A Comparison of Two Elementary Approximation Methods

HARVEY DIAMOND
West Virginia University
Morgantown, WV 26506

LOUISE RAPHAEL
Howard University
Washington, DC 20059

1. Introduction There are two familiar methods one learns in calculus for approximating the value of a differentiable function near one or more data points where the function is known and/or easy to evaluate. The first method is linear interpolation and the second is the differential approximation. A natural question about the the two methods is: "Which is better?" This question was posed by Leon Henkin during a visit to a seminar on curve fitting given by one of the authors in the Summer Mathematics Institute for talented minority students at Berkeley.

In what follows, we study this question from several viewpoints, using only elementary methods. What we find interesting is that, starting with an intuitive but mathematically ill-formulated question, we are lead to a fairly rich investigation touching on several mathematical approaches, including a result on existence/uniqueness that applies to convex functions, an exact/local analysis based on the quadratic case, and a discussion of numerical methods.

2. Formulation and preliminary observations The function f is assumed to be differentiable. We begin by setting out the two alternatives:
Differential approximation: Given $f(a), f'(a)$ we approximate

$$f(x) \cong f(a) + f'(a)(x-a) \equiv f_D(x). \tag{1}$$

Linear interpolation: Given $f(a), f(b)$, we approximate

$$f(x) \cong f(a) + \frac{f(b)-f(a)}{b-a}(x-a) \equiv f_L(x). \tag{2}$$

FIGURE 1 illustrates $f(x)$ and its two approximations.

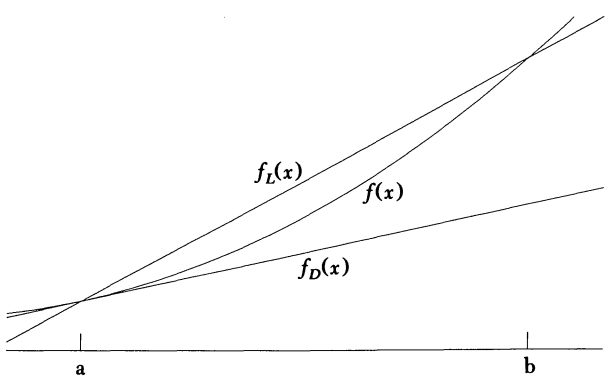


FIGURE 1

The *approximation error* of an approximation method at a particular value of x is simply the difference of $f(x)$ and its approximate value, in this case either $f(x)-f_D(x)$ or $f(x)-f_L(x)$. The *size* of the approximation error is its absolute value. We would be entitled to say that one method is better than the other if for all $a \le x \le b$ its approximation error had smaller size than that of the other method. We will see shortly, however, that this situation never occurs in our problem. The approximation method having the error of smaller size depends on the value of x, and one must try to say something about *where* (i.e. for which values of x) one method is better than another, and for which functions.

We begin with some simple observations. If x is near a, f_D is better, while if x is near b, f_L is better. In fact, the only way this could not be true is if the two approximations are identical. To explain this, note first that both approximations are the same at $x=a$. Near $x=b$, if $f_D(b) \neq f(b)$ then f_L is better near b, since $f_L(b)=f(b)$. On the other hand, if $f_D(b)=f(b)$ then both approximations agree at $x=a$ and $x=b$ and hence are identical since they are linear functions. Next, using the definition of the derivative, one can show that $f_D(x)$ is the best linear approximation of $f(x)$ near $x=a$. Consider all lines through the point $(a, f(a))$ that have the form $y=f(a)+m(x-a)$. The error in approximating $f(x)$ using such a line, at $x \neq a$, is then

$$f(x)-(f(a)+m(x-a))=(x-a)\left\{\frac{f(x)-f(a)}{(x-a)}-m\right\},$$

so that the term in brackets goes to zero as x goes to a (and hence gives the smallest size error) if, and only if, m is equal to $f'(a)$. Thus either f_D is better than f_L near a or else f_L also has derivative $f'(a)$, in which case it is identically equal to f_D.

Having shown that f_D is better near $x=a$ and f_L is better near $x=b$, the question of which approximation method is better becomes, "For what values of x is f_D better than f_L and for what values is the opposite true?" In turn, the answer to this question hinges on the answer to the question "At what points are the approximation errors of the two methods the same size?" We note that at any point where the approximation errors have the same size, the approximation errors must be of opposite sign or else (again) the approximations are identical. This occurs because if the approximation errors are the same at some point then $f_D=f_L$ at that point and, along with $f_D(a)=f_L(a)$, which is always true, we see that both approximations correspond to the same line. Thus if the functions f_D and f_L are not identical, the set of points at which the approximation errors have the same size is therefore precisely where the approximation errors sum to zero, which is precisely the solution set of the equation

$$f(x)-f_D(x)+f(x)-f_L(x)=0 \quad \text{or} \tag{3}$$

$$f(x)=[f_D(x)+f_L(x)]/2. \tag{4}$$

A geometric interpretation of (4) is shown in FIGURE 2, namely the solution set of (4) is the set of points where the line $y=[f_D(x)+f_L(x)]/2$ intersects the curve $y=f(x)$. The label c in FIGURE 2 is the location of x at the point of intersection in the interval (a, b).

3. Qualitative results Continuing with the graph in FIGURE 2, we observe that $f(x)$ is concave up and there is exactly one value of x, $x=c$ in the figure, where the sizes of the two approximation errors are the same. The following theorem shows that this is true for any function that is strictly concave up or concave down and has two continuous derivatives.

THEOREM 1. *Let $f''(x)$ be continuous and either strictly positive or strictly negative for all x in some open interval I. Then for any a, b in I, f_D and f_L cannot be identical, and, aside from the point $x = a$, there is exactly one other point in I, lying between a and b, at which the approximation errors have the same size.*

Proof. Motivated by equation (3), we define

$$g(x) = f(x) - f_D(x) + f(x) - f_L(x).$$

First, f_D and f_L cannot be identical for if they are, $g(a) = 0$, $g(b) = 0$, and $g'(a) = 0$. The first two facts (by Rolle's theorem) imply the existence of a point α between a and b for which $g'(\alpha) = 0$. Since $g'(a) = 0$ there exists another point β between a and α at which $g''(\beta) = 0$. But $g''(x) = 2f''(x)$ since f_D and f_L are linear, so $f''(\beta) = 0$. This contradicts the hypothesis that $f''(x)$ is either strictly positive or strictly negative in I.

Since f_D and f_L are not identical, the approximation errors have the same size if, and only if, (3) is satisfied, i.e., $g(x) = 0$. Now $g(a) = 0$ is always true. If g has two other distinct roots in I there are, by Rolle's theorem, at least two distinct points in I where $g'(x)$ is zero, and so at least one point in I where $g''(x) = 0$, again contradicting the hypothesis on $f''(x)$. Thus g has at most one other root in I, but since we showed that f_D and f_L are not identical, by previous arguments there is at least one point between a and b where $g(x) = 0$. Thus, aside from $x = a$, there is a unique root of g and it lies between a and b. This proves the theorem.

We might ask what can happen if the hypotheses of the theorem do not hold and f has one inflection point, say $x = c$. In this case f_D and f_L can be identical, but if not, there is, interestingly, still only one point where the approximation errors have the same size.

THEOREM 2. *Let $f''(x)$ be continuous and suppose that the closed interval $I = [\alpha, \beta]$ contains an interior point c with $f''(c) = 0$ and $f''(x) > 0$ for $x > c$, $f''(x) < 0$, for $x < c$, where $x \in I$ is assumed in each case. Then the following holds:*

i) *There are infinitely many pairs a, b, with $a < c < b$, for which $f_D(x)$ and $f_L(x)$ are identical.*

ii) *If f_D and f_L are not identical for a particular choice of $a, b \in I$ then there is exactly one point between a and b at which the approximation errors have the same size.*

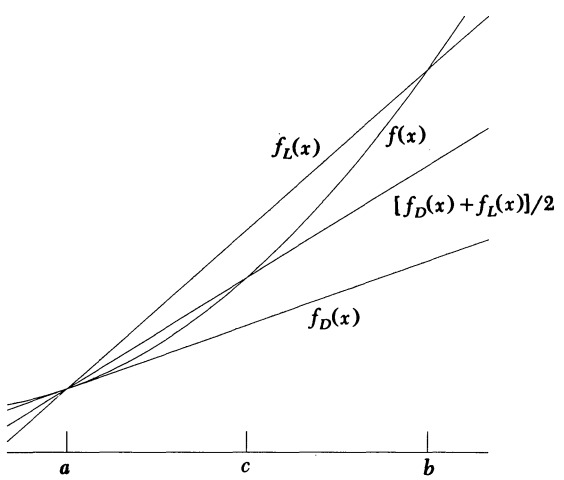


FIGURE 2

Proof. We sketch a proof of part i) with the aid of FIGURE 3. The idea is that, using values of a close to but less than c, the graph of $f_D(x)$ (the tangent from $(a, f(a))$), will intersect the graph of $f(x)$ at some $x = b$ with $b > c$ and $b \in I$. Then f_L will be identical to f_D. To effect this idea as a proof, we begin by explicitly exhibiting the dependence of f_D on the point a by using the notation $f_D(x; a)$. Now we can show, using Rolle's theorem, that

$$f_D(x; c) < f(x) \quad \text{for } x > c, \tag{5}$$

i.e., the tangent through the turning point is below the right, concave-up, portion of the curve. Similarly, we can show that

$$\text{if } a < c,\ f_D(x; a) > f(x) \quad \text{for } x \le c, \tag{6}$$

i.e., a tangent constructed on the left, concave-down, portion of the curve is above the left portion of the curve. Now consider $f_D(\beta; a)$ as a function of a; it is clearly a continuous function of a. When $a = c$, we have $f_D(\beta; c) < f(\beta)$, by virtue of (5), so for $a < c$ and $|a - c|$ sufficiently small, $f_D(\beta; a) < f(\beta)$ holds by continuity. On the other hand, $f_D(c; a) > f(c)$ by virtue of (6). The last two inequalities imply, by the intermediate value theorem, the existence of a point b with $c < b < \beta$, such that $f_D(b; a) = f(b)$, completing part i).

The proof of part ii) is more difficult and is left to the reader, although elementary techniques are still adequate. FIGURE 4 provides a "generic" picture of why the theorem is true. The unique value of x between a and b where the approximation errors have the same size is denoted by c; we observe that at $x = c'$ the approxima-

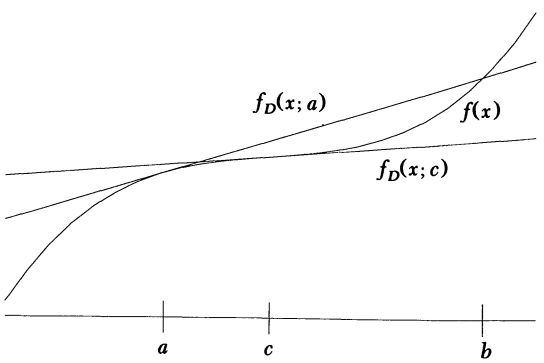


FIGURE 3

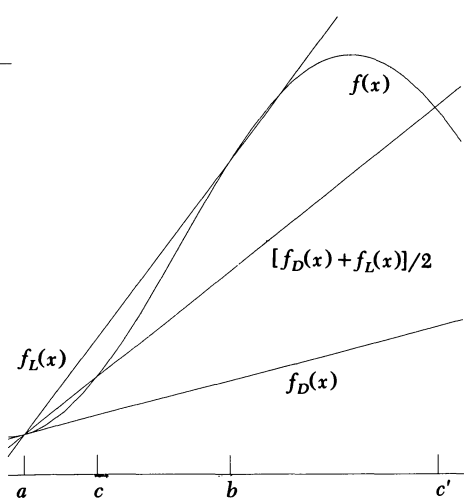


FIGURE 4

tion errors also have the same size, but this value of x lies outside the interval $[a, b]$.

4. An important special case In investigating our problem further, we will completely and explicitly solve the problem in the case when f is quadratic. "Solving the problem" means finding the solution of (3), which we know from Theorem 1 is unique, but in any case will be observed to be unique from the calculations below.

If f is quadratic, say $f(x) = px^2 + qx + r$, then $f(x)$ can be written as $f(x) = f(a) + f'(a)(x-a) + p(x-a)^2$.

This representation can most easily be obtained from Taylor's theorem after noting that $p = f''(a)/2$, since the coefficient of x^2 must be p. It then follows from (1) that

$$f(x) - f_D(x) = p(x-a)^2. \tag{7}$$

On the other hand, with f quadratic and f_L linear, $f(x) - f_L(x)$ is a quadratic function that is zero at $x = a$ and $x = b$ so we must have

$$f(x) - f_L(x) = p(x-a)(x-b). \tag{8}$$

The constant factor can be identified as p since subtracting a linear function from f does not change the coefficient of x^2.

At the x we seek, the solution of (3), the sum of (7) and (8) is zero,

$$p\left[(x-a)^2 + (x-a)(x-b)\right] = 0 \text{ or } p(x-a)(2x-a-b) = 0$$

with solutions $x = a$ and $x = (a+b)/2$. Thus for quadratic functions, differentials are better until halfway between a and b and then the linear interpolant is better. This is a pleasing result and provides a good rule of thumb, as we briefly explain next.

The solution of the quadratic case provides a local analysis of the general problem, applicable when a and b are close together and the hypotheses of Theorem 1 hold for some open interval I containing $[a, b]$. Under these conditions, f is well-approximated by a quadratic function and $x = (a+b)/2$ will be a good approximation to the unique solution of (3). A precise formulation and proof of this fact is somewhat difficult, however, and will not be attempted here.

5. Numerical considerations As discussed in section 3, if $f(x)$ is concave up or down, the problem of determining where each approximation method is better than the other reduces to finding the unique solution of (3) for values of x in the interval (a, b). If we define

$$g(x) \equiv f(x) - f_D(x) + f(x) - f_L(x), \tag{9}$$

then we are seeking the solution of $g(x) = 0$ in the interval (a, b). It turns out that Newton's method is guaranteed to converge to the desired solution if we use $x_0 = b$ as the initial guess for the root of g.

THEOREM 3. *If $f(x)$ satisfies the hypotheses of Theorem 1, then Newton's method applied to the function $g(x)$ in (9), namely, the iteration*

$$x_{n+1} = x_n - g(x_n)/g'(x_n),$$

with starting value $x_0 = b$, will converge to the unique solution $x = c$ of $g(x) = 0$ in the interval (a, b).

Proof. We will not provide a proof here. Results of this sort for Newton's method are well known; see for instance [1, p. 62]. The result basically follows from the observations depicted graphically in FIGURE 5:

i) The sequence $\{x_n\}$ is a decreasing sequence with $x_n > c$. This comes from the concavity of f (and therefore g) in the interval (a, b).

ii) As a decreasing sequence bounded from below, $\{x_n\}$ has a limit, say x^*, and this limit must satisfy $x^* = x^* - g(x^*)/g'(x^*)$ so that $g(x^*) = 0$. It follows that $x^* = c$.

Finally, we briefly consider the possibility of solving (4) by iteration. If we solve (4) for x on the (linear) right side, we can obtain an equation of the form $x = h(x)$. It is tempting to try to solve this equation numerically by the iteration $x_{n+1} = h(x_n)$, or what is the same thing, the iteration

$$[f_D(x_{n+1}) + f_L(x_{n+1})]/2 = f(x_n). \tag{10}$$

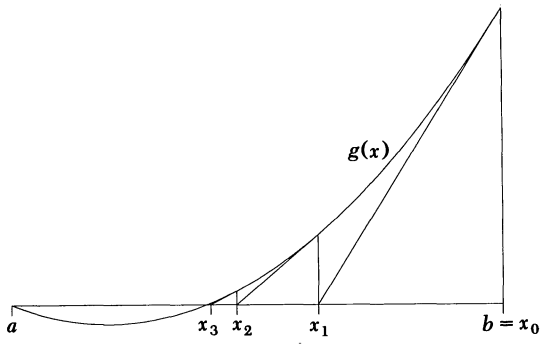


FIGURE 5

This turns out to be rather dangerous, even in the simplest case, where f satisfies the hypotheses of Theorem 1. In the situation depicted in FIGURE 2, for instance, it is easy to see graphically that for the iteration (10), $x = a$ is a stable fixed point, whereas the solution we seek, $x = c$, is an unstable fixed point. Two other possibilities, both with $f'(a) > 0$, $f''(x) < 0$, are depicted in FIGURE 6. If $f'(x) > 0$, $a < x < b$ (see FIGURE 6a), then the solution $x = c$ is a stable fixed point and any x_0 in the interval (a, b) results in a monotone sequence converging to $x = c$. Perhaps most interesting is the case depicted in FIGURE 6b, in which f has a maximum to the left of $x = c$. In this case $x = a$ is an unstable fixed point, but $x = c$ may or may not be stable. In fact this generic situation leads, under certain conditions, to the well-known chaotic maps of an interval onto itself, of which the most famous example is

$$x_{n+1} = 4\lambda x_n(1 - x_n), \tag{11}$$

where chaos occurs for certain values of the parameter λ lying between 0 and 1. (See [2] for a recent discussion at an elementary level.) Indeed, this example arises if we take $f(x) = x - x^2$, $a = 0$ and b a parameter, so that $f_D(x) = x$, $f_L(x) = (1 - b)x$ and (10) becomes (11) with $4\lambda = (1 - b/2)^{-1}$.

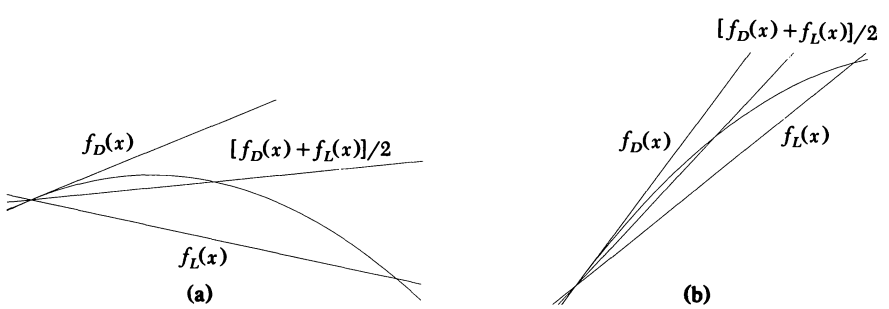


FIGURE 6

6. Concluding remarks We have shown in the preceding discussion, how a simple question concerning the comparison of linear interpolation and the differential approximation leads to some interesting mathematics, using only elementary methods at the calculus level. We discovered classes of functions for which the answer to the question could be well-characterized qualitatively; and obtained a nice exact answer in the case of quadratic functions, which also serves as a local approximation. Finally we showed that Newton's method can be reliably used to find the "cross-over" point in the case of concave up/down functions, while the method of iteration, naively applied, can converge, diverge, or lead to chaotic behavior even in the simplest examples.

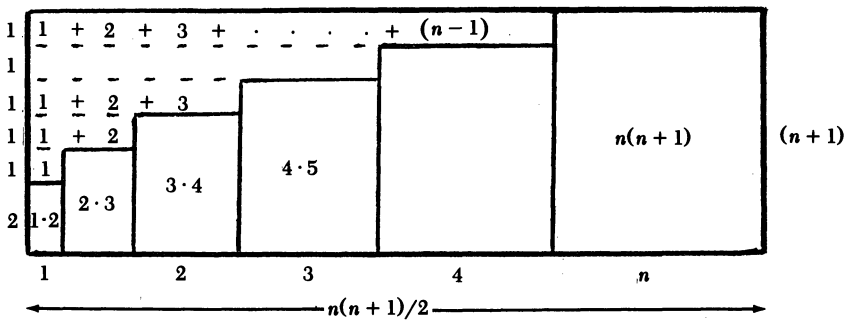
The authors were partially supported by an Air Force Office of Scientific Research contract 91MN062.

REFERENCES

1. Kendall E. Atkinson, *An Introduction to Numerical Analysis*, 2nd edition, John Wiley & Sons, New York, 1989.
2. Robert L. Devaney, The orbit diagram and the Mandelbrot set, *The College Math. J.* 22 (1991), 23–38.

Proof without Words
Sum of Products of Consecutive Integers

$$1\cdot 2+2\cdot 3+3\cdot 4+\cdots+n(n+1)=n(n+1)(n+2)/3.$$



$$1\cdot 2+2\cdot 3+3\cdot 4+\cdots+n(n+1)$$

$$=[n(n+1)/2][n+1]-\left[\begin{array}{l}1\\+1+2\\+1+2+3\\+\\\vdots\\+1+2+3+\cdots+(n-1)\end{array}\right.$$

$$=\frac{n(n+1)^2}{2}-\frac{(n-1)n(n+1)}{6}=\frac{n(n+1)(n+2)}{3}.$$

—James O. Chilaka
Long Island University
Greenvale, NY 11548