous result (2) that the derivative of x^4 is $4x^3$. We write this as

$$\frac{d}{dx} x^4 = 4x^3.$$

Using the rules of algebra we could write this function in a number of ways. We can also think of $x^4 = x^2 \times x^2$. Alternatively, $x^4 = x \times x^3$ or even $x^4 = x \times x \times x \times x$.

Taking the first of these, since the derivative of x^2 is $2x$, we may apply the *product rule* to give

$$\frac{d}{dx}(x^4) = \frac{d}{dx}(x^2 \times x^2) = 2x \times x^2 + x^2 \times 2x = 2x^3 + 2x^3 = 4x^3.$$

We could also apply the product rule to one of the other representations of this function. In particular, we could calculate the derivative as

$$\frac{d}{dx}(x^4) = \frac{d}{dx}(x \times x^3) = 1 \times x^3 + x \times 3x^2 = x^3 + 3x^3 = 4x^3.$$

Similarly we could think of x^4 being a composition of two functions: as " x -squared, all squared", that is to say, $x^4 = (x^2)^2$. In this case we may apply the *chain rule* to see that

$$\frac{d}{dx}(x^4) = \frac{d}{dx}((x^2)^2) = 2(x^2) \times 2x = 4x^3.$$

Notice that in each case we get the same correct answer.

We need not stop there. For example, we could write $x^4 = x^6/x^2$ and apply the quotient rule. If we do this we have

$$\frac{d}{dx} \left(\frac{x^6}{x^2} \right) = \frac{x^2 \times \frac{d}{dx}(x^6) - \frac{d}{dx}(x^2) \times x^6}{(x^2)^2} = \frac{x^2 \times 6x^5 - 2x \times x^6}{x^4} = \frac{6x^7 - 2x^7}{x^4} = 4x^3.$$

The point is that in order to find the derivative of x^4 we may do *anything algebraically legitimate* and apply *any of the rules for differentiation correctly*. The way we choose to find the derivative, as long of course as it

is applied correctly, does not matter. We could ask: How many ways are there of differentiating x^4 ? Can you think of others?

About the author



Chris in the Volcanoes National Park, Hawaii, Summer 2003

Chris Sangwin is a member of staff in the School of Mathematics and Statistics at the University of Birmingham. He is a Research Fellow in the Learning and Teaching Support Network centre for Mathematics, Statistics, and Operational Research. His interests lie in mathematical Control Theory.

Chris would like to thank Mr Martin Brown, of Thomas Telford School, for his helpful advice and encouragement during the writing of this article.



Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.