Topics in Generalized Differentiation

J. Marshall Ash

Abstract

The course will be built around three topics:

(1) Prove the almost everywhere equivalence of the L^p n-th symmetric quantum derivative and the L^p Peano derivative.

(2) Develop a theory of non-linear generalized derivatives, for example of the form $\sum a_n f(x+b_n h) f(x+c_n h)$.

(3) Classify which generalized derivatives of the form $\sum a_n f(x+b_n h)$ satisfy the mean value theorem.

1 Lecture 1

I will discuss three types of difference quotients. The first are the additive linear ones. These have been around for a long time. One can see their shadow already in the notation Leibnitz used for the dth derivative,

$$\frac{d^d}{dx^d}.$$

For an example with $d = 2$, let $h = dx$, and consider the Schwarz generalized second derivative

$$\lim_{h\to 0} \frac{d^2 f(x)}{h^2} = \lim_{h\to 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{1}$$

The difference quotient associated with the generalized additive linear derivative has the form

$$\begin{aligned} D_{uw} f(x) &= \lim_{h\to 0} D_{uw} f(x,h) \\ &= \lim_{h\to 0} \frac{\Delta_{uw} f(x,h)}{h^d} \\ &= \lim_{h\to 0} \frac{\sum_{i=0}^{d+e} w_i f(x+u_i h)}{h^d}. \end{aligned} \tag{2}$$

Here f will be a real-valued function of a real variable. The w_i's are the weights and the $u_i h$'s are the base points, $u_0 > u_1 > \cdots > u_{d+e}$. This is a generalized dth derivative when w and u satisfy

$$\sum_{i=0}^{d+e} w_i u_i^j = \begin{cases} 0 & j = 0, 1, ..., d-1 \\ \\ d! & j = d \end{cases} . \tag{3}$$

The *scale* of such a derivative will be the closest that two base points come when h is set equal to 1. Thus the scale of the Schwarz generalized second derivative is

$$\min \{ |(x+1) - x| , |(x+1) - (x-1)| , |x - (x-1)| \} = 1.$$

That is why we will systematically write, for example, the third symmetric Riemann derivative as

$$\lim_{h \to 0} \frac{f\left(x + \frac{3}{2}h\right) - 3f\left(x + \frac{1}{2}h\right) + 3f\left(x - \frac{1}{2}h\right) - f\left(x - \frac{3}{2}h\right)}{h^3} \tag{4}$$

instead of the completely equivalent

$$\lim_{h \to 0} \frac{f(x+3h) - 3f(x+\ h) + 3f(x-\ h) - f(x-3h)}{8h^3}.$$

In other words, there are an infinite number of ways to write the difference quotient for a single generalized derivative, but only at most 2 if the scale is fixed at 1. (Note that the substitution $h \to -h$ does not change (1), it does change the standard first difference quotient, $\frac{f(x+h)-f(x)}{h} \neq \frac{f(x-h)-f(x)}{-h}$.) Henceforth the scale will always be taken to be 1.

We will systematically abuse notation, referring to any of the difference quotient, or its numerator, or its limit, as the derivative. Thus, the second symmetric Riemann derivative (also called the Schwarz derivative) will refer either to the quotient

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

or to its numerator

$$f(x+h) - 2f(x) + f(x-h),$$

or to its limit as given in equation (1) above.

The third symmetric Riemann derivative, given by the quotient (4) is an instance of the kth symmetric Riemann derivative. There is one of these as well as a forward Riemann derivative for every natural number k. The first three forward Riemann derivatives are given by

$$\frac{f(x+h)-f(x)}{h},$$

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

and

$$\frac{f(x+3h)-3f(x+2h)+3f(x+h)-f(x)}{h^3}.$$

Zygmund and Marcinkiewicz in the middle 1930s did some of the best early work on generalized derivatives. Here is an important family of generalized derivatives that they introduced. They began with a family of differences.

$$\tilde{\Delta}_1 = f(x+h) - 2^0 f(x),$$

$$\begin{aligned}\tilde{\Delta}_2 &= [f(x+2h)-f(x)] - 2^1[f(x+h)-f(x)] \\ &= f(x+2h)-2f(x+h)+f(x),\end{aligned}$$

$$\begin{aligned}\tilde{\Delta}_3 &= [f(x+4h)-2f(x+2h)+f(x)] \\ &\qquad - 2^2[f(x+2h)-2f(x+h)+f(x)] \\ &= f(x+4h)-6f(x+2h)+8f(x+h)-3f(x).\end{aligned} \tag{5}$$

The next one is formed by taking the third difference at $2h$ and subtracting 2^3 times the third difference from it; and thus has base points $8h, 4h, 2h, h$, and 0. It should be clear how to inductively create the entire family of these differences. The dth difference corresponds to a dth derivative, after multiplication by a constant. Other generalized derivatives show up in numerical analysis. One measure for evaluating how good a derivative is comes from considering Taylor expansions. If f is sufficiently smooth, we write

$$f(x+h) = f(x) + f'(x)h + f''(x)\frac{h^2}{2} + f'''(x)\frac{h^3}{3!} + \ldots$$

Substitute this into, say, the third difference (5) to get

$$\tilde{\Delta}_3 = \begin{pmatrix} f(x+4h) \\ -6f(x+2h) \\ +8f(x+h) \\ -3f(x) \end{pmatrix}$$

$$= \begin{pmatrix} 1 \cdot f(x) & +1 \cdot 4f'(x)h & +1 \cdot 4^2 f''(x)\frac{h^2}{2} & +1 \cdot 4^3 f'''(x)\frac{h^3}{3!} + ... \\ -6 \cdot f(x) & -6 \cdot 2f'(x)h & -6 \cdot 2^2 f''(x)\frac{h^2}{2} & -6 \cdot 2^3 f'''(x)\frac{h^3}{3!} + ... \\ +8 \cdot f(x) & +8 \cdot 1f'(x)h & +8 \cdot 1^2 f''(x)\frac{h^2}{2} & +8 \cdot 1^3 f'''(x)\frac{h^3}{3!} + ... \\ -3 \cdot f(x) & & & \end{pmatrix}.$$

Observe that the right hand side simplifies to $4f'''(x)h^3 + ...$ so we get a generalized third derivative here when we divide by $4h^3$. (In other words, $\frac{1}{4}\tilde{\Delta}_3$ corresponds to a generalized third derivative.) This calculation gives an explanation for the conditions (3). To make this more explicit we will consider a very basic generalized derivative, the dth Peano derivative. Say that f has an dth Peano derivative at x, denoted $f_d(x)$ if for x fixed there are numbers $f_0(x), f_1(x), ..., f_{d-1}(x)$ such that

$$f(x+h) = f_0(x) + f_1(x)h + \cdots \\ + f_{d-1}(x)\frac{h^{d-1}}{(d-1)!} + f_d(x)\frac{h^d}{d!} + o(h^d) \qquad (6)$$

as $h \to 0$. This notion, the notion of approximation near x to dth order, is very basic indeed. In fact a serious argument can be made that this is a better definition of higher order differentiation than the standard one which defines the dth derivative iteratively with the dth derivative $f^{(d)}(x)$ being defined as the ordinary first derivative of $f^{(d-1)}(x)$. I will have more to say about this later on. Note that $f_0(x) = f(x)$ if and only if f is continuous at x, and that $f_1(x)$ and $f'(x)$ are identical by definition. Also,

$$\text{if } f^{(d)}(x) \text{ exists, then } f_d(x) \text{ exists and equals } f_d(x). \qquad (7)$$

(This is the content of Peano's powerful, but little known version of Taylor's Theorem.) This implication is irreversible for every $d \geq 2$. I will give a simple example of this in my next lecture.

Here is a connection between the two generalizations considered so far.

$$\text{If } f_d(x) \text{ exists, then every } D_{uw}f(x) \text{ with } w \text{ and } u \\ \text{satisfying (3) exists and equals } f_d(x). \qquad (8)$$

This fact follows immediately from substituting (1) into (3) and interchanging the order of summation. This calculation is the basis for the conditions (3). The implication (8) is also irreversible; for example, $D_{\left(\frac{1}{2},-\frac{1}{2}\right)(1,-1)}|x| = 0$ at $x = 0$, but $|x|$ is not differentiable there. (This D_{uw} is the symmetric first derivative.) Nevertheless, the implication (8) is "more reversible" than is the implication (7). Much of what I have to say in these lectures will be about the possibilities of partial converses to implication (8), to generalizations of implication (8), and to partial converses of those generalizations.

From a certain numerical analysis point of view, the ordinary first derivative is "bad", while the symmetric first derivative is "good". More specifically, for the ordinary first derivative we have the expansion

$$\begin{aligned}\frac{f(x+h)-f(x)}{h} &= \frac{f(x)+f'(x)h+f_2(x)\frac{h^2}{2}+o\left(h^2\right)-f(x)}{h} \\ &= f'(x)+E,\end{aligned}$$

where

$$E = f_2(x)\frac{h}{2}+o(h);$$

while for the symmetric first derivative of the same scale we have

$$\frac{f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)}{h} = \frac{1}{h}\left\{\begin{array}{c} f\left(x+\frac{h}{2}\right) \\ -f\left(x-\frac{h}{2}\right)\end{array}\right\}$$

$$\begin{aligned}&= \frac{1}{h}\left\{\begin{array}{ccccc} f(x) & +f'(x)\frac{h}{2} & +f_2(x)\frac{h^2}{8} & +f_3(x)\frac{h^3}{48} & +o\left(h^3\right) \\ -f(x) & +f'(x)\frac{h}{2} & -f_2(x)\frac{h^2}{8} & +f_3(x)\frac{h^3}{48} & +o\left(h^3\right)\end{array}\right\} \\ &= f'(x)+F,\end{aligned}$$

where

$$F = f_3(x)\frac{h^2}{24}+o\left(h^2\right).$$

Comparing the error terms E and F gives a measure of the goodness of approximation of the first derivative at x by these two difference quotients for fixed small h decidedly favoring the symmetric first derivative.

There are many generalized first derivatives. One way to make new ones is to "slide", that is to add a constant multiple of h to every base point. Thus, sliding by $\frac{1}{2}h$,

$$\left\{x-\frac{1}{2}h, x+\frac{1}{2}h\right\} \to \left\{x-\frac{1}{2}h+\frac{1}{2}h, x+\frac{1}{2}h+\frac{1}{2}h\right\} = \{x, x+h\},$$

changes the first symmetric derivative into the ordinary first derivative. So sliding does not preserve the numerical version of "good" that we have just mentioned. Furthermore, the function $|x|$ does have a first symmetric derivative at $x = 0$, but it fails to be differentiable there, so sliding does not preserve differentiability either. Sliding *does* preserve the property of being a derivative. To see this, note that if c is any real number, then the equation system

$$\sum_{i=0}^{d+e} w_i \left(u_i + c\right)^j = \begin{cases} 0 & j = 0, 1, ..., d-1 \\ \\ d! & j = d \end{cases} .$$

is equivalent to the equation system (3), so that $\sum_{i=0}^{d+e} w_i f\left(x + u_i h\right)$ is a generalized dth derivative if and only if $\sum_{i=0}^{d+e} w_i f\left(x + \left(u_i + c\right) h\right)$ is.

There is a probably apocryphal story from the middle of the nineteenth century that when H. A. Schwarz was examining a Ph.D. candidate who wrote the formula for a general quadratic in two variables as

$$ax^2 + bxy + cy^2 + dx + ey + f, \text{ where } e \text{ is not the natural logarithm,}$$

he insisted that the student be more precise by inserting "necessarily" after "not". The symbol e here stands for excess and is a nonnegative integer. The minimum number of points for a dth derivative is

first derivative	–	2 points
second derivative	–	3 points
...		...
dth derivative	–	$d + 1$ points

and these correspond to the cases of zero excess, $e = 0$. When $e = 0$, the system of equations (3) may be thought of as a system of $d + 1$ equations in the $d + 1$ unknowns $w_0, w_1, ..., w_d$, which could be written in the matrix notation

$$Aw = \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \\ d! \end{pmatrix} .$$

Then the matrix A is a Vandermonde matrix and it is easy to see that there is a unique solution for w [1, p. 182]. But if the excess $e > 0$, then there are many w's for a given u. Some examples of first derivative with positive excess $e = 1$ are

$$\lim_{h\to 0} \frac{(1-2a)\, f\,(x+(a+1)\, h) + 4af(x+ah) - (1+2a)\, f\,(x+(a-1)\, h)}{2h} \tag{9}$$

for any constant a. The case $a = 1/\sqrt{3}$ is special.

2 Lecture 2

There are three "best" generalized first derivatives based on three points. The first is the derivative defined by relation (9) with $a = \frac{1}{\sqrt{3}}$. Performing Taylor expansions here yields

$$\frac{1}{2h}\left[\left(1-\frac{2}{\sqrt{3}}\right) f\left(x+\left(\frac{1}{\sqrt{3}}+1\right) h\right) + \frac{4}{\sqrt{3}} f(x+(\frac{1}{\sqrt{3}})h) \right.$$
$$\left. -\left(1+\frac{2}{\sqrt{3}}\right) f\left(x+\left(\frac{1}{\sqrt{3}}-1\right) h\right)\right] \approx f'\,(x) + .016 f_4\,(x)\, h^3 + o\left(h^3\right).$$

The second, given here together with its approximating relation is

$$\frac{f\left(z+h/\sqrt{3}\right) + \omega^2 f\left(z+\omega h/\sqrt{3}\right) + \omega f\left(z+\omega^2 h/\sqrt{3}\right)}{\sqrt{3}h}$$
$$\approx f'\,(z) + .008 f_4\,(z)\, h^3 + o\left(h^3\right),$$

where ω is a complex cube root of 1 so that $\omega^2 + \omega + 1 = 0$, and the third may be written scaled as

$$\frac{-5f(x+2h) + 32f\,(x+h) - 27f(x-\frac{2}{3}h)}{h} \approx f'\,(x) + .056 f_4\,(x)\, h^3 + o\left(h^3\right).$$

How are these best? For any generalized first derivative the dominant error term in approximating f' by the difference quotient will have the form $Cf_{n+1}h^n$, where $n \geq 2$ is an integer. We will say that a difference quotient with dominant error term $Cf_{n+1}h^n$ is better than one with dominant error term $Df_{m+1}h^m$ if $n > m$, or if $n = m$ and $C < D$. All three have an error term that is one power of h higher than might be expected from linear algebra considerations of the system (3). The first minimizes truncation error if we restrict ourselves to the general setting of these talks, that of real-valued functions of a real variable. The second only makes sense in the setting of analytic functions of a complex variable. It is perhaps out of place here, but arises naturally from the algebra involved with attempts to minimize. The third example, which can be written in equivalent but unscaled form as

$$\frac{-5f(x+6h)+32f\,(x+3h)-27f(x-2h)}{3h},$$

is almost as good as the other two. It is the best you can do if you are limited to functions defined as tabulated values. There are other notions of goodness that might be considered. These are discussed in reference [6]. In reference [8], best additive linear generalized first and second derivatives for each excess e are found.

Let us return to the ordinary dth derivative $f^{(d)}$, the Peano dth derivative f_d, and the generalized (linear) dth derivatives $D^d_{uw}f$. Explicitly, $f_d\,(x)$ is defined as

$$f_d\,(x)=\lim_{h\to 0}\frac{f\,(x+h)-f_0\,(x)-f'\,(x)\,h-f_2\,(x)\,\dfrac{h^2}{2!}\ldots-f_{d-1}\,(x)\,\dfrac{h^{d-1}}{(d-1!)}}{\dfrac{h^d}{d!}}.$$

As we mentioned above, for each fixed x we have the implications

$$\exists f^{(d)}\,(x)\Longrightarrow\exists f_d\,(x) \tag{10}$$

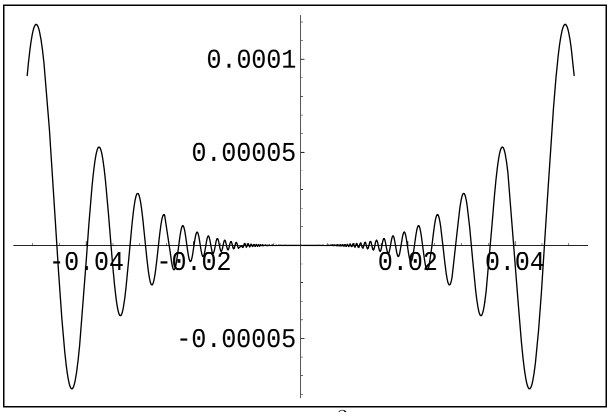
and

$$\exists f_d\,(x)\Longrightarrow\exists D^d_{uw}f\,(x)\,. \tag{11}$$

The first implication is reversible when $d=1$ since the definitions of $f'\,(x)=f^{(1)}\,(x)$ and $f_1\,(x)$ coincide. The second implication is reversible for the same reason when $d=1$ and uw is either $(c,0)\,(1/c,-1/c)$ or $(0,-c)\,(1/c,-1/c)$, for some positive c. Except for these trivial cases, no other implication is reversible. Here is an example which shows why the first implication is irreversible. Let $f\,(x)=\begin{cases} x^3\sin(1/x) & x\neq 0\\ 0 & x=0\end{cases}$. Then for nonzero h,

$$f\,(0+h)=0+0h+0\frac{h^2}{2}+h^2\,(h\sin\,(1/h))$$

where the last term is $o\left(h^2\right)$ since $h\sin\,(1/h)\to 0$ as $h\to 0$. In particular, $f_2\,(0)=0$.



The function $x^3 \sin(1/x)$

However, by direct calculation, $f'(x) = \begin{cases} -x\cos(\frac{1}{x}) + 3x^2\sin(\frac{1}{x}) & x \neq 0 \\ 0 & x = 0 \end{cases}$

so that

$$f''(0) = \lim_{h\to 0} \frac{-h\cos(1/h) + 3h^2\sin(1/h)}{h} = \lim_{h\to 0} -\cos(1/h)$$

does not exist. An instance where the second implication is irreversible given above was the function $|x|$ which has first symmetric derivative at $x = 0$, but which does not have a first Peano derivative there. Another interesting example showing the second implication to be irreversible is the function $\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$. This has a Schwarz generalized second derivative equal to 0 at $x = 0$ (see definition (1)), but certainly is not well approximated by a quadratic polynomial near $x = 0$, so that sgn does not have a second Peano derivative at $x = 0$.

My thesis advisor, Antoni Zygmund, used to say that the purpose of counterexamples is to lead us to theorems, so since the exact converses of implications (10) and (11) are false, let us consider the following partial converses.

$$\text{For all } x \in E, \exists f_d(x) \Longrightarrow \text{for almost every } x \in E, \exists f^{(d)}(x) \qquad (12)$$

$$\text{For all } x \in E, \exists D^d_{uw} f(x) \Longrightarrow \text{for almost every } x \in E, \exists f_d(x). \qquad (13)$$

Except for the trivial $d = 1$ case, converse (12) is still false. H. William Oliver proved that for any interval I and any open, dense subset O of I, there is a function f with the second Peano derivative $f_2(x)$ existing for every $x \in I$, the ordinary second derivative $f''(x)$ existing for all $x \in O$, and $f''(x)$ existing at no point of $I \setminus O$.This example is especially daunting since the measure of $I \setminus O$ can be close to the measure of I. In fact, if $\epsilon > 0$ is given, let $\{r_n\}$ be the rational numbers of I. Then the open set $\left(r_1 - \frac{\epsilon}{4}, r_1 + \frac{\epsilon}{4}\right) \cup \left(r_2 - \frac{\epsilon}{8}, r_2 + \frac{\epsilon}{8}\right) \cup \cdots$ has Lebesgue measure $\le \frac{\epsilon}{2} + \frac{\epsilon}{4} + \cdots = \epsilon$ and its intersection with I is open and dense in I. But Oliver was Zygmund's student, so he found and proved this partial converse.

Theorem 1 (Oliver). *If the dth Peano derivative $f_d(x)$ exists at every point of an interval I and if $f_d(x)$ is bounded above, then the ordinary dth derivative $f^{(d)}(x)$ also exists at all points of I.*

So the hoped for converse (12) is false. But the second converse is true! Almost all of the deep results that I will discuss involve this converse or analogues of it. The first big paper here is the 1936 work of Marcinkiewicz and Zygmund that I have already mentioned. It is "On the differentiability of functions and summability of trigonometric series." [9] In this paper, along with many other things, is proved the validity of the second converse (13) for the special cases when D_{uw} is the dth symmetric Riemann derivative,

$$D_d f(x) = \lim_{h \to 0} \frac{1}{h} \sum_{n=0}^{d} (-1)^n \binom{d}{n} f(x + (d/2 - n) h).$$

Later on, in my thesis, the second converse (13) is proved for every D_{uw}.[1]

Let $d \ge 2$. Define an equivalence relation on the set of all the dth derivatives that we have defined here by saying that two derivatives are equivalent if for any measurable function, the set where one of the derivatives exists and the other one does not has Lebesgue measure zero. Then there are only two equivalence classes. One is the singleton consisting of only the ordinary dth derivative. The other contains everything else, the dth Peano derivative and every D_{uw} satisfying the conditions (3). Earlier I said that a case could be made that the standard dth derivative should be the dth Peano derivative, polynomial approximation to order d, rather than the ordinary dth derivative. The fact that ordinary differentiation stands alone from this almost everywhere point of view is one of the reasons for this.

3 Lecture 3

In 1939 Marcinkiewicz and Zygmund proved that for all d, the existence of the dth symmetric Riemann derivative,

$$D_d f(x) = \lim_{h\to 0} \frac{1}{h^d} \sum_{n=0}^{d} (-1)^n \binom{d}{n} f\left(x + \left(\frac{d}{2} - n\right) h\right) = \lim_{h\to 0} \frac{1}{h^d} \Delta_d \quad (14)$$

for every x in a set E implies the existence of the Peano derivative $f_d(x)$ at almost every $x \in E$.[9] One of the pieces of their proof is a sliding lemma.

Lemma 2. *If*

$$\Delta_d = \sum_{n=0}^{d} (-1)^n \binom{d}{n} f\left(x + \left(\frac{d}{2} - n\right) h\right) = O\left(h^d\right) \quad (15)$$

for all $x \in E$, then any of its slides is also $O\left(h^d\right)$ at almost every point of E.

For example, from

$$\Delta_2 = f(x+h) - 2f(x) + f(x-h) = O\left(h^2\right) \quad (16)$$

for all $x \in E$, follows

$$f(x+2h) - 2f(x+h) + f(x) = O\left(h^2\right) \quad (17)$$

for a. e. $x \in E$.

We cannot drop the "a. e." from the lemma's conclusion. For $h > 0$, $\operatorname{sgn}(0+h) - 2\operatorname{sgn}(0) + \operatorname{sgn}(0-h) = 0 = O\left(h^2\right)$, but $\operatorname{sgn}(0+2h) - 2\operatorname{sgn}(0+h) + \operatorname{sgn}(0) = -1 \neq O\left(h^2\right)$ as $h \to 0$.

If we start from the existence of the second symmetric derivative on a set, so that

$$f(x+h) - 2f(x) + f(x-h) = D_d f(x) h^2 + o\left(h^2\right)$$

there, then pass to the weaker condition (16), and then pass to condition (17), we lose information at both steps. At the first step we lose the limit itself and at the second step we lose a set of measure 0. The second loss obviously does not matter, since we are only aiming for an almost everywhere result. The first loss will not matter either, since we may lower our sights from the goal of achieving the existence of $f_d(x)$ a. e. to that of achieving merely

$$f(x+h) = f(x) + f_1(x)\,h + \cdots + f_{d-1}(x)\,\frac{h^{d-1}}{(d-1)!} + O(h^d) \qquad (18)$$

a. e. This is because Marcinkiewicz and Zygmund prove a lemma stating that if condition (18) holds on a set, then we also have the existence of the Peano derivative $f_d(x)$ itself after discarding a subset of zero measure. Such a lemma is a generalization of the fact that Lipschitz functions are differentiable almost everywhere.

The second big idea is the introduction of the derivatives

$$\tilde{D}_d f(x) = c_d \lim_{h\to 0} \frac{\tilde{\Delta}_d}{h^d},$$

based on the $d+1$ points

$$x + 2^{d-1}h, x + 2^{d-2}h, \ldots, x + 2h, x + h, x.$$

I mentioned how these differences $\tilde{\Delta}_d$ were formed in the first lecture. (See formula (5) for $\tilde{\Delta}_3$.) The constants c_d are required to normalize the last equation of conditions (3), for example, we showed in lecture 1 that $c_3 = \frac{1}{4}$. It turns out that it is much easier to prove their next lemma,

$$\tilde{\Delta}_d = O\left(h^d\right) \text{ on } E \implies \text{formula (18) holds a. e. on } E, \qquad (19)$$

than to prove the same thing starting from $\Delta_d = O\left(h^d\right)$ on E.

Already I have mentioned enough results to produce the $d = 2$ case of the theorem. Suppose $D_2 f(x)$ exists for all $x \in E$. By the sliding lemma we know that condition (17) holds a. e. on E. But Δ_2 and $\tilde{\Delta}_2$ coincide, so by implication (19), formula (18) holds a. e. on E and this is enough to get $f_2(x)$ a. e. on E. For $d \geq 3$, there is only one more step needed. To completes the logical flow of their argument, Marcinkiewicz and Zygmund show that $\tilde{\Delta}_d$ is a linear combination of slides of Δ_d. An example of this last lemma is the equation

$$\begin{aligned}
&f(x+4h) - 6f(x+2h) + 8f(x+h) - 3f(x) \\
&= f(x+4h) - 3f(x+3h) + 3f(x+2h) - f(x+h) \\
&+ 3\left\{f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)\right\}
\end{aligned}$$

or

$$\tilde{\Delta}_3 = (\Delta_3 \text{slid up by } h) + 3\Delta_3.$$

The first generalization of the work of Marcinkiewicz and Zygmund is this.

Theorem 3. *If a generalized dth derivative $D_{uw}f$ exists on a set E, then the dth Peano derivative of f exists for a. e. $x \in E$.*

To prove this, one more big idea is needed. I will restrict my discussion to applying that idea to one very simple case not covered by Marcinkiewicz and Zygmund.

Assume that the derivative

$$Df(x) = \lim_{t \to o} \frac{f(x+3t) + f(x+2t) - 2f(x+t)}{3t}$$

exists for all $x \in E$. We will show that $f'(x)$ exists for a. e. $x \in E$.

Note that D is indeed a first derivative with excess $e = 1$, since $\frac{1}{3} + \frac{1}{3} + \left(\frac{-2}{3}\right) = 0$ and $\frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 2 + \left(\frac{-2}{3}\right) \cdot 1 = 1$.

Taking into account what we know about Marcinkiewicz and Zygmund's approach, we start with

$$f(x+3t) + f(x+2t) - 2f(x+t) = O(t) \tag{20}$$

on E. The plan is to show that

$$F(x) = \int_a^x f(t)\,dt$$

satisfies a. e. on E,

$$F(x+2h) - 2F(x+h) + F(x) = O\left(h^2\right). \tag{21}$$

We will suppress various technical details, such as showing that our hypothesis (20) allows us to assume that f is actually locally integrable in a neighborhood of almost every point of E. Then by the Marcinkiewicz and Zygmund result, F will have two Peano derivatives a. e. on E. It is plausible and actually not hard to prove from this that $F' = f$ a. e. and that f has one less Peano derivative than F does almost everywhere. So our goal is reduced to proving relation (21).

The idea is to exploit the noncommutativity of the operations of integrating and sliding. The substitution $u = x+bt, du = bdt$ allows $\int_0^h f(x+bt)\,dt =$

$\int_x^{x+bh} f(u)\frac{du}{b} = \frac{1}{b}\left[F(x+bh) - F(x)\right]$. So integrating our assumption (20), taking $\int_0^h O(t)dt = o\left(h^2\right)$ into account gives

$$\frac{1}{3}\left[F(x+3h) - F(x)\right] + \frac{1}{2}\left[F(x+2h) - F(x)\right] - 2\left[F(x+h) - F(x)\right] = O\left(h^2\right)$$

or

$$2F(x+3h) + 3F(x+2h) - 12F(x+h) + 7F(x) = O\left(h^2\right) \tag{22}$$

a. e. on E. So integrating and then sliding by h gives

$$2F(x+4h) + 3F(x+3h) - 12F(x+2h) + 7F(x+h) = O\left(h^2\right) \tag{23}$$

for a. e. $x \in E$.

Now start over from assumption (20). This time slide by t,

$$f(x+4t) + f(x+3t) - 2f(x+2t) = O(t)$$

and then integrate to get

$$\frac{1}{4}\left[F(x+4h) - F(x)\right] + \frac{1}{3}\left[F(x+3h) - F(x)\right] - 2\frac{1}{2}\left[F(x+2h) - F(x)\right] = O\left(h^2\right)$$

or

$$3F(x+4h) + 4F(x+3h) - 12F(x+2h) + 5F(x) = O\left(h^2\right) \tag{24}$$

for a. e. $x \in E$.

The crucial point is that the left hand sides of equations (23) and (24) are different. The rest is arithmetic. Think of the three expressions on the left sides of relations (22), (23), and (24) as three vectors in the span of five basis elements $v_i = F(x+iH), i = 0, 1, ..., 4$. We can eliminate v_3 and v_4 by taking the following linear combination of these three vectors.

$$\begin{pmatrix} 6\left\{2v_4 + 3v_3 - 12v_2 + 7v_1\right\} \\ -4\left\{3v_4 + 4v_3 - 12v_2 + 5v_0\right\} \\ -\left\{2v_3 + 3v_2 - 12v_1 + 7v_0\right\} \end{pmatrix} =$$

$$\begin{pmatrix} 12 \\ -12 \\ \ \end{pmatrix} v_4 + \begin{pmatrix} +18 \\ -16 \\ -2 \end{pmatrix} v_3 + \begin{pmatrix} -72 \\ +48 \\ -3 \end{pmatrix} v_2 + \begin{pmatrix} +42 \\ \ \\ +12 \end{pmatrix} v_1 + \begin{pmatrix} \ \\ -20 \\ -7 \end{pmatrix} v_0$$

$$= -27\left(F\left(x+2h\right)-2F\left(x+h\right)+F\left(x\right)\right)=O\left(h^{2}\right)$$

for a. e. $x \in E$. We have reached the desired goal of relation (21).

The next step in the program is to move to L^p norms, $1 \le p < \infty$. Here is a restatement of the Marcinkiewicz and Zygmund Theorem.

Theorem 4. *If*

$$\left|\sum_{n=0}^{d}(-1)^{n}\binom{d}{n}f\left(x+\left(\frac{d}{2}-n\right)h\right)-D_{d}f\left(x\right)h^{d}\right|=o\left(h^{d}\right)$$

for every $x \in E$, then for a. e. $x \in E$,

$$\left|f\left(x+h\right)-\left(f\left(x\right)+f_{1}\left(x\right)h+\cdots+f_{d}\left(x\right)\frac{h^{d}}{d!}\right)\right|=o\left(h^{d}\right).$$

And here is the L^p version.

Theorem 5. *Let $p \in [1,\infty)$. If*

$$\left(\frac{1}{h}\int_{0}^{h}\left|\sum_{n=0}^{d}(-1)^{n}\binom{d}{n}f\left(x+\left(\frac{d}{2}-n\right)t\right)-D_{d}^{p}f\left(x\right)t^{d}\right|^{p}dt\right)^{1/p}=o\left(h^{d}\right),$$

for every $x \in E$, then for a. e. $x \in E$,

$$\left(\frac{1}{h}\int_{0}^{h}\left|f\left(x+t\right)-\left(f_{0}^{p}\left(x\right)+f_{1}^{p}\left(x\right)t+\cdots+f_{d}^{p}\left(x\right)\frac{t^{d}}{d!}\right)\right|^{p}dt\right)^{1/p}=o\left(h^{d}\right).$$

This is true. Even a more general version, the obvious analogue of Theorem 3, which we will call Theorem 3^p, is true. When I went to Zygmund in 1963 for a thesis problem, I wanted to try the then unsolved question of the almost everywhere convergence of Fourier series. My plan was to try to find a counterexample, an L^2 function with almost everywhere divergent Fourier series. But Zygmund wanted me to try to prove Theorem 4. Luckily, I worked on Theorem 4 and Lennart Carleson worked on (and proved) the almost everywhere convergence of Fourier series of L^2 functions. Otherwise, I still might not have a Ph.D.

You cannot prove Theorem 4 by just changing norms. A lot of the lemmas go through easily, but the sliding lemma does not. At least I have never been able to find a direct proof of it. In fact, here is an important open question; perhaps the best one I will offer in these lectures.

Problem 6. *It is true that if* $\int_0^h |f(x+t) - f(x-t)|\, dt = O\left(h^2\right)$ *for all* $x \in E$*, then* $\int_0^h |f(x+2t) - f(x)|\, dt = O\left(h^2\right)$ *for almost every* $x \in E$*. Prove this by a "direct" method.*

The use of the word direct is a little vague and will be explicated in a minute. The method that I used to prove Theorem 4 immediately required Theorem 3. Furthermore, once I had proved Theorem 3, Theorem 3^p was no harder to prove than was Theorem 4. Let me show you where Theorem 3 comes into play. The $d = 1$ case is easy, but already the $d = 2$ case is not. Let $p = 1$, so our assumption implies

$$\frac{1}{h}\int_0^h |f(x+t) + f(x-t) - 2f(x)|\, dt = O\left(h^2\right).$$

From this and the triangle inequality follows

$$\frac{1}{h}\int_0^h \{f(x+t) + f(x-t) - 2f(x)\}\, dt = O\left(h^2\right),$$

or, equivalently,

$$F(x+h) - F(x-h) - 2f(x)\, h = O\left(h^3\right) \tag{25}$$

where $F(x+h) = \int_x^{x+h} f(u)\, du$. If we can remove the $2f(x)\, h$ term, then we be in the realm of generalized derivatives of the form $D_{uw}F$. The obvious thing to do is to write formula (25) with h replaced by $2h$ and then subtract twice formula (25) from that. We get.

$$F(x+2h) - F(x-2h) - 2f(x)\, 2h - 2\left(F(x+h) - F(x-h) - 2f(x)\, h\right)$$

$$= F(x+2h) - 2F(x+h) + 2F(x-h) - F(x-2h) = O\left(h^3\right).$$

Note that $1-2+2-1 = 0, 1\cdot 2 - 2\cdot 1 + 2\cdot(-1) - 1\cdot(-2) = 0, 1\cdot 2^2 - 2\cdot 1^2 + 2\cdot(-1)^2 - 1\cdot(-2)^2 = 0$, and $1\cdot 2^3 - 2\cdot 1^3 + 2\cdot(-1)^3 - 1\cdot(-2)^3 = 12 = 2\cdot 3!$, so that the left side corresponds to the third derivative $D_{(2,1,-1,-2)(1/2,-1,1,-1/2)}F$. Thus arose the necessity of extending Marcinkiewicz and Zygmund's theorem to Theorem 3. Now it should be a little more clear what I mean by a direct proof in the problem given above: do not prove it by passing to the indefinite integral F, then proving that F has two Peano derivatives, etcetera.

4 Lecture 4

Stefan Catoiu, Ricardo Ríos, and I recently proved an analogue of the theorem of Marcinkiewicz and Zygmund. The work will soon appear in the Journal of the London Mathematical Society.[4]

Assume throughout our discussion that

$$x \neq 0.$$

Making the substitution $qx = x + h$ and noting that $h \to 0$ if and only if $q \to 1$ either as

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \text{ or as } \lim_{q \to 1} \frac{f(qx) - f(x)}{qx - x}.$$

Already something new happens if we replace the additively symmetric points $x + h, x - h$ used to define the symmetric first derivative with the multiplicatively symmetric points $qx, q^{-1}x$ and consider

$$D_{1,q} f(x) = \lim_{q \to 1} \frac{f(qx) - f\left(q^{-1}x\right)}{qx - q^{-1}x}.$$

Similarly,

$$D_{2,q} f(x) = \lim_{q \to 1} \frac{f(qx) - (1+q) f(x) + q f\left(q^{-1}x\right)}{q(qx - x)^2}$$

is something new. Notice that in both cases there would be nothing at all defined if x were 0.

These are the first two instances of q-analogues of the generalized Riemann derivatives. Using the polynomial expansion of f about x allows us to write

$$f(qx) = f(x + (q-1)x) = f(x) + f_1(x)(qx - x) + f_2(x)\frac{(qx - x)^2}{2} + \cdots$$

and thereby quickly prove that if at each point x, the first Peano derivative $f_1(x)$ exists, then at that point $D_1^q f$ does also and $f_1(x) = D_1^q f(x)$; similarly the existence of $f_2(x)$ implies $f_2(x) = D_2^q f(x)$. As you would expect, the pointwise converses to these two implications are false. In our paper, we show that the converses do hold on an almost everywhere basis. We do this by more or less establishing the analogue of every step of the Marcinkiewicz and Zygmund proof.

The reason that we don't go on to prove the q-analogue of Theorem 3 is that we cannot find an analogue of the non-commutivity of integrating and sliding idea. This seems to be a great impediment to further progress. I will spend a lot of time in this lecture on posing questions. So far, we can do nothing in the L^p setting. I will specialize to $p = 1$, because in all the classical work involving the additive generalized derivatives nothing special happed to distinguish the $p = 1$ cases from any corresponding cases with other finite values of p, and things are a little easier to write here.

Problem 7. *Prove: If at every $x \in E$,*

$$\frac{1}{q-1}\int_1^q \left|f(px) - f\left(p^{-1}x\right) - \left(px - p^{-1}x\right) D^1_{1,q} f(x)\right| \frac{dp}{p} = o\left((qx - x)\right),$$

then at a. e. $x \in E$,

$$\frac{1}{h}\int_0^h |f(x+t) - f(x)|\, dt = O(h).$$

Problem 8. *Prove: If at every $x \in E$,*

$$\frac{1}{q-1}\int_1^q \left|f(px) - (1+p) f(x) + pf\left(p^{-1}x\right) - p\left(px - p^{-1}x\right)^2 D^1_{2,q} f(x)\right| \frac{dp}{p}$$
$$= o\left(\left(qx - q^{-1}x\right)^2\right),$$

then at a. e. $x \in E$,

$$\frac{1}{h}\int_0^h \left|f(x+t) - f(x) - f_1^p(x)\, t\right| dt = O\left(h^2\right).$$

A q-version of an L^p sliding lemma would immediately solve question 7. It would also give a major boost to solving lots of other problems. Here is what a general sliding lemma might say. Again, I will keep $p = 1$ for simplicity of statement and I really do not think that the value of p makes any difference.

Problem 9. *Prove: If at every $x \in E$,*

$$\frac{1}{q-1}\int_1^q \left|\sum w_k(p) f\left(p^{n_k}x\right)\right| \frac{dp}{p} = o\left((q-1)^\alpha\right),$$

then, for any fixed c, at a. e. $x \in E$,

$$\frac{1}{q-1}\int_1^q \left|\sum w_k(p) f\left(p^{n_k - c}x\right)\right| \frac{dp}{p} = o\left((q-1)^\alpha\right).$$

The very first instance of Problem 9 would immediately resolve Problem 7, so I will state it separately.

Problem 10. *Prove: If at every* $x \in E$,

$$\frac{1}{q-1}\int_1^q \left|f(px) - f\left(p^{-1}x\right)\right| \frac{dp}{p} = O(q-1),$$

then at a. e. $x \in E$,

$$\frac{1}{q-1}\int_1^q \left|f\left(p^2x\right) - f(x)\right| dt = O(q-1).$$

Incidentally, the change of variable $p^2 \to p$, $\frac{dp}{p} \to \frac{dp}{p}$ shows that the last equation is equivalent to

$$\frac{1}{q-1}\int_1^q |f(px) - f(x)|\, dt = O(q-1),$$

since $q^2 - 1$ is equivalent to $q-1$ for q near 1. The reason that I think that even Problem 10 is very hard is that, as I mentioned before, even in the simplest additive case the L^p sliding lemma is by no means a staightforward extension of the usual L^∞ sliding lemma. In fact, Problem 10 is truly an open question, the issue is not just to find a "direct" proof.

Problem 11. *Fix* $\alpha > 0$. *Prove: If at every* $x \in E$,

$$\frac{1}{h}\int_0^h |f(x+t) - f(x-t)|\, dt = O(h^\alpha),$$

then at a. e. $x \in E$,

$$\frac{1}{h}\int_0^h |f(x+2t) - f(x)|\, dt = O(h^\alpha).$$

As I discussed in the third lecture, there is a very tricky indirect proof of this when $\alpha = 1$ using the methods of my thesis. If it could be solved for general α, way, the method of the solution might very well lead to a full L^p additive sliding lemma.

Problem 12. *Fix* $\alpha > 0$.*Prove: If at every* $x \in E$,

$$\frac{1}{h}\int_0^h \left|\sum w_k f(x + u_k t)\right| dt = O(h^\alpha),$$

then, for any fixed c, *at a. e.* $x \in E$,

$$\frac{1}{h}\int_0^h \left|\sum w_k f(x + (u_k - c)t)\right| dt = O(h^\alpha).$$

Note that there are no conditions on the w_k's and u_k's here.

If in 1964 I could have solved Problem 12, I would have simply copied the Marcinkiewicz and Zygmund's proof, lemma by lemma into L^p format without any need for the generalized derivatives D_{uw}.

Let us return now to the question of using generalized derivatives for approximation. This time we will consider multilinear, in particular bilinear, generalized derivatives. The idea is to approximate a derivative of order d by means of finite sums of the form

$$\sum w_{ij} f\left(x + u_i h\right) f\left(x + v_i h\right).$$

Proceeding formally from Taylor series expansions we get this expression equal to

$$\left(\sum\nolimits_{i,j} w_{ij}\right) f^2 + \left(\sum\nolimits_i w_{ij} u_i + \sum\nolimits_j w_{ij} v_j\right) f f_1 h + \begin{pmatrix} \left(\sum\nolimits_i w_{ij} u_i^2 + \sum\nolimits_j w_{ij} v_j^2\right) f f_2 \\ + \left(\sum\nolimits_{i,j} w_{ij} u_i v_j\right) f_1^2 \end{pmatrix} \frac{h^2}{2} + \cdots$$

To see how to make use of such an expression, we will given an example involving analytic functions of a complex variable. This is the second, and last, time in these lectures that I depart from talking about real functions of a real variable.

Let $\omega = \left(1 + \sqrt{3} i\right)/2$ be a cube root of -1, so that $\omega^2 = \omega - 1$ and $\omega^3 = -1$. If f is analytic, Taylor expansion yields

$$f(z+h) f(z+\omega h) - f(z-h) f(z-\omega h)$$

$$= 2(\omega+1) f(z) f'(z) h + (2\omega - 1) f'(z) f''(z) h^3$$

$$+ (10 f'' f''' - f f^{(5)}) (w-2) h^5/60 + O\left(h^7\right)$$

or

$$\frac{f(z+h) f(z+\omega h) - f(z-h) f(z-\omega h)}{2(\omega+1) f(z) h} = \tag{26}$$

$$f'(z) + \frac{(2\omega-1) f'(z) f''(z)}{2(\omega+1) f(z)} h^2 + \frac{(10 f'' f''' - f f^{(5)}) (w-2)}{120 (\omega+1) f(z)} h^4 + O\left(h^6\right)$$

Note that the main (first) error term involves only derivatives of order ≤ 2. If one had used a linear generalized derivative instead to generate a similar estimate, the main error term would necessarily involve $f^{(n)}(x)$ with $n \geq 3$. So it is possible that this kind of approximation could prove useful for functions with small low order derivative and large high order derivatives. This is the best non-linear example I've found so far, but I don't think it compares very favorably with, say, this excellent linear approximation for $f'(x)$,

$$\frac{-f\left(x+\frac{3}{2}h\right)+27f\left(x+\frac{1}{2}h\right)-27f\left(x-\frac{1}{2}h\right)+f\left(x-\frac{3}{2}h\right)}{24h} =$$

$$f'(x) - \frac{3}{640}f^{(5)}(x)\,h^4 + O\left(h^6\right),$$

which requires only 4 function evaluations.[8] So the problem here is to improve substantially on the example I have given.

Problem 13. *Find at least one bilinear (or multilinear, or even more general) numerical difference quotient which approximates the first derivative in a way that compares favorably with the known good linear quotients.*

The last topic will be generalizations of the mean value theorem. Let

$$Df(x) = \lim_{h\to 0} h^{-1}\sum_{n=0}^{1+e} w_n f(x+u_n h)$$

be a generalized first derivative, so that $\sum w_n = 0$ and $\sum w_n u_n = 1$. Recall that $u_0 > u_1 > \cdots$. Let A and $H > 0$ be fixed real numbers. Assume that $Df(x)$ exists for every $x \in [A, A+H]$. The question is, does there necessarily exist a number $c \in (A, A+H)$ such that

$$Df(c) = \frac{\sum_{n=0}^{1+e} w_n f(A+u_n H)}{H}. \tag{27}$$

When $uw = (1,0)(1,-1)$, the derivative $D = D_{uw}$ reduces to the ordinary derivative and formula (27) reduces to the usual mean value formula. The standard hypothesis is that f be continuous on $[A, A+H]$ and differentiable on $(A, A+H)$.

The first non-standard derivative that comes to my mind is the first symmetric derivative, $uv = \left(\frac{1}{2}, -\frac{1}{2}\right)(1,-1)$. Here we quickly run into trouble. For example, the absolute value function is continuous on $[-1,2]$ and its symmetric derivative exists (and is equal to $\operatorname{sgn} x$) at each point of $(-1, 2)$. Nevertheless, with $A = -1$ and $H = 3$, we have

$$\frac{|A+H| - |A|}{H} = \frac{2-1}{3} = \frac{1}{3},$$

which is not equal to any of $-1, 0, 1$, the three possible values of $\operatorname{sgn} x$.

To sharpen the question, let us make three definitions.

Definition 14. Say that D_{uw} has the weak mean value property if the existence of f' on $[A, A+H]$ implies that there is a $c \in (A, A+H)$ so that the mean value formula (27) holds.

Definition 15. Say that D_{uw} has the mean value property if the continuity of f on $[A, A+H]$ and the existence of D_{uw} on $(A, A+H)$ implies that there is a $c \in (A, A+H)$ so that the mean value formula (27) holds.

Definition 16. Say that D_{uw} has the strong mean value property if the existence of D_{uw} on $[A, A+H]$ implies that there is a $c \in (A, A+H)$ so that the mean value formula (27) holds.

When Roger Jones and I noticed the counterexample involving the symmetric derivative and the absolute value function, our response was to prove that the symmetric derivative has the weak MVP. We went on to prove that many other generalized first derivatives also have this property. However, some first derivatives do not have the MVP. In our paper [7], we give a sufficient condition, which is also necessary when there are two or three base points. A complete classification has not been achieved.

If D_{uw} has the MVP, then it has the weak MVP. Ordinary differentiation has all three MVPs. Even though symmetric differentiation does have the weak MVP, the last example shows that it does not have either the MVP, nor the strong MVP. Since the symmetric derivative is so close to ordinary differentiation, Jones and I conjectured that there might very well be no first derivatives other than the ordinary derivative with either the MVP or the strong MVP. Just as I was giving these lectures Ricardo Ríos and I discovered that there are some derivatives with the MVP. In particular, we have the ironical result that of all two base point generalized derivatives, the symmetric derivative stands out as the only one not having the MVP. Explicitly, the generalized first derivative,

$$\lim_{h \to 0} \frac{f(x+h) - f(x+ah)}{h - ah}, a \neq 1$$

has the mean value property for each value of a except $a = -1$.

Problem 17. *Which generalized derivatives have the weak mean value property, which have the mean value property, and which have the strong mean value property?*

References

[1] Ash, J. M., Generalizations of the Riemann derivative, *Trans. Amer. Math. Soc.*, **126**(1967), 181–199.

[2] ––, A characterization of the Peano derivative, *Trans. Amer. Math. Soc.,* **149**(1970), 489–501.

[3] ––, Very generalized Riemann derivatives, generalized Riemann derivatives and associated summability methods, *Real Analysis Exchange,* **11**(1985-86), 10–29.

[4] Ash, J. M., Catoiu, S., and Ríos, R., On the nth quantum derivative, *J. London Math. Soc.,* **66**(2002), 114-130.

[5] Ash, J. M., Erdös P., and Rubel, L. A., Very slowly varying functions, *Aequationes Math.,* **10**(1974), 1–10.

[6] Ash, J. M., Jones, R., Optimal numerical differentiation using three function evaluations, *Math. Comp.,* **37**(1981), 159–167.

[7] ––, Mean value theorems for generalized Riemann derivatives, *Proc. Amer. Math. Soc.,* **101**(1987), 263–271.

[8] Ash, J. M., Janson, S., and Jones, R., Optimal numerical differentiation using n function evaluations , *Estratto da Calcolo,* **21**(1984), 151–169.

[9] Marcinkiewicz, J., and Zygmund, A., On the differentiability of functions and summability of trigonometric series, *Fund. Math.,* **26**(1936), 1–43.

Author's address:

J. Marshall Ash
Department of Mathematical Sciences
DePaul University
2320 North Kenmore Ave.
Chicago, IL, 60614, U. S. A.
E-mail: `kmash@condor.depaul.edu`