Lecture Notes for Math 5630/6630 **Fall 2024**

Note 11: Numerical Differentiation

Tags: math.na *Date: 10/17/2024*

Disclaimer: *This lecture note is for math 5630/6630 class only.*

1 Differentiation with Finite Difference

Let $f \in C^2([a,b])$, we recall the Taylor expansion with reminder term,

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

where $\xi = \xi(x) \in [a,b]$, therefore we can compute the derivative by

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

This approximation offers a way to evaluate the derivative $f'(x)$ with the error term $\mathcal{O}(h)$. In addition, the above formula is *exact* when f is a polynomial of degree 1. We say that an approximation has degree k accuracy if the approximation is *exact* for any polynomial of degree k.

Another important terminology is the *order*. It describes the error term of the approximation. In the above case, the error term scales like $\mathcal{O}(h)$ as $h \to 0$, then the approximation is of order 1 or first order. In general, if the error term behaves like $\mathcal{O}(h^p)$, then we can say that it is the p-th order approximation.

The *stencil* refers to a set of nodes used in the approximation. In the above example, we have used $x, x+h$. We can of course create its sibling

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

which uses the nodes $x-h, x$. When all nodes are $\geq x$ or $\leq x$, we say the scheme is forward or backward, respectively.

1.1 Finite Difference from Taylor Expansion

All results related to finite difference can be easily derived from the Taylor expansion. Suppose we would like to approximate a high-order derivative $f^{(m)}(x)$ with some nodes scattered around x in the following form.

$$\frac{1}{h^m}\sum_{j=0}^{n} c_j f(x+a_j h) = f^{(m)}(x) + E(x,h),$$

where E is the error term and $a_j \in \mathbb{Z}$ (sometimes half integers are used). Since $h \to 0$, we can expand all $f(x+a_j h)$ locally by Taylor series and truncate at least order $(m+1)$.

$$\frac{1}{h^m}\sum_{j=0}^{n} c_j \left(\sum_{p=0}^{m}\frac{1}{p!}f^{(p)}(x)a_j^p h^p + \frac{f^{(p+1)}(\xi_j)}{(p+1)!}a_j^{p+1}h^{p+1}\right).$$

We need all lower (and maybe higher than m-th) order derivatives of f canceled in the above summation, which is (using Kronecker delta),

$$\sum_{j=0}^{n} c_j a_j^p = m!\cdot\delta_{pm},\quad 0\le p\le m.$$

It is straightforward that $n \ge m$ is necessary; otherwise, the first equation system (Vandermonde matrix) must have a zero solution. Suppose that we have found a solution (c_j, a_j), $j = 0,\dots,n$, to the above system, then in the sequel, we try to estimate the error term $E(x,h)$. Especially, when $n = m$, there are two cases. Let the constant $C = \sum_{j=0}^{n} c_j a_j^{m+1}$, then

1. If $C = 0$, then the error term can be estimated by expanding to $(m+2)$-th derivative.

   $$E(x,h) = \frac{1}{h^m}\sum_{j=0}^{n} c_j \frac{f^{(m+2)}(\xi_j)}{(m+2)!}a_j^{m+2}h^{m+2} = h^2\left(\sum_{j=0}^{n} c_j a_j^{m+2}\frac{f^{(m+2)}(\xi_j)}{(m+2)!}\right).$$

   One can expect higher-order accuracy when more terms are involved.

2. If $C \neq 0$, then the error term is

   $$E(x,h) = \frac{1}{h^m}\sum_{j=0}^{n} c_j \frac{f^{(m+1)}(\xi_j)}{(m+1)!}a_j^{m+1}h^{m+1} = h\left(\sum_{j=0}^{n} c_j a_j^{m+1}\frac{f^{(m+1)}(\xi_j)}{(m+1)!}\right).$$

Remark 1.1. *The abscissa ξ_j, $j = 0,\dots,n$ are in general distinct, but it is possible to choose a single ξ to simplify the representation through the intermediate value theorem.*

Lemma 1.2. *If $f \in C^{m+1}(\mathcal{I})$, where $\xi_j \in \mathcal{I}$, then there exists $\xi \in \mathcal{I}$ such that*

$$\sum_{j=0}^{n} c_j a_j^{m+1}\frac{f^{(m+1)}(\xi_j)}{(m+1)!} = \sum_{j=0}^{n} c_j a_j^{m+1}\frac{f^{(m+1)}(\xi)}{(m+1)!}, \tag{0.1}$$

if $c_j a_j^{m+1} \ge 0$ (or ≤ 0) for all $j = 0,\dots,n$.

Proof. Define

$$\psi(x) = \sum_{j=0}^{n} c_j a_j^{m+1}\frac{f^{(m+1)}(x) - f^{(m+1)}(\xi_j)}{(m+1)!},$$

then $\max_j \psi(\xi_j) \ge 0$ and $\min_j \psi(\xi_j) \le 0$. Then apply the intermediate value theorem. $\square$

Example 1.3. *Let $n = 2, m = 2$ for an example. Then the equation system becomes*

$$\begin{aligned} c_0 + c_1 + c_2 &= 0, \\ c_0 a_0 + c_1 a_1 + c_2 a_2 &= 0, \\ c_0 a_0^2 + c_1 a_1^2 + c_2 a_2^2 &= 2. \end{aligned}$$

Then using Gauss elimination, one can find that

$$c_2(a_2 - a_0)(a_2 - a_1) = 2.$$

The above formula can be generalized. The constant $C = c_2(a_2 - a_0)(a_2 - a_1)(a_0 + a_1 + a_2)$. We list a few possible choices.

1. *$(a_0, a_1, a_2) = (-1, 0, 1)$, $c_2 = 1$. Then $c_0 = 1$ and $c_1 = -2$ are derived. In this case $C = 0$. We will have the error term*

$$E(x, h) = \frac{h^2}{24}\left(f^{(4)}(\xi_0) + f^{(4)}(\xi_2)\right) \underbrace{\Longrightarrow}_{??} \frac{h^2}{12} f^{(4)}(\xi).$$

   This is called the central difference scheme.

2. *$(a_0, a_1, a_2) = (0, 1, 2)$, $c_2 = 1$. Then $c_1 = -2$, $c_0 = 1$, $C \neq 0$. The error is*

$$E(x, h) = \frac{h}{6}(f^{(4)}(\xi_1) + 8f^{(4)}(\xi_2)) \underbrace{\Longrightarrow}_{??} \frac{3h}{2} f^{(4)}(\xi).$$

   This is the forward difference scheme.

The combination of the coefficients is not unique. The central scheme has better approximation due to its symmetry. Any combination satisfying $a_0 + a_1 + a_2 = 0$ should have the same order of error.

The general scheme with $a_j = j$ (or $-j$) can be derived from the following theorem.

Theorem 1.4. *In general, if $n = m$, then the forward difference scheme satisfies*

$$\begin{aligned} \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{i} j^p &= 0, \qquad 0 \le p < n \\ \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{i} j^n &= n!\,. \end{aligned}$$

Proof. It is the easiest to prove by a binomial transform. It can also be proved through induction easily. Let

$$P_0(x) = (x-1)^n, \quad P_k(x) = x P'_{k-1}(x),$$

then one can show inductively that P_k, $k \geq 1$, has following form

$$P_k(x) = n(n-1)\cdots(n-k+1)(x-1)^{n-k}x^k + (x-1)^{n-k+1}F(x) \tag{$**$}$$

with $F(x)$ as a polynomial of the highest degree $k-1$. This can be easily proved since

$$\begin{aligned} P_{k+1}(x) = xP_k'(x) = n(n-1)\cdots(n-k)(x-1)^{n-k-1}x^{k+1} +& \\ + (x-1)^{n-k}n(n-1)\cdots(n-k+1)kx^k& \\ + (x-1)^{n-k}(n-k+1)F(x)& \\ + (x-1)^{n-k}(x-1)F'(x).& \end{aligned}$$

The last three terms can merge into the form $(**)$. Therefore, $P_k(1) = 0$ unless $k = n$ and $P_n(1) = n!$ are immediately obtained.

Now if we expand the polynomial P_0 as a monomial,

$$P_0(x) = \sum_{j=0}^{n} \binom{n}{j} x^j (-1)^{n-j},$$

then it is not difficult to show that

$$P_k(x) = \sum_{j=0}^{n} \binom{n}{j} j^k x^j (-1)^{n-j}$$

through induction as well, which is exactly our conclusion by setting $x = 1$. $\square$

Theorem 1.5 (forward difference)**.**

$$f^{(n)}(x) = \frac{1}{h^n}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j} f(x+jh) + \mathcal{O}(h).$$

The corresponding schemes of backward and central differences can be derived similarly.

1.2 Finite Difference from Polynomial Interpolation

Another idea to derive the finite difference scheme for derivatives is using the polynomial interpolation.

Let p_n be the interpolation polynomial for data points $(x_0, f(x_0))$, $(x_1, f(x_1)$, $\cdots$, $(x_n, f(x_n))$. If $p_n(x)$ approximates $f(x)$ well, then $p_n'(x)$ can be a potential candiate approximation for $f'(x)$.

Using the Lagrange interpolation,

$$p_n(x) = \sum_{j=0}^{n} f(x_j)L_j(x) \Rightarrow f'(x) \approx p_n'(x) = \sum_{j=0}^{n} f(x_j)L_j'(x).$$

In particular, if x coincides with a node x_k, the coefficients are

$$\begin{aligned}
L_j'(x_k) &= L_j(x_k) \sum_{l=0, l\neq j}^{n} \frac{1}{x_k - x_l} \\
&= \prod_{l=0, l\neq j}^{n} \frac{x_k - x_l}{x_j - x_l} \sum_{l=0, l\neq j}^{n} \frac{1}{x_k - x_l} \\
&= \frac{1}{x_j - x_k} \prod_{l=0, l\neq j, l\neq k}^{n} \frac{x_k - x_l}{x_j - x_l}, \qquad j \neq k.
\end{aligned}$$

and

$$L_k'(x_k) = L_k(x_k) \sum_{l=0, l\neq k} \frac{1}{x_k - x_l} = \sum_{l=0, l\neq k} \frac{1}{x_k - x_l}.$$

Example 1.6. *For instance, if the nodes are equally spaced, $x_k = x_0 + kh$, then*

$$L_j'(x_k) = \frac{1}{h} \left(\frac{1}{j-k} \prod_{l=0, l\neq j, l\neq k}^{n} \frac{k-l}{j-l} \right), \qquad j \neq k$$

and

$$L_k'(x_k) = \frac{1}{h} \sum_{l=0, l\neq k} \frac{1}{k-l}$$

The derivative $f'(x_k)$ can be approximated by the finite difference

$$f'(x_k) \approx \sum_{j=0}^{n} f(x_j) L_j'(x_k).$$

The error estimate for $|f'(x_k) - p_n'(x_k)|$ can be derived using Newton's form.

$$f(x) - p_n(x) = f[x_0, x_1, \cdots, x_n, x](x - x_0) \cdots (x - x_n).$$

Then differentiate both sides,

$$f'(x) - p_n'(x) = \left(\frac{d}{dx} f[x_0, x_1, \cdots, x_n, x] \right) (x - x_0) \cdots (x - x_n) + (f(x) - p_n(x)) \sum_{j=0}^{n} \frac{1}{x - x_j}.$$

At $x = x_k$, only one term is nonzero,

$$f'(x_k) - p_n'(x_k) = f[x_0, x_1, \cdots, x_n, x] \prod_{l=0, l\neq k}^{n} (x_k - x_l).$$

Therefore,

$$|f'(x_k) - p_n'(x_k)| \leq \frac{\max_\zeta |f^{(n+1)}(\zeta)|}{(n+1)!} \prod_{l=0, l\neq k}^{n} |x_k - x_l|.$$

Remark 1.7. *For an arbitrary x, the error can be estimated by*

$$|f'(x) - p_n'(x)| \leq \prod_{j=0}^{n} |x - x_j| \left(\frac{\max_\zeta |f^{(n+2)}(\zeta)|}{(n+2)!} + \frac{\max_\zeta |f^{(n+1)}(\zeta)|}{(n+1)!} \sum_{j=0}^{n} \frac{1}{x - x_j} \right).$$

1.3 Rounding Error Issue

The finite difference formula provides a simple and effective way to evaluate the derivatives, however, its formulation would be sensitive to rounding errors. Take the central difference scheme for $f''(x)$ as an example, one can derive a similar estimate for higher-order derivatives.

Example 1.8.

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

The error comes from two sources. The truncation error term from $\frac{h^2}{12} f^{(4)}(\xi)$ and the rounding error from the evaluation of the first term by the basic arithmetic operations. Suppose the addition/subtraction is implemented by Kahan sum which almost does not introduce errors in the arithmetic operations. Then the rounding error of $f(x+h) - 2f(x) + f(x-h)$ is at most $4 \max_{x \in \mathcal{I}} |f(x)| \eta$. Therefore the total error

$$|E_{total}| \leq \frac{4 \max_{x \in \mathcal{I}} |f(x)| \eta}{h^2} + \frac{h^2}{12} \max_{x \in \mathcal{I}} |f^{(4)}(x)| \qquad (***)$$

Remark 1.9. *Minimizing the right-hand-side of $(***)$, let $M = \max_{x \in \mathcal{I}} |f(x)| \max_{x \in \mathcal{I}} |f^{(4)}(x)|$, we obtain*

$$\min_{h \in \mathbb{R}} |E_{total}| \leq \sqrt{\frac{4}{3} \eta M}.$$

The optimal achieves at $h^* = \sqrt[4]{48 \eta M}$. For example, if $f(x) = \exp(x)$ and evaluate its second derivative around $x = 0$, then $M \sim 1$, the error is around 1.3×10^{-8} for $h^* \sim 2.5 \times 10^{-4}$.

For higher-order derivatives, the rounding error would have an even greater impact on the finite difference schemes. Then it is much more important to avoid h being too small.

1.4 Improving by Extrapolation

We can combine the previously discussed extrapolation technique to acquire a higher-ordered scheme. We use a very simple example to show how this works.

Example 1.10. *Suppose $A(f, h)$ is the central difference scheme for $f'(x)$, which is*

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

the previous discussion has claimed that $A(f,h) = f'(x) + \mathcal{O}(h^2)$. Now we try to fit the formulation in the framework of extrapolation. Formally, we can expand $f(x \pm h)$ with Taylor series with infinite terms (might not converge though), that is,

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

Therefore $A(f,h) = f'(x) + \frac{h^2}{6} f''(x) + \frac{h^4}{120} f''''(x) + \cdots$. Here the coefficients are all formal since the convergence is not guaranteed. In the next, we take $A(f, \frac{h}{2})$, which uses a smaller step length to approximate $f'(x)$, then

$$
A(f, h/2) = f'(x) + \frac{h^2}{24} f''(x) + \frac{h^4}{1920} f''''(x) + \cdots
$$

Cancel the $\mathcal{O}(h^2)$ term by

$$
\frac{1}{3} \left(4A(f, h/2) - A(f,h) \right) = f'(x) - \frac{h^4}{480} f''''(x) + \cdots .
$$

In this way we have built a more accurate formula $A_1(f,h) = \frac{4A(f,h/2) - A(f,h)}{3}$ for $f'(x)$, the error is fourth order. Bring the definition of the finite difference scheme into A_1, then

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

This central difference scheme has $\mathcal{O}(h^4)$ error.

The above example can still iterate through the extrapolation process, since $A_1(f,h) = f'(x) + \mathcal{O}(h^4)$, we can use $A_2(f,h) = \frac{16}{15} A_1(f, h/2) - \frac{1}{15} A_1(f,h)$ to cancel out the $\mathcal{O}(h^4)$ term which leads to a $\mathcal{O}(h^6)$ error. However one should also notice this process requires more nodes for computation: A_1 needs nodes $x \pm h$, $x \pm h/2$, A_2 will acquire additional nodes $x \pm h/4$ for evaluation. Such a higher precision evaluation method takes more computational time, sometimes we need to trade off the efficiency and accuracy.

Remark 1.11. *One of the advantages of using extrapolation is that h does not have to be too small which is sensitive to numerical rounding errors. The potential issue would be the growth of derivative in order, for sufficiently smooth functions, extrapolation usually produces quite accurate evaluations. The potential limitation of the Richardson extrapolation is the requirement of known asymptotic expansion (formally only), while the Wynn ε method does not have such a limitation.*