Hermite versus Simpson: the Geometry of Numerical Integration

Andy Long and Cliff Long

May 1, 2008

Abstract

An examination of current calculus and numerical analysis texts shows that when composite numerical integration rules are developed, the link to parametric curve fitting (what we call the *geometry* of an integration rule) is frequently ignored, or at least not exploited to its fullest.

In particular, the popular Simpson's rule is not mined for its close connection with parametric curve fitting. When any relationship is developed it is often done so badly: the composite Simpson's Rule is usually derived as the integral of a continuous piecewise quadratic spline interpolant of the function values or data points, but without even slope-continuity. We provide a better geometry to associate with Simpson's Rule.

Beyond that, however, we suggest that the emphasis on Simpson's Rule is outdated: we prefer another rule, the Corrected-Trapezoid method (which we call *Hermite's Rule*, although others might prefer *Hermite Cubic quadrature* [8] (p. 161)), as it has several pedogogical and logistical advantages over Simpson's Rule (especially a more interesting and useful geometry). Hermite's Rule is more general than Simpson's Rule, as it is based on capturing derivative information (true or approximate) as well as function information.

We derive one approximation to Hermite's Rule whose error term is slightly better than that of Simpson's Rule, and compare the integration schemes on a number of standard calculus test functions while focusing on the geometric aspects of each method.

keywords: Simpson's Rule, Hermite's Rule, Hermite interpolation, finite differences, parametric curve fitting, splines.

1 Introduction

Consider $n+1$ data values $y_0, y_1, \ldots, y_n$ associated with $n+1$ equally spaced points, $x_0, x_1, \ldots, x_n$. Think of the y_i as values of a function f at the points x_i. How might you approximate the definite integral

$$I = \int_{x_0}^{x_n} f(x)dx?$$

The "justly famous"[6] Simpson's Rule may come to mind: certainly many calculus students will think of it, because Simpson's Rule is usually the end-all of integration rules in introductory calculus class. As for numerical analysis students, "[w]here would any book on numerical analysis be without Mr. Simpson and his 'rule'?"[10]

Mr. Simpson was Thomas Simpson (1710-1761), "...an able and self-taught English mathematician,... author of several text-books, ... active in perfecting trigonometry as a science."[1] This author of text books would perhaps be pleased to be so well-remembered by so many of them. Simpson is credited with "[t]he first application of the Newton-Raphson[2] process to the solution of transcendental equations" and was certainly an early advocate of the arithmetic mean, as we can see from the title of his paper "An attempt to show the advantage arising by taking the mean of a number of observations, in practical astronomy". Did he go beyond the mean to the weighted average known as Simpson's rule? There is apparently no evidence that he developed the numerical integration rule which bears his name.

Simpson's rule, although tremendously popular, is not mined for its close connection with parametric curve fitting (which we call the *geometry* of an integration rule). In fact, when the relationship is made clear - when the geometry is connected with the composite Simpson's Rule - it is often done so in a "bad" way: the composite rule is usually derived by pasting together the elemental quadratic interpolants, leading thus to an integral of a continuous

[1]This and other quotes concerning Simpson are from Florian Cajori's "A History of Mathematics"[4].

[2]aka "Newton's method", for root-finding

piecewise quadratic spline interpolant of the function values or data points, without even slope-continuity. The interpolant is continuous (C^0), but not differentiable (C^1).

We demonstrate first of all that one may associate a better geometry with the composite Simpson's Rule, that of a C^1 cubic spline interpolant. But we have a more ambitious goal than simply putting Simpson's Rule in closer contact with a good parameterization: we propose that the focus on Simpson's Rule endemic to calculus, engineering mathematics, and numerical analysis texts might be better placed on another rule (which we call *Hermite's Rule*). This rule is also associated with a C^1 cubic spline interpolant, but has several additional useful pedogogical and logistical features.

Hermite's Rule itself requires derivative information. This may dissuade the casual reader from reading any further. We hope, however, to reach the imaginative reader, who will suspend disbelief long enough to follow us through to the various approximations to Hermite's Rule based on finite difference approximations to the derivatives.

George Polya, in How to Solve It[9], posits that the solution to a specific problem may be best obtained by considering a more general problem. In many cases requiring numerical integration we have only function values (or data values), so why, then, introduce derivatives? This argument is really specious: all of numerical integration is based on numerical differentiation, in some sense, it is just that we ordinarily are not so bold as to act as if we have the derivative values at hand. Let us do so, nonetheless: the advantage is that we will derive a method that works if one is given the derivative information, and furthermore develop an approximation to the method which beats Simpson's Rule on "level ground" (at least for a suite of calculus test functions), even when the true derivative is unknown.

2 Simpson's Rule

Simpson's Rule is undoubtably king of the numerical integration rules: open a calculus, numerical analysis, or engineering mathematics textbook and check the index - there it is. For an integrable function f on the closed interval $[a,b]$, Simpson's Rule makes the approximation

$$\int_a^b f(x)dx \approx \frac{h}{3}(y_0 + 4y_1 + y_2), \tag{1}$$

where $h = (b-a)/2$; $x_0 = a$, $x_1 = a+h$, and $x_2 = b$; and $y_i = f(x_i)$ (or is a data value corresponding to the point x_i). If f is C^4, then the error term (defined as the exact integral minus the approximation) can be written as

$$-\frac{h^5}{90} f^{(4)}(\mu),$$

with $\mu \in (a,b)$. In short, Simpson's Rule divides the interval into two equal subintervals, computes the function at the three boundary points, and then takes a weighted average, giving an error of the fifth-order in h proportional to the fourth derivative of f. The pair of adjacent subintervals $[x_0, x_1]$ and $[x_1, x_2]$ constitutes what is popularly known as a *panel*.

This elemental Simpson's Rule leads to the so-called composite Simpson's rule, created by dividing the interval $[a,b]$ into a large number of panels (that is, an even number n of equal subintervals), invoking the elemental rule (1) on each panel, and then adding up the results. The composite Simpson's rule approximation is

$$S = \frac{h}{3}(y_0+4y_1+2y_2+4y_3+2y_4+\ldots+2y_{n-4}+4y_{n-3}+2y_{n-2}+4y_{n-1}+y_n), \quad (2)$$

where $h = (b-a)/n$ and $y_i = f(a+ih)$. If this were a dance, we might call it the 14242-step; Burden and Faires[3] call it "...the most frequently used general-purpose quadrature algorithm."

It can be shown that the composite Simpson's Rule is of fourth-order in h. More precisely, the composite Simpson's Rule has an error term of the form

$$-\frac{(b-a)}{180} h^4 f^{(4)}(\mu), \quad (3)$$

with $\mu \in (a,b)$ ([3], p. 186). The fact that the error is proportional to the fourth derivative implies that Simpson's Rule gives exact results for polynomials of degree three or less (as their fourth derivatives are identically zero). The order of the stepsize h in the error term is one measure of a rule, while the degree of polynomial that it integrates exactly is another. We say that the composite Simpson's Rule is a fourth-order method having degree of precision three.

3 What's Wrong with Simpson's?

Although Simpson's rule is frequently presented in beginning calculus as an improvement over the rectangle and trapezoidal rules for integral approximation, we have three logistical or pedogical concerns about the resulting composite rule:

- it only applies to an even number of subintervals;
- it gives the (mistaken) impression that some interior points are twice as important as others; and
- the geometry that is frequently used to derive the composite Simpson's Rule (e.g. [3], Figure 2.7, p. 185) involves piecewise parabolic data fitting which doesn't even have slope continuity at the joints (Figure (1), upper left).

3.1 Odd subintervals out

This first concern may seem rather frivolous, especially if you're a calculus professor accustomed to a formula for every function. But in today's high-tech society, problems requiring approximate integrals often stem from automatically sampled input data rather than from functions defined by formulas. Hence one can never assume that an even number of subintervals will be provided or available. The subintervals themselves are frequently equal, however, making simple composite rules attractive. Thus we are faced with a situation in which we want to use a rule like the composite Simpson's rule, but would rather not be bothered with the even/odd headache.

A recent consulting experience associated with the design of instrumentation for determining the quality of automotive windshield glass brought this issue to our attention. In the quality assessment process laser scanners provide glass surface information (e.g. angles of reflection) at fixed time intervals; certain integrals naturally arose, which we wanted to calculate on an ongoing basis; the question thus became how, and although we started with Simpson's Rule, we migrated to Hermite's initially simply to avoid the problem of odd intervals.

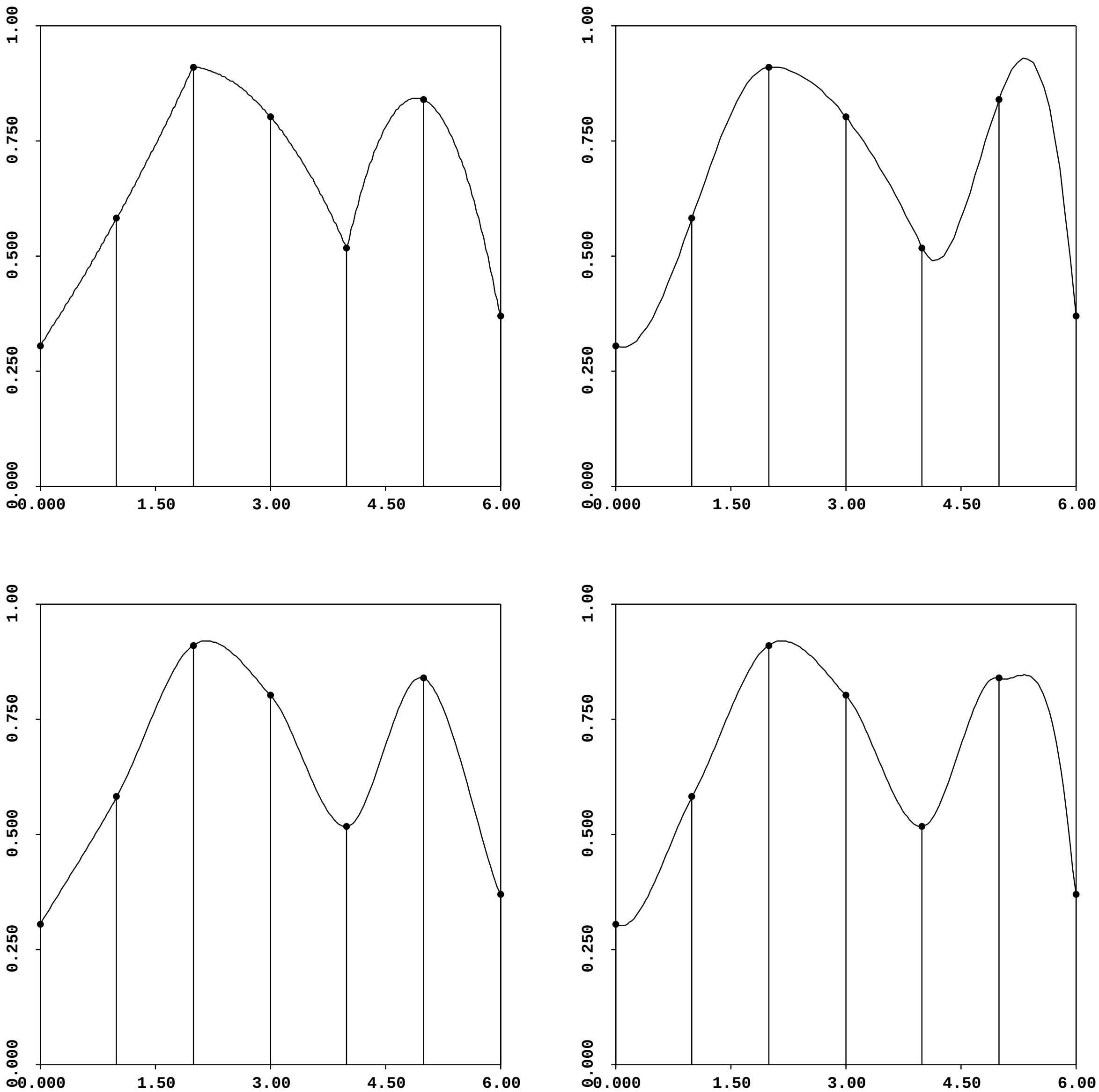

Figure 1: The top two functions are geometries associated with the composite Simpson's Rule on three panels (C^0 quadratic patches, and a C^1 cubic spline with the left-slope given by the five-point rule (eq. (9)), the same slope used by H_5). The bottom two functions are geometries associated with approximations H_3 (left) and H_5 (right) to Hermite's Rule.

Many authors go to the trouble of talking their readers through the problem of applying Simpson's rule in the event of an odd number of intervals, often encouraging the use of "Simpson's 3/8 rule" (which is defined on three-subinterval panels) on one end to make up the difference. This means, at the very least, that some authors believe that Simpson's or Simpson's-like Rules are necessary and useful. We took a different approach when we found ourselves in this same situation: we adapted Simpson's for an odd number of subintervals as follows, and so arrived at our first example of a Hermite rule (which we call H_3).

If n is odd, then the first $n-1$ subintervals and the last $n-1$ subintervals can be handled with Simpson's rule, and the results averaged. This neglects, however, the fact that the end subintervals are added in only once, rather than twice as are all the internal subintervals. To correct for that, we need to add in an additional approximation for the two end subintervals before averaging. We approximated the function on the end subintervals by using a quadratic function that interpolated the data values (2 conditions) and also matched an approximation to the slope at the internal node (given by the centered finite difference approximation). For the first interval $[x_0, x_1]$, for example, this function is

$$q_0(x) = y_1 + \frac{y_2 - y_0}{2h}(x - x_1) + \frac{y_2 - 2y_1 + y_0}{2h^2}(x - x_1)^2.$$

These two functions, integrated over their respective subintervals, give contributions of

$$\frac{h}{12}(5y_0 + 8y_1 - y_2)$$

and

$$\frac{h}{12}(-y_{n-2} + 8y_{n-1} + 5y_n).$$

Averaging these various elements via the matrix product

$$\frac{1}{2}\begin{bmatrix} \frac{h}{3} & \frac{h}{3} & \frac{h}{12} & \frac{h}{12} \end{bmatrix} \begin{bmatrix} 1 & 4 & 2 & 4 & 2 & \cdots & 4 & 2 & 4 & 1 & \\ & 1 & 4 & 2 & 4 & \cdots & 2 & 4 & 2 & 4 & 1 \\ 5 & 8 & -1 & & & & & & & & \\ & & & & & & & & -1 & 8 & 5 \end{bmatrix}$$

yields the weights

$$\frac{h}{24}\begin{bmatrix} 9 & 28 & 23 & 24 & 24 & \cdots & 24 & 24 & 23 & 28 & 9 \end{bmatrix}; \qquad (4)$$

the inner product of these weights with the vector of points $\left[\begin{array}{cccc} y_0 & y_1 & \dots & y_n \end{array}\right]$ gives the approximation which we have since named Hermite's Rule H_3 (we will show why in a moment); and H_3 ultimately inspired us to write this paper.

3.2 Who says odd points are better than even points?

It is clear why students may get the impression that Simpson's rule values the odd numbered points twice as much as the even points: 4 is, after all, twice 2. This impression can be rectified by rewriting (2) as

$$S = h\sum_{i=0}^{n} y_i + \frac{h}{3}(-2y_0 + y_1 - y_2 + y_3 - y_4 + \dots - y_{n-1} + y_n - 2y_n). \qquad (5)$$

But this form, too, provokes troubling questions: how was the phase of this -2+1-1+1-1-wave chosen? Why does it multiply even-numbered points by -1 and odd-numbered points by 1? Does this sinusoidal wave, superimposed on the data, solicite contrast information across the domain of the function? The authors of Numerical Recipes[10] have a theory all their own: "Many people believe that the wobbling alternation somehow contains information about the integral of their function which is not apparent to mortal eyes."

One way to explain away this wave is to claim that it is merely an artifact of the method of construction (pasting together elemental units to generate a composite rule). However, as we are about to show, the composite Simpson's rule can be derived without pasting anything at all.

3.3 The geometry of Simpson's Rule

All the usual numerical integration rules are associated with some type of approximating function. The rectangle rules are so named because the data are treated as though they are a sample of a step-function, and the integral of the step function (a sum of areas of rectangles) serves as an approximation to the integral of the function itself. The trapezoidal rule

$$T = \sum_{i=0}^{n-1} \frac{h}{2}(y_{i+1} + y_i) = h\sum_{i=0}^{n} y_i - \frac{h}{2}(y_0 + y_n), \qquad (6)$$

(the simplest *Newton-Cotes* rule) is derived by creating a linear spline between data points: that is, on the i^{th} subinterval a linear function (the linear

Lagrange interpolant[3]) joins the points (x_i, y_i) and (x_{i+1}, y_{i+1}), and the composite trapezoidal rule is derived by joining these linear segments together to form a C^0 spline, then integrating the result (a sum of areas of trapezoids).

Simpson's Rule is also a member of the Newton-Cotes family of intergration formulae, each of which is derived (in elemental form) by integrating Lagrange interpolating polynomials. This is the positive aspect of the relationship between parametric curve fitting and Newton-Cotes numerical integration rules. Unfortunately, this sensible treatment does not usually extend to the development of the composite integration rules.

The elemental Simpson's Rule is derived by integrating the Lagrange quadratic interpolant on a panel, rather than on a subinterval. That is, we consider the panel composed of the three points $\{(x_0, y_0), (x_1, y_1), (x_2, y_2)\}$, for which the Lagrange quadratic interpolant is given by

$$q(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2.$$

Integrating q from x_0 to x_2 gives the elemental Simpson's Rule (eq. (1)).

Moving on to the composite rule, wherein we divide the interval $[a, b]$ into n panels, the inclination is to simply calculate the Lagrange quadratic for each panel and paste them together to give rise to the geometry of the composite rule. This results again in a C^0 spline, but is unsatisfactory due to the slope discontinuities (as shown, for example, in Figure (1)). To interpolate data points or function values geometrically with piecewise curves which are not even slope continuous sends a strange message to our students (especially to any with interests in engineering or computer-aided design, for example).

Note that our concern about geometry is somewhat pedogogical (some might even say pedantic!): first of all, the geometry does not affect the numerical answer; and secondly, Simpson's Rule is not tied to a single geometry, although some authors tend to act as though it were. For example, one textbook gushes that "[Simpson's rule] yields exact results for cubic polynomials even though it is derived from a parabola!"[5] Our chief concern is that the temptation is great to carelessly adopt the "quadratic geometry" of the elemental Simpson's Rule (a geometry which we do appreciate) to the composite rule, rather than to rethink the issue.[4] We now establish that the geometry of

[3]These will be defined shortly. The impatient will want to refer to eq. (8).

[4]We will demonstrate how to "think the issue" below, with the derivation of Hermite's Rule.

Simpson's Rule is not unique, and that we can, without much ado, associate instead a C^1 cubic-spline interpolant to the composite Simpson's Rule.

Since Simpson's error of approximation is proportional to the fourth-derivative, we can add any cubic c to the quadratic q such that $c(x_i) = 0$ at the panel values x_0, x_1, and x_2. This clearly leaves the approximation of Simpson's Rule unchanged,

$$\int_{x_0}^{x_2} (q(x) + c(x))dx = \frac{h}{3}[q(x_0) + c(x_0) + 4(q(x_1) + c(x_1)) + q(x_2) + c(x_2)]$$
$$= \frac{h}{3}[q(x_0) + 4q(x_1) + q(x_2)],$$

and since Simpson's Rule is exact for cubics, the exact integral of this cubic $q(x) + c(x)$ is given by Simpson's Rule. Thus we may turn the tables on the original derivation of Simpson's Rule as the integral of a quadratic and agree that Simpson's Rule will henceforth be associated with any cubic interpolant of the three points on the panel (including the degenerate cubic which is the Lagrange interpolating quadratic). Examine the top two plots in Figure (1): these two functions – one a quadratic C^0 spline, the other a cubic C^1 spline – have the same integral value on the intervals shown!

In the case of a (non-degenerate) cubic, interpolation of the three data values leaves one degree of freedom to be used as we wish (a cubic is determined by four constraints). Why not, therefore, use the extra degree of freedom when constructing the composite rule? For example, we could prescribe the slope at each left endpoint of a panel. This gives us just enough freedom to create a C^1 spline, as follows: the choice of the slope at the left endpoint of the first cubic (along with the three constraints of interpolation) dictates the slope at the right endpoint; we then choose the left slope of the second panel's cubic to match the first's right slope, which in turn dictates the second's right slope; and so on. Continuing in this manner we see that the choice of the first cubic's slope at the left endpoint completely determines all subsequent cubics, resulting in a C^1 cubic-spline interpolant.

These cubic splines on panels are in the same family as quadratic splines with slope continuity on subintervals (which are unpopular due to our inherent inability to control slope behavior at both ends of the spline). For these splines "the tail wags the dog": the slope at one side of the interpolant wags the other. Imagine grasping the shipwright's spline[5], this C^1 interpolant,

[5]A *spline* was originally a flexible rod serving as a design tool for ship builders.

which is pegged to all the data points, at one end and varying the slope: you would have no control over the slope at the other end. It would wiggle as it pleases. Quartics defined on three-subinterval panels, quintics on fours, etc. all can be made C^1, but will “wag” in this manner: from a design standpoint, one is at the mercy (rather than in command) of the curve.

4 Hermite’s Rule

The Hermite of Hermite’s Rule is Charles Hermite (1822-1901), about whom Poincaré (his student) said “Talk with M. Hermite. He never evokes a concrete image, yet you soon perceive that the most abstract entities are to him like living creatures.”[6] In 1873 he proved the transcendence of e (meaning that e cannot be expressed as the root of a polynomial equation with integral coefficients), and having done so wrote to a friend that “I shall risk nothing on an attempt to prove the transcendence of the number π. If others undertake this enterprise, no one will be happier than I at their success, but believe me...it will not fail to cost them some efforts.” We may be excused for imagining that he actually said this in French.

Hermite’s name is well-known in quantum mechanics (Hermite functions, Hermite polynomials), in algebra (Hermitian forms), and in numerical analysis (Hermite interpolation). As so often happens, the physicists must be grateful to a man who was doing number theory - nothing really “practical” at all. Hermite corresponded voluminously with mathematicians all over Europe, and was known for his kindly, encouraging manner. At the time of his death, he was “loved the world over”. One thinks, perhaps, of Erdos in our own time.

We now proceed to the derivation of Hermite’s Rule, so named because of its relationship to Hermite interpolation. Hermite’s Rule and its associated approximations

- work for even or odd numbers of subintervals;
- use derivative information (if available, or approximations if not);
- are associated with a very satisfying geometry; and

[6]This and following quotes are from [12] and [1].

- have error terms comparable to (and, in some cases, superior to) Simpson's Rule.

We will derive Hermite's Rule based on the geometry of the C^1 cubic-spline interpolant.

Consider $n+1$ data values $y_0, y_1, \ldots, y_n$ associated with $n+1$ equally spaced points, $x_0, x_1, \ldots, x_n$. We will suppose that the slopes m_i at the points x_i for $i \in \{0, \ldots, n\}$ are given. Then between any two data points a unique cubic is completely defined by the four constraints of matching the data and slopes for that subinterval.

This is the Hermite cubic interpolator[13], and it provides us with another elemental method for numerical integration. The cubic approximation to the function on the interval $[x_i, x_{i+1}]$ can be expressed as

$$c_i(x) = y_i + m_i(x - x_i) + \frac{1}{h^2}[3(y_{i+1} - y_i) - (m_{i+1} + 2m_i)h](x - x_i)^2$$

$$+\frac{1}{h^3}[-2(y_{i+1} - y_i) + (m_{i+1} + m_i)h](x - x_i)^3,$$

as one can verify by calculation ($c_i(x_i) = y_i$, $c_i(x_{i+1}) = y_{i+1}$, $c_i'(x_i) = m_i$, and $c_i'(x_{i+1}) = m_{i+1}$). This cubic has an especially nice feature as a bonus: its derivative serves as a piecewise quadratic interpolator of the derivative function of f at the locations x_i and x_{i+1}.

We integrate c_i on the interval $[x_i, x_{i+1}]$ to give the elemental integration rule

$$\frac{h}{12}[6(y_i + y_{i+1}) - h(m_{i+1} - m_i)],$$

which, when applied to an interval divided into many subintervals, yields the elegantly simple composite rule

$$H \equiv h\sum_{i=0}^{n} y_i - \frac{h}{2}(y_0 + y_n) - \frac{h^2}{12}(m_n - m_0). \tag{7}$$

Cancellation of the internal derivative terms leads to this beautiful result. We have written the approximation in this form to illuminate its similarity to the trapezoidal rule (eq. (6)):

$$H = T - \frac{h^2}{12}(m_n - m_0).$$

One may well marvel at the simplicity of this formula: it is the trapezoidal rule with a slight slope adjustment (which explains why this rule is more commonly known as the "Corrected trapezoid(al) rule")[7]. Note that the values of the internal slopes, while fundamental to the geometry, have become irrelevant from the point of view of the estimate: they disappear from the formula (see Kahaner, Moler and Nash[8] (p. 162) for more on "the case of the disappearing internal slopes").

Higher-order (and more complicated) methods can be created by considering Hermite interpolators for $N > 2$ points (matching data and slopes over panels, rather than subintervals). The Hermite polynomial is given in general by

$$H(x) = \sum_{i=1}^{N} l_i^2(x)\{[1 - 2l_i'(x_i)(x - x_i)]f_i + (x - x_i)f_i'\},$$

where the l_i are Lagrange polynomials, which can be written as

$$l_i(x) = \prod_{j=1, j\neq i}^{N} \frac{(x - x_j)}{(x_i - x_j)} \tag{8}$$

(see Buchanan and Turner[2]). The error of approximation of these Hermite polynomials is

$$f(x) - H(x) = \frac{f^{(2N)}(\xi)}{(2N)!}[L_N(x)]^2,$$

where

$$L_N(x) = \prod_{j=1}^{N}(x - x_j).$$

and where ξ is an element of the smallest interval containing x and all the x_i[8].

For our purposes we consider only two-point (N=2) Hermite interpolation - that is, Hermite cubics. Hermite's Rule thus has local error on each subinterval of

$$\frac{f^{(4)}(\xi)h^5}{720}.$$

[7]Additional details concerning this formula are given in the appendix.

[8]Greek letters like ξ will usually designate unknown elements of these sorts of intervals (unspecified) in the following.

When we build a composite rule with this elemental rule, we integrate over the subintervals of the interval $[a, b]$ to produce an error of

$$\frac{1}{4}\left[\frac{(b-a)}{180}h^4 f^{(4)}(\xi)\right],$$

where $\xi \in (a, b)$. This fourth-order scheme with degree of precision three has approximately one-fourth the error of Simpson's Rule.

While Hermite's Rule thus beats Simpson's Rule in terms of error, and has a more pleasing and sensible geometry, the obvious criticism of this rule is that it requires derivative information (at least at the boundary points x_0 and x_n). Take that away and what have we got?

In the absence of the actual derivative information, we simply approximate it: we replace the true slopes with approximations using forward and backward difference formulae. If we make our goal a method of fourth-order and degree of precision three (for comparison with Simpson's Rule), then we will not succeed by choosing the most obvious difference methods. The usual "calculus class" approximations, e.g.

$$f'(x) = \frac{f(x+h) - f(x)}{h} + \frac{h}{2}f''(\xi),$$

will achieve neither fourth order, nor precision three. To do both, we must use at least the three-point forward and backward difference formulae

$$f'(x_0) = \frac{1}{2h}[3y_0 - 4y_1 + y_2] + \frac{h^2}{3}f^{(3)}(\xi_0))$$

and

$$f'(x_n) = \frac{1}{2h}[y_{n-2} - 4y_{n-1} + 3y_n] + \frac{h^2}{3}f^{(3)}(\xi_n)).$$

Using these approximation we arrive at the formula

$$H_3 = T - \frac{h}{24}(3y_0 - 4y_1 + y_2 + y_{n-2} - 4y_{n-1} + 3y_n),$$

where the subscript (3) indicates that we have used the three-point difference scheme to approximate H. This is exactly the same weighted sum as that seen earlier (eq. (4)), derived by adapting Simpson's Rule to odd numbers of subintervals.

One other surprising link to Simpson's rule can be made: if one applies the H_3 rule to a single panel, it yields exactly Simpson's Rule:

$$H_3 = \frac{h}{2}(y_0+2y_1+y_2) - \frac{h}{24}(3y_0-4y_1+y_2+y_0-4y_1+3y_2) = \frac{h}{3}(y_0+4y_1+y_2).$$

Once we push beyond a single panel, however, this connection to Simpson's rule is lost. Even so, it suggests that we could associate a quadratic geometry with endpoint subintervals (remember that we first derived H_3 by tacking on integrals of quadratics on the ends).

Rather than think quadratically, however, we continue to associate the Hermite cubic-spline geometry with these Hermite rules created via derivative approximations. Rather than the exact slopes, we use the best approximations available. Because the internal slopes "disappear", we can use any approximation whatsoever to them: an often reasonable estimate is provided by centered finite differences, such that at each of the internal points ($i \in \{1, \ldots, n-1\}$) we estimate the slope at (x_i, y_i) to be the symmetric difference

$$\frac{y_{i+1} - y_{i-1}}{2h}.$$

With these slopes given, along with the interpolation conditions, the internal cubics are completely defined. We are left to determine the cubics associated with the endpoints.

At the endpoints themselves ($i = 0$ and $i = n$) we use the three-point forward and backward difference approximation to the derivative for the H_3 Hermite rule. Putting it all together we get the approximate Hermite cubic-spline interpolator, which is still C^1 and is, from a practical standpoint, an interpolator whose derivative continues to provide a reasonable approximation to the derivative function of f. See the bottom left plot in Figure (1) for this interpolator to our sample data.

The order of error of H_3 can be derived by combining the error of Hermite's Rule with the error of the derivative approximations:

$$I - H_3 = \frac{1}{4}\left[\frac{(b-a)}{180}h^4 f^{(4)}(\xi)\right] - \frac{h^4}{36}(f^{(3)}(\xi_n) - f^{(3)}(\xi_0))$$

$$= \frac{1}{4}\left[\frac{(b-a)}{180}h^4 f^{(4)}(\xi)\right] - \frac{h^4}{36}f^{(4)}(\xi_m)(\xi_n - \xi_0)$$

(which is a consequence of treating the last term by the fundamental theorem of calculus). Then

$$I - H_3 = \frac{1}{4}\left[\frac{(b-a)}{180}h^4 f^{(4)}(\xi)\right] - \frac{h^4}{36}f^{(4)}(\xi_m)\rho(b-a)$$

$$= -\frac{(b-a)}{36}h^4\left[\rho f^{(4)}(\xi_m) - \frac{1}{20}f^{(4)}(\xi)\right],$$

where $\rho \to 1$ as $h \to 0$.

Thus, in the limit, this method does not have as good an error profile as Simpson's Rule (by a factor of about five). A disappointing start, but "a journey of a thousand miles begins with a single step." Let's take another step (well, two steps): instead of H_3, we consider H_5, using the five-point difference schemes applied to the endpoints. That is, we approximate the endpoint derivatives by

$$f'(x_0) = \frac{1}{12h}[-25y_0 + 48y_1 - 36y_2 + 16y_3 - 3y_4] + \frac{h^4}{5}f^{(5)}(\xi_0) \qquad (9)$$

and

$$f'(x_n) = \frac{1}{12h}[3y_{n-4} - 16y_{n-3} + 36y_{n-2} - 48y_{n-1} + 25y_n] + \frac{h^4}{5}f^{(5)}(\xi_n).$$

These approximations lead to the rule

$$H_5 = T - \frac{h}{144}[25(y_0+y_n) - 48(y_1+y_{n-1}) + 36(y_2+y_{n-2}) - 16(y_3+y_{n-3}) + 3(y_4+y_{n-4})].$$

As for the error,

$$I - H_5 = \frac{(b-a)}{720}h^4 f^{(4)}(\xi) - \frac{h^5}{720}(f^{(5)}(\xi_n) - f^{(5)}(\xi_0)),$$

which we can write as

$$= \frac{(b-a)}{720}h^4[f^{(4)}(\xi) - h\rho f^{(6)}(\xi_m)]$$

provided f is six times differentiable, and where $\rho \to 1$ as $h \to 0$. This, too, is a fourth-order method with degree of precision three. As $h \to 0$ the error of this method tends to the error of Hermite's Rule, as we see here and as we shall see again when we test the schemes on some standard functions from calculus while forcing h to get very small.

Once again we stick with the geometry which has served us so well, approximating slopes at internal points by centered differences and at the endpoints by the five-point forward and backward difference approximations (see the bottom right plot of Figure (1)).

5 Test Results

For purposes of comparing the integration rules, it is necessary to revert to functions with known integrals. Burden and Faires[3] provide a set of test functions which we used to compare the methods, but we also added several more of our own, as well as a number chosen from the integral tables in the back of a favorite[9] calculus text[11]. We wanted to give a rather complete set of functions, such as one is likely to encounter in calculus class.

As one can see from the following tables, Simpson's Rule does better than Hermite's Rule H_3 by a factor of 4 or so in general, while Hermite's Rule H_5 beats Simpson's by a similar factor. In the limit as $h \to 0$ Hermite's Rule itself shows that it is about 4 times better than Simpson's Rule, as expected.

We have included a couple of functions which are not differentiable, or not twice differentiable on the domain: results for these functions are harder to classify. It seems fairly clear, though, that for a reasonable number of subintervals (e.g. 80) the results of H_5 are superior to those of Simpson's Rule on almost all functions. And what if the data come from an automated process? Certainly much industrial-strength integral estimation from data involves at least 81 points; thus, provided that the date come from relatively smooth processes, the better rule to use is H_5.

6 Conclusions

Putting the right geometry with an integration rule is just good mathematics. A careless connection (or, worse yet, no connection) of integration techniques

[9] Favorite in spite of the fact that no geometry is associated with the derivation of the composite Simpson's Rule!

to function approximation and interpolation is a wasted opportunity. One might argue that the geometry really doesn't matter (especially if it isn't even unique); that what matters is the numerical answer. But what matters in our minds is that students make the strong connection between subjects which really should go hand-in-hand.

We present Hermite's Rule as a useful replacement for the overworked and at times unwieldy Simpson's Rule. Hermite's Rule avoids the three concerns of Simpson's rule mentioned earlier, namely the problem of even/odd numbers of subintervals, the curious emphasis on every other point, and the C^0 geometry commonly associated with the composite Simpson's Rule. Hermite's Rule has the distinct advantage of emphasizing modern data sampling problems while introducing piecewise differentiable cubic interpolating curves with known slopes at the nodes, which play such a significant role in computer-aided geometric design.

It is clear that the H_5 approximation to Hermite's Rule has a better error than Simpson's rule on most of the standard test functions, yet does not require any explicit derivative information, has a formula as simple to apply as Simpson's, and is associated with a useful geometry whose derivative also serves as an approximation to the derivative of f.

So next time you find yourself using or teaching numerical integration techniques, consider Hermite's Rule, and consider teaching your rules in conjunction with piecewise interpolation of data points. Integration techniques provide one more opportunity to broach this important subject, and to demonstrate to our students the interdependence of important topics in mathematics and modern life. And perhaps the next time they look out an automobile windshield they won't just see telephone poles and lines - they'll also see integration schemes and interpolators.

7 Appendix

Another derivation of Hermite's Rule is provided by consideration of the asymptotic error in the trapezoidal rule as $h \to 0$ (the following is based on [2]):

$$\lim_{n\to\infty} \frac{\epsilon_T(f)}{h^2} = \lim_{n\to\infty} \left[-\frac{h}{12} \sum_{k=0}^{n-1} f''(\eta_k) \right] = -\frac{1}{12} \lim_{n\to\infty} \sum_{k=0}^{n-1} f''(\eta_k) h,$$

Table 1: The first six functions in this table were used by Burden and Faires [3] to compare integration schemes, and integrated from 0 to 2. Additional functions were added to these to offer a more complete sample of typical functions from calculus. We used 4 panels (i.e. 8 subintervals) to obtain these results. It appears that the error in Hermite's Rule H_3 is often around 3 to 4 times the error in Simpson's Rule, while Hermite's Rule H_5 beats Simpson's by a similar factor. Results have not settled down for the majority of these rules (n is still rather small).

Function	True	$\epsilon_{H_3}/\epsilon_S$	$\epsilon_S/\epsilon_{H_5}$	ϵ_S/ϵ_H
x^2	2.6667	NaN	NaN	NaN
x^4	6.4000	3.8125	-4.0000	-4.0000
$1/(x+1)$	1.0986	2.7018	2.7302	-3.6191
$\sqrt{x^2+1}$	2.9579	-16.9978	0.0816	-3.8235
$\sin(x)$	1.4161	4.0152	-2.7090	-4.0240
e^x	6.3891	3.6435	-6.6074	-3.9763
$\ln(x+1)$	1.2958	2.9649	4.4211	-3.7525
$1/(x^2+1)$	1.1071	46.3668	-0.0414	-6.8899
$1/\sqrt{x^2+1}$	1.4436	49.6661	-0.0372	-5.4080
$\cos(2x)$	-0.3784	4.9928	-1.0510	-4.0977
$\cos(5x)$	-0.1088	2.8500	2.2358	-4.7071
$\cos(10x)$	0.0913	-1.9091	-0.7013	-10.4870
$4x^3$	16.0000	NaN	NaN	NaN
$5x^4$	32.0000	3.8125	-4.0000	-4.0000
$6x^5$	64.0000	3.8125	-4.0000	-4.0000
$7x^6$	128.0000	3.5332	-11.5817	-3.9642
$8x^7$	256.0000	3.2474	12.3380	-3.9283
$(5/2)*x^{3/2}$	5.6569	1.6317	1.2257	-0.5492
$signum(x-pi/4)*(x-pi/4)^2$	0.4358	0.4175	-41.8741	0.4041
$\|x-pi/4\|$	1.0461	0.2741	2.0348	3.6477

which is a Riemann sum; thus

$$\lim_{n\to\infty} \epsilon_T(f) = -\frac{h^2}{12}\lim_{n\to\infty}\int_a^b f''(x)dx = -\frac{h^2}{12}[f'(b)-f'(a)],$$

the asymptotic error.

Table 2: The same table, using 40 panels (i.e. 80 subintervals). We now see good convergence of the schemes toward the expected values of about 5, for the ratio of error in H_3 to Simpson's, to 4 for the other two ratios. Again, the non-differentiable functions aren't giving consistent results.

Function	True	$\epsilon_{H_3}/\epsilon_S$	$\epsilon_S/\epsilon_{H_5}$	ϵ_S/ϵ_H
x^2	2.6667	NaN	NaN	NaN
x^4	6.4000	4.6563	-4.0000	-4.0000
$1/(x+1)$	1.0986	4.3964	-4.5356	-3.9952
$\sqrt{x^2+1}$	2.9579	2.2768	-4.7247	-4.0125
$\sin(x)$	1.4161	4.6884	-3.9713	-4.0002
e^x	6.3891	4.6283	-4.0284	-3.9998
$\ln(x+1)$	1.2958	4.4715	-4.3259	-3.9971
$1/(x^2+1)$	1.1071	14.6165	-2.0563	-3.9996
$1/\sqrt{x^2+1}$	1.4436	12.6549	-2.4243	-4.0002
$\cos(2x)$	-0.3784	4.8299	-3.8806	-4.0010
$\cos(5x)$	-0.1088	4.8519	-3.3541	-4.0060
$\cos(10x)$	0.0913	3.1861	-3.3110	-4.0240
$4x^3$	16.0000	NaN	NaN	NaN
$5x^4$	32.0000	4.6563	-4.0000	-4.0000
$6x^5$	64.0000	4.6563	-4.0000	-4.0000
$7x^6$	128.0000	4.6115	-4.0432	-3.9996
$8x^7$	256.0000	4.5668	-4.0874	-3.9993
$(5/2)*x^{3/2}$	5.6569	1.6569	1.2316	-0.5522
$signum(x-pi/4)*(x-pi/4)^2$	0.4358	0.1439	6.9472	0.0655
$\|x-pi/4\|$	1.0461	1.5809	0.6325	0.6325

This process can be continued, leading to the *IMT formula*[7], or *Euler-Maclaurin summation formula*[10]:

$$\int_a^b f(x)dx = T - \sum_{r=1}^{m} \frac{h^{2r}B_{2r}}{(2r)!}[f^{(2r-1)}(b) - f^{(2r-1)}(a)] + R_m,$$

where

$$R_m = \frac{h^{2m+1}}{(2m)!}\int_0^1 B_{2m}(t)\left[\sum_{k=0}^{n-1} f^{(2m)}(a+kh+ht)\right]dt,$$

and where $B_n(t)$ are the Bernoulli polynomials of degree n and the B_n are

Table 3: Using 100 panels.

Function	True	$\epsilon_{H_3}/\epsilon_S$	$\epsilon_S/\epsilon_{H_5}$	ϵ_S/ϵ_H
x^2	2.6667	NaN	NaN	NaN
x^4	6.4000	4.7125	-4.0000	-4.0000
$1/(x+1)$	1.0986	4.6021	-4.0891	-3.9992
$\sqrt{x^2+1}$	2.9579	3.6907	-4.6279	-4.5680
$\sin(x)$	1.4161	4.7257	-3.9953	-4.0000
e^x	6.3891	4.7010	-4.0048	-4.0001
$\ln(x+1)$	1.2958	4.6346	-4.0552	-3.9996
$1/(x^2+1)$	1.1071	8.7061	-3.7718	-4.0003
$1/\sqrt{x^2+1}$	1.4436	7.9190	-3.8348	-3.9987
$\cos(2x)$	-0.3784	4.7834	-3.9808	-4.0002
$\cos(5x)$	-0.1088	4.7996	-3.8821	-4.0010
$\cos(10x)$	0.0913	4.1501	-3.6818	-4.0038
$4x^3$	16.0000	NaN	NaN	NaN
$5x^4$	32.0000	4.7125	-4.0000	-4.0000
$6x^5$	64.0000	4.7125	-4.0000	-4.0000
$7x^6$	128.0000	4.6941	-4.0070	-3.9999
$8x^7$	256.0000	4.6757	-4.0141	-3.9999
$(5/2)*x^{3/2}$	5.6569	1.6574	1.2319	-0.5523
$signum(x-pi/4)*(x-pi/4)^2$	0.4358	0.0738	13.5577	-0.0250
$\lvert x-pi/4\rvert$	1.0461	1.1938	0.8376	0.8376

Bernoulli numbers ($B_0 = 1$, $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, $B_8 = -1/30$, and so on), generated by

$$\frac{t}{e^t - 1} = \sum_{n=0}^{\infty} B_n \frac{t^n}{n!}.$$

According to [10], this is "...an asymptotic expansion whose error when truncated at any point is always less than twice the magnitude of the first neglected term."

References

[1] E. T. Bell. *Men of Mathematics.* Simon and Schuster, New York, 1980.

[2] J. L. Buchanan and P. R. Turner. *Numerical Methods and Analysis.* McGraw Hill, New York, 1992.

[3] R. L. Burden and J. D. Faires. *Numerical Analysis.* PWS Publishing Company, Boston, fifth edition, 1993.

[4] F. Cajori. *A History of Mathematics.* The MacMillan Company, London, second edition, 1926.

[5] S. C. Chapra and R. P. Canale. *Numerical Methods for Engineers.* McGraw-Hill Book Company, New York, second edition, 1988.

[6] S. D. Conte and C. de Boor. *Elementary Numerical Analysis.* McGraw Hill, New York, third edition, 1980.

[7] K. Itô, editor. *Encyclopedic Dictionary of Mathematics.* The MIT Press, Cambridge, MA, second edition, 1993.

[8] D. Kahaner, C. Moler, and S. Nash. *Numerical Methods and Software.* Prentice Hall, Englewood Cliffs, N.J., 1989.

[9] G. Polya. *How to Solve It.* Princeton University Press, second edition, 1957.

[10] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: the Art of Scientific Computing.* Cambridge University Press, New York, 1986.

[11] S. L. Salas and E. I. Hille. *Calculus: One and Several Variables.* Wiley, New York, third edition, 1978.

[12] G. F. Simmons. *Differential Equations with Applications and Historical Notes.* McGraw-Hill, Inc., New York, second edition, 1991.

[13] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, MA, first edition, 1986.