Math 361S Lecture Notes
Differentiation and Richardson Extrapolation

Jeffrey Wong

March 6, 2019

Topics covered

- Differentiation
    - Finite differences
    - Error analysis, effect of rounding error
    - Deriving formulas using Taylor series...
    - ...and using Lagrange interpolation
- Richardson extrapolation
    - Asymptotic error series
    - Use in deriving differentiation formulas
    - Estimating error using two approximations

1 Differentiation

We now consider the problem of deriving formulas for the derivative $f'(x)$ of a function at a point x using its values at a finite set of nearby points. The simplest is the forward difference

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

defined for $h > 0$ (which is obvious from the definition of f'). To determine, the error, we simply use a Taylor series to expand

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(\xi_x).$$

This gives the forward difference formula with error term:

$$f'(x) = \frac{f(x+h) - f(x)}{h} + \frac{h}{2}f''(\xi_x) \tag{1}$$

where ξ_x lies between x and $x+h$. In particular, the error is $O(h)$, which is not very good; this formula is about as inaccurate as a formula can be (while still valid). To get better formulas, we can use more points and/or arrange them more optimally. There are several approaches.

1.1 Basic Formulas: Taylor series method

A formula using $n+1$ points will have the form

$$f'(x) = \sum_{k=0}^{n} c_k f(x+a_k h) + \text{error}$$

for coefficients c_k and a_k (important point: why must the formula be a linear combination of function values?). Typically the derivation goes as follows:

- Decide which points $x+a_k h$ to use (e.g. x and $x+h$). This set of points is the **stencil** for the approximation (see Figure 1).
- Taylor expand the terms $f(x+a_k h)$ around x, keeping as many terms as needed:

  $$f(x+a_k h) = f(x) + c_k a_k h f'(x) + \cdots$$

  Given $n+1$ coefficients, we can expect to control the first $n+1$ terms; the Taylor series should then be taken to $n+1$ terms plus an error term. **Note:** In some cases, one more term should be taken; see below.
- Choose c_k's so that $f'(x)$ remains and as many other terms cancel as possible.
- The result will be that

  $$f'(x) = \sum_{k=0}^{n} c_k f(x+a_k h) + C f^{(p+1)}(\xi) h^p$$

  for some ξ in the interval spanned by the points and some constant C (that would be found in the derivation). The formula has degree p (see subsection 3.1), i.e. it is exact for polynomials up to degree p, and the error is $O(h^p)$ as $h \to 0$.

1.2 The most important formula: centered difference

Consider a 'centered difference formula'

$$D(h) = af(x-h) + bf(x+h). \tag{2}$$

We seek a and b so that

$$f'(x) = af(x-h) + bf(x+h) + \text{error}. \tag{3}$$

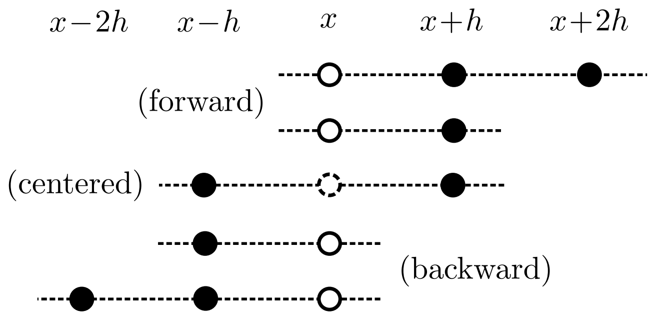

Figure 1: Some stencils for derivative approximation $f'(x)$ at a point x. Note that the centered formula evaluates f' at x but only uses the points at $x \pm h$.

First, expand $f(x-h)$ and $f(x+h)$ in a Taylor series:

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(\xi_2).$$

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(\xi_1),$$

Grouping terms by powers of h, our proposed formula (2) becomes

$$D(h) = (a+b)f(x) + h(-a+b)f'(x) + h^2(a+b)f''(x) + \frac{h^3}{6}(af'''(\xi_1) + bf'''(\xi_2)).$$

To cancel the terms at $O(h^k)$ we need:

$$\begin{aligned} O(1): &\quad a+b=0 \\ O(h): &\quad h(-a+b)=1 \\ O(h^2): &\quad a+b=0. \end{aligned}$$

The first two equations determine a and b:

$$a = -\frac{1}{2h}, \quad b = \frac{1}{2h}.$$

We have no more coefficients to choose, but the $O(h^2)$ term cancels anyway. The resulting formula with error term is

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{12}(f'''(\xi_1) + f'''(\xi_2)).$$

The error can be simplified further into a single term (see subsection 3.2), leading to

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6}f'''(\xi).$$

Key point (symmetry): The $O(h^2)$ term cancels for free due to the symmetry of the approximation. We can guarantee that $D(h)$ has degree of accuracy 1 (exact for constant and linear functions) by choosing a and b.

But observe that the approximation is also exact for

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

It follows that it is exact for all quadratic polynomials (why?). The symmetry provides extra accuracy. Note that this formula uses the same number of points as the forward difference (1) but has much better error - it is a superior formula when it is available.

1.3 Another typical example

We find the three-point **backward difference** using points $x - 2h, x - h$ and x:

$$D(h) = \frac{af(x) + bf(x-h) + cf(x-2h)}{h}.$$

The $1/h$ is put in here knowing that the coefficients will have this factor. There are three coefficients and no symmetry, so we need three terms plus an error term. Expand:

$$\begin{aligned} f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(\xi_1) \\ f(x-2h) &= f(x) - 2hf'(x) + 2h^2 f''(x) - \frac{4h^3}{3}f'''(\xi_2). \end{aligned}$$

Plugging in, we get

$$D(h) = \frac{(a+b+c)}{h}f(x) - (b+2c)f'(x) + h(\frac{b}{2} + 2c)f''(x) - h^2\left(\frac{b}{6}f'''(\xi_1) + \frac{4c}{3}f'''(\xi_2)\right)$$

which gives

$$\begin{aligned} 0 &= a + b + c \\ 1 &= b + 2c \\ 0 &= b/2 + 2c. \end{aligned}$$

Solving for the coefficients, we obtain the formula

$$f'(x) = \frac{-3f(x) + 4f(x-h) - f(x-2h)}{h} + h^2\left(\frac{1}{3}f'''(\xi_1) - \frac{2}{3}f'''(\xi_2)\right).$$

Using (without proof) the rule that the ξ_i's can be taken to be the same (subsection 3.2),

$$f'(x) = \frac{-3f(x) + 4f(x-h) - f(x-2h)}{h} - \frac{h^2}{3}f'''(\xi). \tag{4}$$

Notation (finite difference operators): In this section, we have foregone defining symbols for the various differences (using the placeholder $D(h)$ instead). The forward, backward and centered differences are common enough that they are sometimes given notation. For instance,

$$\delta_h[f](x) = \frac{f(x+h) - f(x-h)}{2h}$$

is common for centered differences, with Δ and ∇ for forward and backward. These symbols are operators (mapping a function to a function) and are analogous to the derivative operator:

$$\delta_h \leftrightarrow \frac{d}{dx}.$$

A 'calculus' of finite differences can be developed, just as you have done in calculus for continuous derivatives.

1.4 Error analysis with rounding error

Observe that both the numerator and denominator of a differentiation formula go to zero as the points get closer together, e.g

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h). \tag{5}$$

If there is an error ϵ in computing f due to rounding (or if f was given empirical data with som eerror), then the error in $f'(x)$ is on the order of ϵ/h. Thus as $h \to 0$ the error goes to ∞. See homework for details. This problem is more or less inherent to computing f', since it is a limit of difference quotients by definition. A key lesson is that it is dangerous to estimate derivatives from data, and calculating derivatives is best avoided when possible.[1]

To give a concrete example, consider the forward difference (5), but assume that the computed function values are $\tilde{f}$ rather than f:

$$\tilde{f}(x) = f(x) + \delta, \qquad |\delta| < \eta$$

where $\eta approx 1.2 \times 10^{-16}$ is the rounding unit. Also assume that

$$|f''(x)| \leq M.$$

Using the error formula we have that the computed derivative, $\tilde{D}(h)$, satisfies

$$\tilde{D}(h) = \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h} = f'(x) + \frac{h}{2}f''(\xi_x) + \frac{\delta_1 - \delta_0}{h}$$

---

[1] Unfortunately, derivatives are sometimes unavoidable, and show up all the time in numerical calculations. In this case, we often accept that there will be some limited accuracy, and use the knowledge of this section to avoid disaster (e.g. taking h too small). When high accuracy is needed and some aspects of the function are known, there is a technique in between exact calculation and finite differences called **automatic differentiation** that can help.

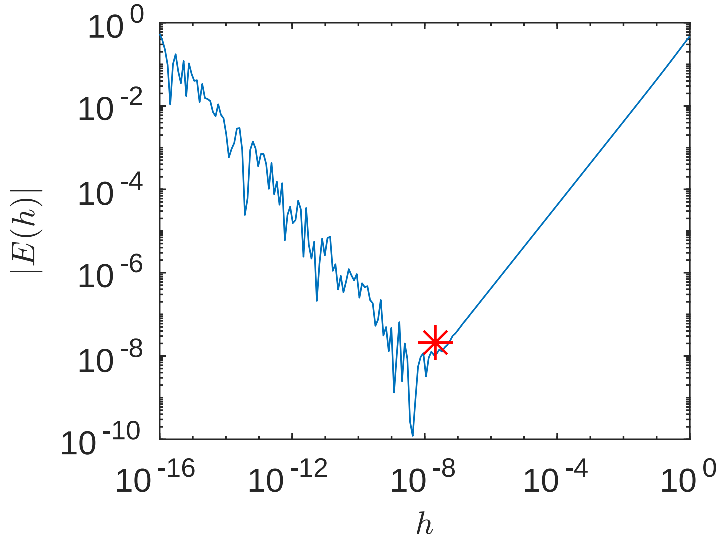

Figure 2: Calculated error $E(h)$ (in Matlab) for a forward difference on a well-behaved function; for h below the size $\sqrt{\eta}$ (up to a constant), rounding error dominates.

where δ_0, δ_1 are the rounding errors in calculating $f(x)$ and $f(x+h)$.

The total error $E(h)$ is therefore has the bound

$$|E(h)| \leq \frac{Mh}{2} + \frac{2\eta}{h}.$$

The minimum occurs at a value $h^* > 0$:

$$0 = E'(h^*) = M/2 - \frac{2\eta}{h^2} \implies h^* = 2\sqrt{\eta/M}.$$

A typical error plot (for $f(x) = \sin x$ and $x = 1$ here) is shown below. The predicted h^* and minimum error are plotted in Figure 2; the red star is the predicted minimum,

$$h^* = 2\sqrt{\eta} \approx 2 \times 10^{-8}, \quad E(h^*) = 2\sqrt{M\eta} \approx 2 \times 10^{-8}$$

When $h > h^*$, the error behaves like the approximation error $E \sim Ch$. When $h < h^*$, the balance of errors switches and the error is dominated by rounding error ($E \sim C/h$).

1.5 Note on differentiating data:

The problem is even worse when trying to estimate derivatives from given data that has noise. The scale of the error δ is then larger than 10^{-16} and causes problems at a larger h. For a k-th derivative (at equally spaced points for simplicity), the formula will be

$$f^{(k)}(x) \approx \frac{\sum c_j f(x + a_k h)}{h^k}.$$

The approximation error will be $O(h)$ or better. From the formula, we see that

$$\delta \text{ sized error in } f \implies \text{error } \sim \frac{\delta}{h^k} \text{ in } f^{(k)}.$$

Not only does this create error, but the nature of the error is 'random' if the error in f is not well-behaved (like random noise), leading to jagged shapes like the one in Figure 2. For a nice example, see the textbook. p425.

1.6 Deriving formulas using interpolation

We can derive the same numerical differentiation formulas using Lagrange interpolation. The process is more mechanical, but cumbersome. Here we switch notation a bit but this is really the same problem as before.

Let $x_0, x_1, \cdots x_n$ be a set of nodes, and let

$$f_k = f(x_k).$$

Recall that the i-th Lagrange basis polynomial is

$$L_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

and the interpolant satisfies

$$f(x) = \sum_{i=0}^{n} f_i L_i(x) + \underbrace{\frac{f^{(n+1)}(\xi_x)}{(n+1)!} w(x)}_{E(x)} \tag{6}$$

where $w(x) = \prod_{j=0}^{n} (x - x_j)$.

Suppose we wish to compute $f'(x)$ at one of the nodes, x_α. The idea is simple: just differentiate the interpolation formula (6). This gives

$$f'(x_\alpha) = \sum_{k=0}^{n} f_i L_i'(x_\alpha) + E'(x_\alpha).$$

The error term can be simplified using the observation that

$$((x - a) f(x))' \Big|_{x=a} = f'(a). \tag{7}$$

which avoids some nasty product rule terms. We have that

$$E(x) = (x - x_\alpha)\frac{f^{(n+1)}(\xi_x)}{(n+1)!}\prod_{k=0,k\neq\alpha}^{n}(x - x_k)$$

so

$$E'(x_\alpha) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}\prod_{k=0,k\neq\alpha}^{n}(x_\alpha - x_k). \tag{8}$$

If the points are all separated by a constant times h then the error is $O(h^n)$. Thus the order of the error is one less than the number of points (also one less than the error for the interpolant itself). Useful rule of thumb: Differentiating worsens the error by one power of h.

Example: To obtain the centered difference formula we derived earlier, first observe that it suffices to derive an approximation for $f'(0)$ given points $-h, 0$ and h. Choose points

$$x_0 = -h, \quad x_1 = 0, \quad x_2 = h$$

and $x_\alpha = 0$. Note that the x_1 must be included here (even though it disappears in the formula) since we must evaluate at an interpolation point. The Lagrange polynomials are

$$L_0 = \frac{x(x-h)}{2h^2}, \quad L_1 = -\frac{(x+h)(x-h)}{h^2}, \quad L_2 = \frac{(x+h)x}{2h^2}.$$

Differentiating and evaluating at zero,

$$L_0'(0) = \left.\frac{x-h}{2h^2}\right|_{x=0} = -\frac{1}{2h}$$

and similarly

$$L_1'(0) = 0, \quad L_2'(0) = \frac{1}{2h},$$

using the general rule (7) to avoid some product rule terms. Thus

$$f'(0) = -\frac{1}{2h}f(-h) + \frac{1}{2h}f(h) + \cdots$$

and the error (8) is

$$E'(0) = \frac{f^{(3)}(\xi)}{6}(0+h)(0-h) = -\frac{f^{(3)}(\xi)}{6}h^2.$$

which yields the same formula as obtained via Taylor series,

$$f'(0) = \frac{f(h) - f(-h)}{2h} - \frac{f^{(3)}(\xi)}{6}h^2.$$

Note that the Lagrange-form error (8) proves some observations from before:

- An approximation with $n+1$ points has an error $O(h^n)$.

- This approximation is exact for polynomials of degree $\leq n$.
- The more centered a formula is, the better the constant in front of the h^n. This is because the product

  $$\prod_{k\neq\alpha}(x_\alpha - x_k)$$

  will the x_k's are distributed symmetrically around the evaluation point x_α (vs. all on one side).

Note that 'number of points here' includes the evaluation points, e.g. the centered difference formula is really a three point formula with one coefficient equal to zero.

1.7 Higher order derivatives

The Taylor series method is well-suited for deriving formulas for higher order derivatives. A centered formula for f'' using $x-h, x$ and $x+h$ can be derived with the form

$$D(h) := af(x-h) + bf(x) + cf(x+h) \approx f''(x).$$

We have that

$$\begin{aligned} f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(\xi_1), \\ f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(\xi_2). \end{aligned}$$

To cancel out the f' term we need $a = c$. This also cancels out the f''' term for free (same symmetry bonus that we saw previously), giving

$$D(h) = (2a+b)f(x) + ah^2 f''(x) + \frac{a}{4!}h^4\left(f^{(4)}(\xi_1) + f^{(4)}(\xi_2)\right).$$

Thus $2a + b = 0$ and $a = 1/h^2$ which gives $a = c = 1/h^2$ and $b = -2/h^2$. Using the IVT trick to simplify the error, the resulting formula is

$$\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = f''(x) + \frac{f^{(4)}(\xi)}{12}h^2$$

This centered difference formula is often used for solving differential equations.

In general, a formula for a k-th derivative requires $k+1$ points. The simplest formula (with $k+1$ points) will have an $O(h)$ or $O(h^2)$ error (with symmetry). Each extra point can improve the order by one.

2 Richardson extrapolation

Here we introduce a powerful, rather general tool for improving the accuracy of a formula with minimum effort. The underlying idea is **of critical importance** in numerical analysis, and will be used often.

The basic setup is this: suppose we have a quantity L to be computed and a formula $A(h)$ defined for $h > 0$ such that

$$\lim_{h\to 0} A(h) = L.$$

We first illustrate via an example: the two-point centered difference formula. We have

$$L = f'(x), \qquad A(h) = \frac{f(x+h) - f(x-h)}{2h}. \tag{9}$$

Consider the Taylor series, but this time with infinite terms:

$$f(x \pm h) = f(x) \pm hf'(x) + \frac{h^2}{2}f''(x) \pm \frac{h^3}{3!}f'''(x) + \cdots .$$

Every other term cancels when computing $f(x+h) - f(x-h)$ (exercise: check this), and so

$$A(h) = L + c_2h^2 + c_4h^4 + \cdots \tag{10}$$

for some constants $c_2, c_4, \cdots$. Their values do not matter here.

The fundamental idea is that by evaluating $A(h)$ at more than one value of h, we obtain systems of equations for the unknowns (like L). These equations can be combined to 'solve' for quantities of interest.

Suppose first that we want to obtain a better approximation to L than $A(h)$. The process of **Richardson extrapolation** is to evaluate (10) at two values of h and solve for L.

Let us pick h and $h/2$. Then we have

$$A(h) = L + c_2h^2 + c_4h^4 + \cdots ,$$

$$A(h/2) = L + \frac{c_2}{4}h^2 + \frac{c_4}{16}h^4 + \cdots$$

To cancel the error term, take 4× the second equation minus the first:

$$4A(h/2) - A(h) = 3L - \frac{3}{4}c_4h^4 + b_6h^6 + \cdots$$

where $\cdots$ is a series in powers of h^2 starting with h^8. Thus

$$\frac{4}{3}A(h/2) - \frac{1}{3}A(h) = L + b_4h^4 + b_6h^6 + \cdots$$

for constants $b_4, b_6, \cdots$. Again, we could compute the b's in terms of the c's but their values are probably not known are not needed here anyway. The result is that

$$A_1(h) = \frac{4}{3}A(h/2) - \frac{1}{3}A(h)$$

is an $O(h^4)$ formula for L. For the centered difference formula (10), this gives

$$\begin{aligned} A_1(h) &= \frac{4}{3}\left(\frac{f(x+h/2)-f(x-h/2)}{h}\right) - \frac{1}{3}\left(\frac{f(x+h)-f(x-h)}{2h}\right) \\ &= \frac{-f(x+h)+8f(x+h/2)-8f(x-h/2)+f(x-h)}{6h} \end{aligned}$$

which is the four point centered difference formula for f' using points $x \pm h$ and $x \pm h/2$.

But we can keep going! The approximation $A_1(h)$ also has an asymptotic error series:

$$A_1(h) = L + b_4h^4 + b_6h^6 + \cdots .$$

Following the same process with $A_1(h)$ and $A_1(h/2)$ we obtain the formula

$$A_2(h) = \frac{16}{15}A_1(h/2) - \frac{1}{15}A_1(h) = \frac{1}{45}A(h) - \frac{20}{45}A(h/2) + \frac{64}{45}A(h/4).$$

The result is an $O(h^6)$ formula using the $O(h^2)$ formula evaluated at $h, h/2$ and $h/4$.

The whole process can be formalized, leading to a table of approximations $A_k(h2^{-j})$ for $j = 0, \cdots k$ that can be computed efficiently using formulas like the above, with

$$A_k(h) = L + O(h^{2k}).$$

It is worth emphasizing that this process is **cheap to compute**, because we only need to compute some formula $A(h)$ a few times to get higher-order accuracy. This makes Richardson extrapolation a convenient way to improve accuracy without much (computational) effort.

Caution: Note that the process tells us the order of each approximation (as $O(h^p)$) but not more precise information like the degree or the constant in front of the h^p. A more detailed analysis would be required to do so. On the other hand, once the formula is known this is not so bad, and in practice is not always necessary.

2.1 Discussion

What is required for Richardson extrapolation? All that is needed is

- An approximation $A(h)$ that can be computed for several values of h
- Prior knowledge that an error series exists of the form

$$A(h) = L + \sum_{k=0}^{\infty} c_k h^{n_k} \tag{11}$$

- The powers of h in the error series (but not the coefficients!)

Nothing else is required. The approximation formula $A(h)$ could even be some ‘black box’ algorithm we cannot change. It turns out that the theory often tells us the error has the form (11) and which powers of h are present.

Series with more powers of h: Consider the forward difference formula

$$D(h) := \frac{f(x+h) - f(x)}{h} \approx f'(x).$$

Suppose we want to derive a formula that uses $x, x+h$ and $x+2h$. We have, using a Taylor series, that

$$D(h) = f'(x) + \frac{h}{2}f''(x) + \frac{h^2}{6}f'''(x) + \cdots = f'(x) + a_1 h + a_2 h^2 + \cdots$$

for the forward difference formula. Thus the error is a series in powers of h. To obtain an $O(h^2)$ approximation using the desired points, we can use $D(2h)$ (which involves x and $x+2h$):

$$D(2h) = f'(x) + 2a_1 h + 4a_2 h^2 + \cdots$$

Combining the two series, we get

$$2D(h) - D(2h) = f'(x) + b_2 h^2 + b_3 h^3 + \cdots$$

so

$$D_2(h) = 2D(h) - D(2h)$$

is an $O(h^2)$ formula for $f'(x)$.

2.2 Variant: error estimation

We can use the same trick to estimate the error in an approximation instead. Consider

$$A(h) = L + E(h)$$

where $E(h)$ is the (unknown) error in the approximation. Suppose $A(h)$ has the series

$$A(h) = L + c_2 h^2 + c_4 h^4 + \cdots .$$

Then we have that

$$A(h/2) = L + c_2 h^2/4 + c_4 h^4/16 + \cdots$$

Rather than try to cancel the error out, we can try to ’solve’ for the error using $A(h)$ and $A(h/2)$ (computable quantities) and cancel out L (unknown). Subtract the two equations:1

$$A(h) - A(h/2) = \frac{3c_2}{4}h^2 + \frac{15c_4}{16}h^4 + \cdots = \frac{3}{4}(c_2 h^2) + O(h^4).$$

But observe that the error is

$$E(h) = c_2h^2 + O(h^4) \approx c_2h^2,$$

so the expression on the right can be written in terms of the error (up to $O(h^4)$):

$$E(h) = \frac{4}{3}(A(h) - A(h/2)) + O(h^4).$$

Alternately, the error in $A(h/2)$ could be found as $\approx (A(h) - A(h/2))/3$.

Key point: Combining two different approximations can yield an estimate of the error if we have some knowledge of the form of the error. This technique is a good way of doing practical error estimation for any method with an asymptotic error series.

3 Background

Some general concepts and details used in this section.

3.1 Degree of accuracy

Suppose we have an approximation

$$f \to A[f]$$

that is **linear**, in that

$$A[c_1 f_1 + c_2 f_2] = c_1 A[f_1] + c_2 A[f_2]$$

for all scalars c_1, c_2 and functions f_1, f_2 that can be approximated.

The **degree of accuracy** n is the largest integer n such that

$$A[f] \text{ is exact for all polynomials of degree } \leq n.$$

For instance, the interpolating polynomial with $n+1$ points has degree of accuracy n, since the interpolant of a degree n polynomial through $n+1$ points is itself.

The forward difference formula has degree of accuracy 1 since

$$f'(x) = \frac{f(x+h) - f(x)}{h} s \text{ when } f \text{ is linear.}$$

In general, if the error in a formula has the form

$$f^{(n+1)}(\xi)\,(\cdots)$$

then it has degree of accuracy n. This is simply because if f is a polynomial of degree n then $f^{(n+1)}(\xi) = 0$, but if it is one degree more then this may not vanish.

Observe that for the differentiation formulas,

$$\text{degree } n \implies \text{error} \approx Ch^n.$$

For instance, the centered formula has one higher degree than the forward one, and its error is $O(h^2)$ instead of $O(h)$. The symmetry of the formula increases its degree by one for free, which leads to the improvement.

3.2 Simplifying error with the IVT

Expressions like

$$c_1 f'''(\xi_1) + c_2 f'''(\xi_2)$$

that arise in calculations like the centered difference formula can be simplified.

This is done using the IVT (intermediate value theorem).

Useful fact: Suppose f is continuous in $[a, b]$, that $x_1, \cdots x_k$ are points in (a, b) and $c_1, \cdots, c_k$ are **positive** numbers such that

$$c_1 + \cdots + c_k = 1.$$

Then there is a point y such that

$$f(y) = c_1 f(x_1) + \cdots + c_k f(x_k).$$

Informally: The average of function values is in the range of f.

Proof. Let $m = \min\{f(a), f(b)\}$ and $M = \max\{f(a), f(b)\}$ (the smaller/larger values of f at the two endpoints) and let

$$z = \sum_{j=1}^{k} c_j f(x_j).$$

We have

$$m = m \sum_{j=1}^{k} c_j \leq z \leq M \sum_{j=1}^{k} c_j = M$$

By the intermediate value theorem, there is a point $y \in (a, b)$ such that $f(y) = z$, i.e.

$$f(y) = \sum_{j=1}^{k} c_j f(x_j).$$

□

Warning: Note that the theorem only applies when the coefficients are positive. However, the Lagrange interpolation approach shows that for differentiation, these terms can be combined anyway, for instance

$$-\frac{1}{3}(-f'''(\xi_1) + 2f'''(\xi_2)) = -\frac{1}{3}f'''(\xi)$$

for the backward difference (4). The theorem does not apply because -2 and 1 do not have the same sign; If deriving with an interpolant, we will get the error on the right.

Thus, the mistake of writing one ξ instead of $\xi_i' s$ turns out to **give the right answer**, even if a rigorous derivation necessitates more complete arguments.

3.3 Mean value theorem (for integrals)

There are two useful variants of the mean value theorem.

Mean value theorem (derivative version): Suppose $f \in C([a,b])$ and f' exists in (a,b). Then there exists a point $\xi \in (a,b)$ such that

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

Equivalently, if f is defined in an interval I and $a, b \in I$ then

$$f(b) - f(a) = f'(\xi)(b - a)$$

for some ξ between a and b.

The equivalent version is just a useful way of interpreting the result: it lets us connect (with an equality) the change in f vs. the change in its arguments. Here's a simple example of how it is typically used:

Claim: Suppose $f \in C^1([a,b])$ and f' is bounded by M:

$$|f'(x)| \leq M \text{ for } x \in [a,b].$$

Then $|f(x_1) - f(x_2)| \leq M|x_1 - x_2|$ for all $x_1, x_2 \in [a,b]$.

Proof. Let $x_1, x_2 \in [a,b]$. By the mean value theorem, there is a $\xi \in (a,b)$ such that

$$f(x_1) - f(x_2) = f(\xi)(x_1 - x_2).$$

Since ξ is in the interval, the bound applies, so

$$|f(x_1 - f(x_2)| \leq |f(\xi)||x_1 - x_2| \leq M|x_1 - x_2|.$$

$\square$

The mean value theorem also has a variant for integrals:

Mean value theorem (integral version): Let $f \in C([a,b])$ and suppose $g(x)$ is a function such that

$$g(x) \geq 0 \text{ for } x \in [a,b].$$

Then there is a point $\xi \in (a,b)$ such that

$$\int_a^b f(x)g(x)\,dx = f(\xi)\int_a^b g(x)\,dx.$$

That is, the factor $f(x)$ can be moved outside the integral if the rest does not change sign (and we replace x with an unknown point).

The integral form will be useful for simplifying the error in integration formulas. It allows us to pull complicated functions outside of an integral, at the cost of evaluating them at some unknown point. Integrating the rest can give nice errors formulas, losing less structure than one would by just bounding terms.

For example, if ξ_x is some point in $(0,h)$ depending on x then

$$\int_0^h f''(\xi_x) x \, dx = f''(\eta) \int_0^h x \, dx = \frac{1}{2} h^2 f''(\eta)$$

for a single value $\eta \in (0,h)$ (this is ξ_{x^*} for some $x^* \in (0,1)$).