Limit-free derivatives

Citation for published version:
Sangwin, CJ 2011, 'Limit-free derivatives' The Mathematical Gazette, vol. 95, no. 534, pp. 469-482.

Limit-free derivatives

Christopher J Sangwin
MSOR Network, School of Mathematics, University of Birmingham,
Birmingham, B15 2TT, UK. Email: C.J.Sangwin@bham.ac.uk

Algebraic double roots are used by [14] to motivate the following limit-free definition of *derivative*:

> A function $f(x)$ has a derivative m at $x = a$ if
> $$f(x) - (mx + c) = q(x)(x - a)^2 \tag{1}$$
> for some value c.

As we shall see later, "function" in this definition will actually be restricted to real polynomials and [14] concludes

> We have shown how an elementary algebraic principle – double roots – can lead to a complete calculus of polynomials and related functions, without the need for a limit concept. And given the difficulty which many beginning students have with limits, this is an idea worth pursuing in some calculus classes. The difficulty of the algebra in our approach can be ameliorated by referring to a computer algebra system.

Here we define the derivative for polynomials in an alterative limit-free way using an elementary algebraic *form*. We also discuss how and when this definition can be extended to other elementary functions. Furthermore, this note will argue that:

- to calculate this form it is only necessary to perform substitutions and expansion of brackets;
- this is a direct result of an algorithm, and does not require us to solve a system of equated coefficients, the technique used by [14] to work operationally with (1);
- our form provides a direct link between algebraic representations and the geometric concept of curves of best approximation at the point $x = a$;
- our definition links to higher order derivatives in a natural and direct way;
- our definition has an interesting link to the historical development of the calculus via Lagrange's attempt to avoid infinitesimals and base analysis on series.

Furthermore, the definition points the way to more advanced areas of mathematical analysis. It is by using these advanced areas that we explain our alternative limit-free derivative here. In particular, we do not propose a complete curriculum sequence or report trials of introducing derivatives in this way.

1 An algebraic form "*about the point* $x = a$"

Rearranging (1) we have

$$f(x) = c + mx + q(x)(x-a)^2 \tag{2}$$

or, in an alternative form,

$$f(x) = \bar{c} + m(x-a) + q(x)(x-a)^2.$$

We therefore see that $\bar{c} = f(a)$ and $m = f'(a)$, or put another way

$$f(x) = f(a) + f'(a)(x-a) + q(x)(x-a)^2. \tag{3}$$

The *Taylor series* of a function $f(x)$ at the point $x = a$ is

$$\begin{aligned} f(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n. \end{aligned}$$

A polynomial has a *finite* Taylor series. The best fit polynomial of order n to $f(x)$ locally at $x = a$ is the truncation of the Taylor series about $x = a$, including only terms in powers of x of order $\leq n$. The smallness of $(x-a)^m$ for x "close" to a can provide both a conceptual and analytic justification for the use of the adjective "best" in this definition. Furthermore, the truncation of the Taylor series about $x = a$ can be found as the remainder after polynomial long division by $(x-a)^{n+1}$. So, the tangent line at $x = a$ is the remainder after division by $(x-a)^2$. Operationally, it is relatively unproblematic to perform polynomial long division by $(x-a)^{n+1}$ to obtain the required remainder using the standard algorithm. For example, to find the tangent line to $p(x) = x^3 - 6x^2 + 10x - 3$ at $x = 2$ we divide by $(x-2)^2 = x^2 - 4x + 4$.

$$\begin{array}{r|rrrr} & & & x & -2 \\ \hline x^2 - 4x + 4 & x^3 & -6x^2 & +10x & -3 \\ & -x^3 & +4x^2 & -4x & \\ \hline & & -2x^2 & +6x & -3 \\ & & 2x^2 & -8x & +8 \\ \hline & & & -2x & +5 \end{array}$$

We could define the derivative as the coefficient of x in the tangent line. A computer algebra system could be employed. However, the calculations can be performed in the following, even more straightforward, way.

We illustrate this by finding the tangent line to $p(x) = x^3 - 6x^2 + 10x - 3$ at $x = 2$. Since we are interested in the point $x = 2$ we shift the origin to this point by evaluating $p(x+2)$ and expand out the brackets.

$$\begin{aligned} p(x+2) &= (x+2)^3 - 6(x+2)^2 + 10(x+2) - 3 \\ &= x^3 - 2x + 1. \end{aligned}$$

Next we truncate this expression and notice that the best fit line at $x = 0$ is the line $l(x) = -2x + 1$. We shift this back to give

$$l(x-2) = -2(x-2) + 1 = -2x + 5,$$

which is the tangent line to $p(x)$ at $x = 2$.

Appreciation of form is an important part of algebraic thinking. There are many forms for polynomials, e.g. gathered terms (i.e. $2x^3$ rather than x^3+x^3), expanded, factored, completed square and so on. We highlight the importance of the form in which a polynomial is "written about the point $x = a$". A polynomial is written *about the point* $x = a$ when it is written as

$$p(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \cdots + a_n(x-a)^n. \quad (4)$$

In our example above,

$$p(x) = x^3 - 6x^2 + 10x - 3 = 1 - 2(x-2) + (x-2)^3. \quad (5)$$

To be definite, we reverse the usual order of terms for polynomials making our form directly comparable to truncations of Taylor's series.

In general, the algorithm to re-write $p(x)$ about the point $x = a$ is

1. make the substitution $x + a$ for x;
2. expand out and gather like terms, reversing the usual order;
3. make the substitution $x - a$ for x.

We will call this process "shift-expand-shift". It involves only substitution of variables, and expansion of binomials (or other terms), and is therefore arguably much simpler than polynomial long division or the identification of double roots. While it is clear that the result is in the form (4), it is a theorem, the proof of which is straightforward, that this algorithm results in a polynomial algebraically equivalent to $p(x)$. A consequence is that this form is *canonical*, which means that each polynomial $p(x)$ can be represented uniquely in this form. The algorithm can be used directly on any form of polynomial, e.g. a factored polynomial can be written about $x = a$ without any preparation. We shall use this form to define our derivative.

All forms have particular advantages, e.g. one advantage of the factored form is the ease with which it can be used to solve the equation $p(x) = 0$. The advantage of writing a polynomial about the point $x = a$ is the ease with which the curves of local best fit at $x = a$ can be obtained by truncating. The result of this truncation process, $r(x)$ say, is also the remainder from the division of $p(x)$ by $(x-a)^{n+1}$. Indeed, when convenient, we may write $p(x)$ as

$$p(x) = r(x) + (x-a)^{n+1}q(x), \quad (6)$$

assuming that $r(x)$ is the unique remainder of $p(x)$ on division by $(x-a)^{n+1}$.

With knowledge of Taylor series it is clear what is happening but our proposal is to reverse this ordering. Hence, to create a limit-free derivative for polynomials we simply work in the other direction.

> **Definition:** The *derivative* of $f(x)$ at $x = a$ is the coefficient of $(x - a)$ when written about the point $x = a$.

The *derived function* is a function which gives the coefficient of $(x - a)$, written as a function of x. For polynomials at least, this coefficient is uniquely defined at $x = a$ so that we may say f is *differentiable* at $x = a$ and simply write $f'(a)$ to represent it. Furthermore, we may use (3) wherever convenient for a differentiable function.

This definition has a geometrical interpretation: the derivative is a function of x which gives the slope of the tangent line. We reiterate however, that tangency is justified by appealing to the existence of a double root, as in [14]. If we wanted the tangent line itself, we simply truncate the polynomial written about $x = a$. Furthermore, we may readily obtain higher order curves of best fit to the function $f(x)$ at the point $x = a$ by truncating the form (4).

Notice in (5) the quadratic term is missing. Hence, for this example, the tangent line is also the best approximating quadratic at $x = 2$. This is because $x = 2$ is the point of inflection of this cubic. Such missing coefficients or degeneracy is accommodated by shift-expand-shift in quite a natural way.

The example $f(x) = x^n$

Applying this definition to $f(x) = x^n$ at $x = a$ we have the first shift as $(x + a)^n$ which we expand out to the first term in x only (as we don't need the other explicit terms here),

$$(x + a)^n = a^n + na^{n-1}x + \cdots .$$

We shift this back to give

$$x^n = a^n + na^{n-1}(x - a) + \cdots .$$

The *derived function* gives the coefficient of $(x - a)$, but written as a function of x. So, the derivative is nx^{n-1}. We have at least made explicit the conceptual shift necessary in turning the coefficient, i.e. gradient, into a new function of x. Put another way, we "unapply" the function which evaluates the derivative at $x = a$, i.e. $f'(a) = na^{n-1}$, to create a new function of x. Algorithmically, we simply make the substitution $a = x$ in the coefficient of $(x - a)$. When [14] evaluates m, in terms of a, the same issue arises.

We have succeeded in obtaining the formula

$$\text{if } f(x) = x^n \text{ then } f'(x) = nx^{n-1} \tag{7}$$

for positive integer n without using any limiting arguments. *"This rule can easily be memorized"* ([5, §152]) and put to use in the traditional way. We have certainly shifted the effort from limits to using the binomial theorem confidently. But fluency in algebra is necessary for any operational success with calculus. The extent to which fluency in algebra or in calculus is desirable in a computer algebra world is not a debate we take up here, but we do argue that appreciation of form, and fluency in the selection and use of forms, will always be critical.

Linearity

Let

$$p_1(x) = r_1(x) + (x - a)^2 q_1(x).$$

$$p_2(x) = r_2(x) + (x-a)^2 q_2(x).$$

Then for $A_1, A_2 \in \mathbb{R}$ we have

$$(A_1 p_1 + A_2 p_2)(x) = (A_1 r_1 + A_2 r_2)(x) + (x-a)^2 (A_1 q_1 + A_2 q_2)(x).$$

So linearity is immediate and our derivative of x^n can be used to obtain the derivatives of all polynomials.

Higher derivatives

The second derivative of $f(x) = x^n$ is found using linearity and (7) as $f''(x) = n(n-1)x^{n-2}$ and by induction we have

$$f^{(k)}(x) = n(n-1)\cdots(n-k+1)x^{n-k}.$$

If we now expand out $(x+a)^n$ further we have

$$(x+a)^n = a^n + na^{n-1}x + \tfrac{n(n-1)a^{n-2}}{2}x^2 + \cdots .$$

$$x^n = a^n + na^{n-1}(x-a) + \tfrac{n(n-1)a^{n-2}}{2}(x-a)^2 + \cdots .$$

With linearity, our attention is therefore drawn to the equality of $\frac{f''(a)}{2}$ with the coefficient of $(x-a)^2$. There are many similar relationships, including those with *finite differences* which [5] combines with “infinitesimals”, dx, to motivate a definition of derivative.

Product rule

To prove the product rule, we simply take $n = 1$ and write $r_i(x) = p_i(a) + p_i'(a)(x-a)$, $i = 1, 2$ and multiply out

$$\begin{aligned}
p_1 \times p_2 &= \big(r_1 + (x-a)^2 q_1\big) \times \big(r_2 + (x-a)^2 q_2\big) \\
&= r_1 r_2 + (x-a)^2 \big(r_1 q_2 + r_2 q_1 + q_1 q_2 (x-a)^2\big) \\
&= (p_1(a) + p_1'(a)(x-a))(p_2(a) + p_2'(a)(x-a)) \\
&\quad +(x-a)^2 (\cdots) \\
&= (p_1(a) + p_2(a)) + (p_1'(a)p_2(a) + p_2'(a)p_1(a))(x-a) \\
&\quad +(x-a)^2 (\cdots) .
\end{aligned}$$

The derivative is the coefficient of $(x-a)$ in the above expression, i.e. $p_1'(x)p_2(x) + p_2'(x)p_1(x)$. We note that [14] writes (1) in the form (2) to prove the product rule in a limit-free manner.

Chain rule

Assume we have two functions f and g, and that g is differentiable at $x = a$ and f is differentiable at $x = g(a)$. We are interested in the composition $f(g(x))$. First, differentiability of f at $x = g(a)$ enables us to write

$$f(x) = f(g(a)) + f'(g(a))(x - g(a)) + \bar{f}(x)(x - g(a))^2. \tag{8}$$

Differentiability of g at $x = a$ enables us to write

$$g(x) - g(a) = g'(a)(x - a) + \bar{g}(x)(x - a)^2.$$

We are interested in the coefficient of $(x - a)$ when $f(g(x))$ is written about $x = a$. To this end we substitute $g(x)$ into (8) as follows

$$\begin{aligned}
f(g(x)) &= f(g(a)) + f'(g(a))(g(x) - g(a)) + \bar{f}(g(x))(g(x) - g(a))^2 \\
&= f(g(a)) + f'(g(a))(g'(a)(x - a) + \bar{g}(x)(x - a)^2) \\
&\quad + \bar{f}(g(x))(g'(a)(x - a) + \bar{g}(x)(x - a)^2)^2 \\
&= f(g(a)) + g'(a)f'(g(a))(x - a) \\
&\quad + (x - a)^2 \left[f'(g(a))\bar{g}(x) + \bar{f}(g(x)) \left(g'(a)^2 + 2g'(a)\bar{g}(x)(x - a) + \bar{g}(x)^2(x - a)^2\right)\right]
\end{aligned}$$

We notice that the coefficient of $(x - a)$, written as a function of x, is $g'(x)f'(g(x))$, which proves the chain rule.

nth roots of $f(x)$

Next we enlarge the range of functions to include nth roots of existing functions. Assume that $f(x)$ is differentiable, so that for each a we have (3). We assume $f(x)^{\frac{1}{n}}$ is differentiable and that for some $\bar{q}(x)$ we have

$$f(x)^{\frac{1}{n}} = c + m(x - a) + \bar{q}(x)(x - a)^2.$$

Raising both sides to the power n, expanding out and equating $f(x)$ with (3) we have

$$\begin{aligned}
f(a) + f'(a)(x - a) + q(x)(x - a)^2 &= (c + m(x - a) + \bar{q}(x)(x - a)^2)^n \\
&= c^n + mnc^{n-1}(x - a) + (x - a)^2(\cdots). \qquad (9)
\end{aligned}$$

Thus $c = f(a)^{\frac{1}{n}}$ and

$$f'(a) = mn\left(f(a)^{\frac{1}{n}}\right)^{n-1} = mnf(a)^{\frac{n-1}{n}}.$$

Hence

$$m = \frac{1}{n}f'(a)f(a)^{\frac{1}{n}-1}.$$

This results in the correct answer for the derivative of $f(x)^{\frac{1}{n}}$, but what is the nature of $\bar{q}(x)$ and the term $(\cdots)$ in (9)?

The quotient rule

The focus of our attention so far has been on polynomials which do not require a quotient rule. However, we now seek to extend the definition to rational functions, and hence need to apply our definition to find the derivative of $\frac{1}{f(x)}$. We assume that f is differentiable and that $f(a) \neq 0$, and we write $f(x)$ in the form (3). We assume that the derivative of $\frac{1}{f(x)}$ exists and equals m. Therefore, for some $\bar{q}$, we equate

$$\frac{1}{f(a) + f'(a)(x - a) + q(x)(x - a)^2} = c + m(x - a) + \bar{q}(x)(x - a)^2.$$

$$\begin{aligned} 1 &= (f(a) + f'(a)(x-a) + q(x)(x-a)^2)(c + m(x-a) + \bar{q}(x)(x-a)^2) \\ &= f(a)c + (mf(a) + cf'(a))(x-a) + (x-a)^2(\cdots). \end{aligned}$$

Equating constants and the coefficient of $(x-a)$ we have

$$c = \frac{1}{f(a)}$$

and

$$0 = mf(a) + \frac{f'(a)}{f(a)} \quad \text{i.e.} \quad m = -\frac{f'(a)}{f(a)^2}.$$

This combined with the product rule gives us the quotient rule.

This proof is significantly shorter and more direct than that of [14]. However, the sleight of hand in this derivation remains: $\bar{q}(x)$ might not be a finite series of terms! Indeed, if $f(x) = x + 1$ then

$$\frac{1}{x+1} = \frac{1}{1+a} - \frac{(x-a)}{(1+a)^2} + \frac{(x-a)^2}{(1+a)^3} - \frac{(x-a)^3}{(1+a)^4} + \frac{(x-a)^4}{(1+a)^5} - \cdots$$

Therefore, if we seek to extend the definition given here for polynomials to rational functions we will encounter infinite series when writing an expression about a point $x = a$. The same problem occurs when taking nth roots.

What holds for the finite...

These observations allow us to reconnect with infinite series, or what [5] treats as "infinite polynomials". Indeed, Euler can often be paraphrased as saying "what holds for the finite holds for the infinite". Assume that

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots + a_k x^k + \cdots.$$

First we differentiate this term by term using (7), avoiding any question of convergence, to obtain

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots + k a_k x^{k-1} + \cdots.$$

Expressions for higher derivatives can easily be obtained: by induction we have

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1)(k-2)\cdots(k-n+1) a_k x^{k-n}.$$

Strictly speaking, we have only defined the form "about $x = a$" for polynomials, but working formally we can apply the technique shift-expand-shift to a series and write a series about $x = a$. Therefore, substituting $x + a$ for x we obtain

$$f(x+a) = a_0 + a_1(x+a) + a_2(x+a)^2 + \cdots + a_k(x+a)^k + \cdots$$

which we can write out in columns to expand as follows

$$\begin{array}{clclclclclcl}
 & a_0 & = & a_0 & & & & & & & & \\
+ & a_1(x+a) & + & a_1 a & + & a_1 x & & & & & & \\
+ & a_2(x+a)^2 & + & a_2 a^2 & + & 2a_2 a x & + & a_2 x^2 & & & & \\
+ & a_3(x+a)^3 & + & a_3 a^3 & + & 3a_3 a^2 x & + & 3a_3 a x^2 & + & a_3 x^3 & & \\
+ & a_4(x+a)^4 & + & a_4 a^4 & + & 4a_4 a^3 x & + & 6a_4 a^2 x^2 & + & 4a_4 a x^3 & + & a_4 x^4 \\
+ & \vdots & + & \vdots & & \vdots & & \vdots & & & & \\
+ & a_k(x+a)^k & + & a_k a^k & + & k a_k a^{k-1} x & + & \frac{k(k-1)}{2} a_k a^{k-2} x^2 & + & \cdots & & \\
+ & \vdots & + & \vdots & & & & & & & &
\end{array}$$

Totalling the first column to the right of the equality symbol gives us

$$a_0 + a_1 a + a_2 a^2 + \cdots + a_k a^k + \cdots = f(a).$$

The second column gives

$$(a_1 + 2a_2 a + 3a_3 a^2 + \cdots + k a_k a^{k-1} + \cdots)x = f'(a)x,$$

The third column gives

$$(a_2 + 3a_3 a + 6a_4 a^2 + \cdots + \tfrac{k(k-1)}{2} a_k a^{k-2} + \cdots)x^2 = \frac{f''(a)}{2} x^2,$$

and so on. In general the $(n+1)$th column from the equality contains terms with x^n. This term arises from the coefficient of x^n in the expansion of $a_k(x+a)^k$ (where $k \geq n$) which is

$$a_k \left(\begin{array}{c} k \\ n \end{array} \right) a^{k-n} \; x^n.$$

The total of these terms in the $(n+1)$th column equals

$$\sum_{k=n}^{\infty} \left(\begin{array}{c} k \\ n \end{array} \right) a_k a^{k-n} \; x^n.$$

Note that since $\left(\begin{array}{c} k \\ n \end{array} \right) = \frac{k!}{(k-n)!n!} = \frac{k(k-1)\cdots(k-n+1)}{n!}$, we have

$$\sum_{k=n}^{\infty} \left(\begin{array}{c} k \\ n \end{array} \right) a_k a^{k-n} = \sum_{k=n}^{\infty} \frac{k(k-1)\cdots(k-n+1)}{n!} a_k a^{k-n} = \frac{f^{(n)}(a)}{n!}.$$

Hence, ignoring all issues of convergence when totalling over the whole table and performing the reverse shift, we have

$$f(x) = \sum_{n=0}^{\infty} \left(\sum_{k=n}^{\infty} \left(\begin{array}{c} k \\ n \end{array} \right) a_k a^{k-n} \right) (x-a)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n,$$

i.e. our shift-expand-shift procedure has derived Taylor's expansion formally.

Exponential, logarithmic and trigonometric functions

We have, so far, avoided any discussion of the exponential, logarithmic and trigonometric functions. Our formulation of the derivative presupposes that the function can be written as an algebraic expression. As such, it is ideal for polynomials and we have shown how it can be extended to nth roots, rational functions via the quotient rule, and used formally with infinite series. However, here we have ignored the issue of convergence and worked mechanically with symbols. To deal with exponential, logarithmic and trigonometric functions we cannot avoid infinite series.

We define the following function by a series

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots . \tag{10}$$

Expanding out each term, just as with Taylor's expansion, we can argue that

$$f(x + y) = f(x)f(y) \tag{11}$$

and, using both shift-expand-shift and term-by-term differentiation, that

$$f'(x) = f(x). \tag{12}$$

If we define $e = f(1)$ then analogy with algebraic notation and (11) suggests strongly we could write $f(x) = e^x$. This is one route to differentials of the exponential function.

Writing $\ln(x)$ as the inverse of e^x we note that $a^x = e^{\ln(a)x}$ and the chain rule enables us to differentiate a^x for any a.

We can at this point expand our thinking into complex numbers. The proof in [15] begins with the traditional elementary definition of trigonometrical functions as ratios in geometric figures and makes use of (12) to prove that $e^{iz} = \cos(z) + i\sin(z)$. When combined with (10), we can readily find the series for $\sin(x)$ and $\cos(x)$, and hence their derivatives using shift-expand-shift. A lot of this kind of "informal analysis" can introduce and perhaps even motivate the ideas which a university analysis course should make more rigorous. However, by now we are using some quite involved concepts and technical algebra. This technical load will certainly cause problems, perhaps even comparable to those of defining the derivative in the traditional way using limits. Furthermore, we have completely ignored any difficulties with convergence. It is for this reason that the limit-free approach ultimately fails mathematically, but for very interesting reasons.

2 Historical background

This section places the preceding suggestions into a brief historical context, with much more detail provided in, for example, [3, 4]. The first approaches to introducing derivatives involved infinitesimals, e.g. [5]. However, until the recent work of non-standard analysis, e.g. [16], the logical basis of infinitesimals was not clear. This is exemplified by the calculation of the derivative of $f(x) = x^2$ by considering a short chord of length h. First we assume $h \neq 0$ to calculate the gradient of the chord

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

Next assume $h = 0$ to give $f'(x) = 2x$. The same argument works for $f(x) = x^n$ using the binomial theorem, as we have seen. But, surely we cannot have both $h \neq 0$ and then $h = 0$? Such criticism of the infinitesimal calculus came early, for example most famously by Berkeley.

> And what are these Fluxions? The Velocities of evanescent Increments? And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the Ghosts of departed Quantities? [1]

Furthermore, the traditional calculation of the derivative of $\sin(x)$, which requires $\lim_{h\to 0} \frac{\sin(h)}{h}$, appeals to geometric reasoning. This potentially confuses algebraic and geometric notions making unclear which notions are basic (i.e. axiomatic).

In an attempt to provide a solid algebraic foundation for the theory of functions, and so to avoid these philosophical difficulties, Joseph Louis Lagrange (1736–1813) attempted to develop analysis as the theory of *series*, [7]. In [11], (see also [18]), he begins by assuming functions can be developed as the series

$$f(x+i) = f(x) + pi + qi^2 + ri^3 + \cdots . \tag{13}$$

The coefficient of i is what Lagrange calls the *first derived function* of $f(x)$, indeed he is the first to introduce the notation $f'(x)$. To identify relationships between coefficients in (13), he considers $f(x+i+o)$. First he replaces i by $i+o$ and expands out the terms in (13) using the binomial theorem. Then he replaces x by $x+o$ in (13) and repeats the procedure. By writing $f(x+i+o)$ in these two different ways he is able to equate coefficients of i and from this derive Taylor's expansion, much as we have done.

Lagrange implicitly assumes every function can be written as a series of the form (13) and that such series essentially behave as polynomials of infinite degree. Furthermore, term by term operations are valid and properties such as continuity are preserved.

Consider the following function

$$f(x) = \sum_{n=1}^{\infty} \frac{x^2}{(1+x^2)^n} = \frac{x^2}{1+x^2} + \frac{x^2}{(1+x^2)^2} + \frac{x^2}{(1+x^2)^3} + \frac{x^2}{(1+x^2)^4} + \cdots .$$

Admittedly this has an infinite number of terms, but it is a thinly disguised geometric progression of rational terms, and is precisely the kind of object studied by the seventeenth and eighteenth century mathematicians such as Euler and Lagrange. Clearly $f(0) = 0$ and for $x \neq 0$ this converges to 1, so that f is well-defined. Each partial sum is continuous, but the limit is discontinuous at $x = 0$. For $a \neq 0$, the function

$$f(x) = \sum_{n=1}^{\infty} \frac{x^2}{a^2 + x^2 n^2} = -\frac{x^2}{2a^2} + \frac{\pi x}{2a} \coth\left(\frac{a\pi}{x}\right)$$

is continuous at $x = 0$ but not differentiable there: indeed $f(x)$ has one-sided derivatives of $\pm\frac{\pi}{2a}$ at $x = 0$.

As a more striking example, we have functions which are continuous but nowhere differentiable, e.g. Weierstrass's example

$$\sum_{n=1}^{\infty} a^n \cos(b^n \pi x)$$

where b is an odd integer, $0 < a < 1$ and $ab > 1 + \frac{3}{2}\pi$. See, e.g. [9]. This example, discovered only much later of course, is conclusive evidence of the failure of Lagrange's approach: in order to use Taylor's theory to write this function as a series we need a point at which we are able to differentiate it infinitely often. However, while we can easily formally differentiate individual terms in the series, we can't even differentiate the function itself once, somewhere!

Even when a function is continuous and infinitely differentiable there are problems. Cauchy's famous example $f(x) = e^{-\frac{1}{x^2}}$, $x \neq 0$ and $f(0) = 0$, has derivatives of all orders at zero which are equal to zero. Hence, the Taylor series about zero converges to the original function only at $x = 0$. In modern terms, a function that can be expressed locally about a point $x = a$ by a convergent power series is called *analytic*. This function fails to be analytic about $x = 0$, not because of a failure to find derivatives of all orders (they exist and equal zero), but because there is no interval on which the series converges to the original function.

In the other direction, we might ask what coefficients can arise in a Taylor's series? Here we find little comfort: any sequence of real numbers (a_n) can arise as the Taylor coefficients. That is to say, there exists a function $f(x)$, analytic on $(-1, 1)$ such that $f^{(n)}(0) = a_n$. [2]

These few examples illustrate why Lagrange's attempt to base a theory of function on infinite series fails. Many further examples are given in [2] and a much more detailed discussion of functions, including more detail of the subsequent development of the modern limit definition, is given by [8] and [17, 13]. However, the reasons for the failure of Lagrange's attempt are precisely those which make analysis such an interesting subject within mathematics. In this paper, we have retreated into the safe territory of polynomials, and the definition given here is consistent with, and equivalent to, that based on limits. But, attempts to expand this into a more general theory of functions ultimately fails for the same reasons.

However, the apparent simplicity of the *proofs* can be rescued, and we end this section by mentioning *little-o* notation. [6, 12]

$f(x)$ is *little-o* of $h(x)$ as x approaches a, i.e. we write $f(x) = o(h(x))$ as $x \to a$, if

$$\lim_{x \to a} \frac{f(x)}{h(x)} = 0.$$

Further, we say that $f(x)$ and $g(x)$ are *linearly the same* at $x = a$, written $f \sim_a g$, if

$$f(x) - g(x) = o(x - a) \text{ as } x \to a.$$

$\sim_a$ is an equivalence relation and f is differentiable at $x = a$ in the traditional sense if and only if $f(x) \sim_a f(a) + m(x - a)$, where m is the traditional derivative of f at a, i.e. $m = f'(a)$. That is to say, we may write

$$f(x) = f(a) + f'(a)(x - a) + o(x - a) \text{ as } x \to a.$$

Notwithstanding the fact that all this relies on the traditional limit concept, this notation enables us to write proofs of the traditional calculus rules in the form outlined in this article. Indeed, these proofs generalize to multi-variable cases in a natural way, see, e.g. [12, pg. 366–376].

Similarly we may consider other curves of local best fit of order n at $x = a$ by defining an equivalence relation

$$f(x) \backsim g(x) \text{ if and only if } f(x) - g(x) = o((x-a)^n) \text{ as } x \to a.$$

Other approaches include using "big-O notation", or *strong derivative*, which are equivalent to the traditional derivative for many common functions but not all, [10]. This is much closer to our approach here as

$$f(x) = f(a) + f'(a)(x-a) + O((x-a)^2).$$

This has a certain appeal, and indeed is proposed as a third alternative to the traditional and limit-free approaches to teaching.

> Students will be motivated to use O notation for two important reasons. First, it significantly simplifies calculations because it allows us to be sloppy, but in a satisfactorily controlled way. [10]

3 Conclusion

In this note we discuss the algebraic form "about a point $x = a$", and point out that shift-expand-shift is a simple way to find this form. The remainder after polynomial long division is equivalent to truncations of this form, and these ideas combine to provide an alternative limit-free definition of the derivative. This definition is equivalent to the traditional definition for polynomials, and working at a formal symbolic level, ignoring convergence, we have extended it to some important elementary functions. Lastly we place this in a historical context, and apply the criticisms of Lagrange's attempt to formulate analysis based on infinite series to our limit-free approach.

Quite whether limit-free declivities can, or should, be turned into an alternative teaching path is a matter for further careful thought and debate. How this might be combined as natural precursor to little-o notation is a further consideration. The author would very much welcome correspondence on this issue.

References

[1] G. Berkeley. *The Analyst; or, a discourse addressed to an Infidel Mathematician.* London, 1734.

[2] G. G. Bilodeau. The origin and early development of non-analytic infinitely differentiable functions. *Archive for History of Exact Sciences*, 27(2):115–135, June 1982.

[3] C. B. Boyer. *The History of the Calculus and Its Conceptual Development*. Dover, 1959.

[4] C. H. Edwards. *The Historical Development of the Calculus*. Springer-Verlag, 1979.

[5] L. Euler. *Foundations of Differential Calculus*. Springer, 2000. Translated by Blanton, J. from the Latin *Institutiones Calculi Differentialis, 1755*.

[6] J. Fowler. little o notation. Private Correspondence.

[7] C. F. Fraser. Joseph Louis Lagrange's algebraic vision of the calculus. *Historia Mathematica*, 14(1):38–53, 1987.

[8] A. Gardiner. *Understanding Infinity: The Mathematics of Infinite Processes*. Dover, 2003.

[9] G. H. Hardy. Weierstrass's non-differentiable function. *Transactions of the American Mathematical Society*, 17(3):301–325, July 1916.

[10] D. E. Knuth. Teach calculus with big O. *Notices of the American Mathematical Society*, 45(6):687–688, June 1998.

[11] J. L. Lagrange. *Théorie des Fonctions Analytiques*. Librarie de Bachelier, Paris, 3rd edition, 1847.

[12] S. Lang. *Undergraduate Analysis*. Springer, 1983.

[13] N. Luzin. The evolution of function: Part II. *The American Mathematical Monthly*, 105(3):263–270, 1998.

[14] A. McAndrew. An elementary, limit-free calculus for polynomials. *The Mathematical Gazette*, 94(529):67–83, 2010.

[15] T. Needham. *Visual Complex Analysis*. Oxford, 1997.

[16] A. Robinson. *Non-standard analysis*. Princeton University Press., revised edition, 1996.

[17] A. Shenitzer and N. Luzin. The evolution of function: Part I. *The American Mathematical Monthly*, 105(1):59–67, 1998.

[18] D. J. Struik. *A source book in Mathematics 1200–1800*. Harvard College, 1969.

Acknowledgements

The author wishes to acknowledge Joe Kyle and an anonymous referee for helpful comments and historical suggestions.