ECE 4960

Spring 2017

Lecture 5

Local Analysis: Differentiation

Edwin C. Kan
School of Electrical and Computer Engineering
Cornell University

Approximation in Local Analysis

- It is often difficult to observe global behavior (weather, experiment, commerce, etc.) because our observation and measurement often have a scope and precision in space and time.
- Critical to know between the known points (interpolation or integration to obtain the mean value) or beyond the known points (extrapolation or differentiation to obtain the slope or trends).
- What are the errors in the interpolation and extrapolation approximation? What can we do about it?

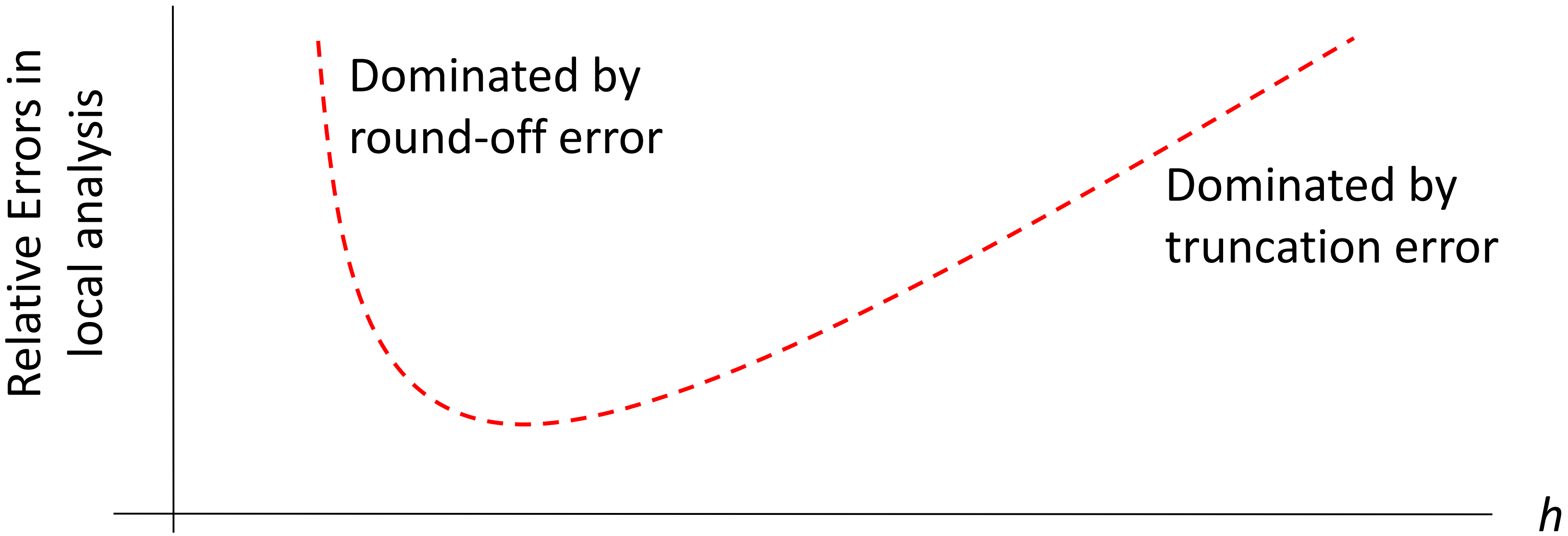
Taylor Series for Local Analysis

- The approximation of a function A is $\hat{A}$.
- Within a resolution limit or step size h, the approximation is **consistent** if $\hat{A} \to A$ as $h \to 0$.
- For the first derivative of a function (slope or margin) where $A = f'(x)$, we can use:

1st –order forward difference: $$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h)$$

2nd–order central difference: $$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

Interplay Between Truncation and Round-off



Hacker Practice

- For $f(x) = x^2$, we know the exact $f'(x=1) = 2$.
- Estimate $f'(x=1)$ by:

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

- Varying the value of h from 0.1 to 10^{-18} to observe the relative error in calculating $f'(x)$.

- Repeat above with $f(x) = x^2 + 10^8$.

- Repeat the above by using

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

Generalized Taylor Approximation

- Assume that in addition to $f(x)$, we have two additional sampling points at $f(x + h_1)$ and $f(x+h_2)$.
- We call x the base point. We know nothing about $f(x)$ except a few sampling point around x, which is thus called the "**local analysis**".
- Taylor expansion to the second order shows:

$$f(x+h_1) = f(x) + h_1 \cdot f'(x) + \frac{1}{2} h_1^2 f''(x) + O(h^3)$$

$$f(x+h_2) = f(x) + h_2 \cdot f'(x) + \frac{1}{2} h_2^2 f''(x) + O(h^3)$$

- $O(h^3)$ above means all terms with h^3 or higher polynomials are truncated.

Second-Order Analysis by Three Points

$$\left\{ f\left(x+h_1\right)= f\left(x\right)+h_1\cdot f'\left(x\right)+\frac{1}{2}h_1^2 f''\left(x\right)+O\left(h^3\right)\right\} \qquad \times h_2^2$$

$$+ \qquad \left\{ f\left(x+h_2\right)= f\left(x\right)+h_2\cdot f'\left(x\right)+\frac{1}{2}h_2^2 f''\left(x\right)+O\left(h^3\right)\right\} \qquad \times\left(-h_1^2\right)$$

$$f'\left(x\right)=\frac{h_1}{h_2\left(h_1-h_2\right)}f\left(x+h_2\right)-\frac{h_1+h_2}{h_1h_2}f\left(x\right)-\frac{h_2}{h_1\left(h_1-h_2\right)}f\left(x+h_1\right)+O\left(h^2\right)$$

- Only possible two-point evaluation: $h_1 = -h_2$
- In general, second-order approximation for $f'(x)$ by three arbitrary points. Third-order approximation for $f'(x)$ by four points, etc.

General Observation from Taylor Series

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

$$= \sum_{n=1}^{\infty} \frac{h^n}{n!} f^{(n)}(x) \cong \sum_{n=1}^{p} \frac{h^n}{n!} f^{(n)}(x) + O\left(h^{p+1}\right)$$

- We can use knowledge of more points (h) to improve the approximation order (p).
- When $h \to 0$, the high-order error terms USUALLY diminish much faster, but not always. Ex.: Odd functions.
- High-order terms can cause local oscillations in larger h.
- There are approximations that are not converging or **consistent** by Taylor expansion.

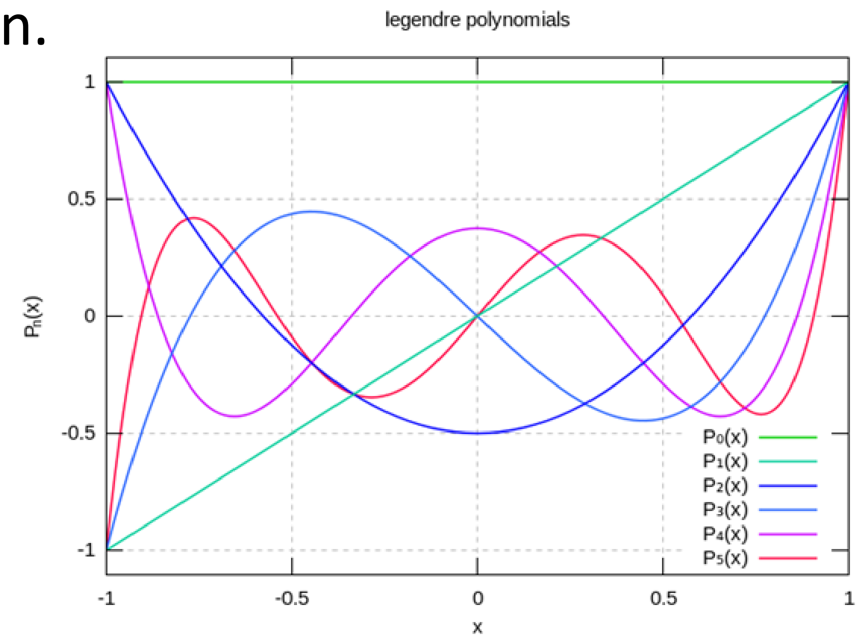
$$\lim_{x \to 0} \frac{e^{-a/x}}{x^n} \to 0; \qquad \lim_{x \to 0} \frac{\exp\left(-\frac{a^2}{x^2}\right)}{x^n} \to 0$$

Other than Taylor Series

- Taylor series are more intuitive, but the base functions of 1, x, x^2, etc. are not orthogonal.
- For polynomials within (–1, 1), we can use orthogonal polynomials such as the Legendre series to improve efficiency in determining the expansion coefficients.
- Additional knowledge can help determine the most appropriate expansion series: coupled equation (how x_1 can affect x_2 in multi-variable case); exponential functions by Hermite series; discontinuity by discrete Galerkin.

$$L_n(x) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}^2 (x-1)^{n-k}(x+1)^k$$

$$L_0 = 1; \quad L_1 = x; \quad L_2 = \frac{1}{2}\left(3x^2 - 1\right); \quad L_3 = \frac{1}{2}\left(5x^3 - 3x\right)\ldots$$



Forward and Backward Euler

- When the local approximation is with respect to time, stability is governed by how we evaluate $f'(t)$.
- Consider the exponential function: $f(t) = C\cdot \exp(at)$, where C is given by the initial values of f at $t = 0$.
- $a < 0$: exponential decay!

$$f'(t) = \frac{df(t)}{dt} = af(t)$$

Forward Euler: $$\frac{f(t) - f(t-\Delta t)}{\Delta t} = af(t-\Delta t) \Rightarrow f(t) = (1 + a\Delta t) f(t-\Delta t)$$

Stable only if: $\Delta t < -\frac{1}{a}$

Backward Euler: $$\frac{f(t) - f(t-\Delta t)}{\Delta t} = af(t) \Rightarrow f(t) = \frac{1}{1 - a\Delta t} f(t-\Delta t)$$

Always stable: $0 < \frac{1}{1 - a\Delta t} < 1$

Hacker Practice

- For $f(t) = \exp(-t)$, i.e., $a = -1$
- Compare the evaluation of $f(t)$ for $0 \le t \le 20$ by three methods:

1. Ground truth: $f(t) = \exp(-t)$
2. Forward Euler with $f(0) = 1$ and march with $\Delta t = 0.5$, $\Delta t = 1.0$ and $\Delta t = 2.0$.

$$f(t) = (1 - \Delta t) f(t - \Delta t)$$

3. Backward Euler with $f(0) = 1$ and march with $\Delta t = 0.5$, $\Delta t = 1.0$ and $\Delta t = 2.0$.

$$f(t) = \frac{1}{1 + \Delta t} f(t - \Delta t)$$

Observe the error in Backward Euler in relation with Δt even with **absolute stability**.

Richardson Extrapolation

- The choice of $h_2 = 2h_1 = 2h$ deserves a closer look:

$$f'(x) = \frac{f(x+h) - f(x)}{h} + E(h); \qquad E(h) = O(h) = \frac{1}{2} h f''(x) + O(h^2) \qquad (1)$$

$$f'(x) = \frac{f(x+2h) - f(x)}{2h} + E(2h); \qquad E(2h) = O(h) = \frac{1}{2} 2h f''(x) + O(h^2) \qquad (2)$$

- By $f(x)$, $f(x+h)$ *and* $f(x+2h)$, we can make a second-order approximation to $f'(x)$:

$$f'(x) = \frac{-1}{2h} f(x+2h) - \frac{3}{2h} f(x) + \frac{2}{h} f(x+h) + O(h^2) \qquad (3)$$

- (3) can be generalized to higher precision by a nested procedure

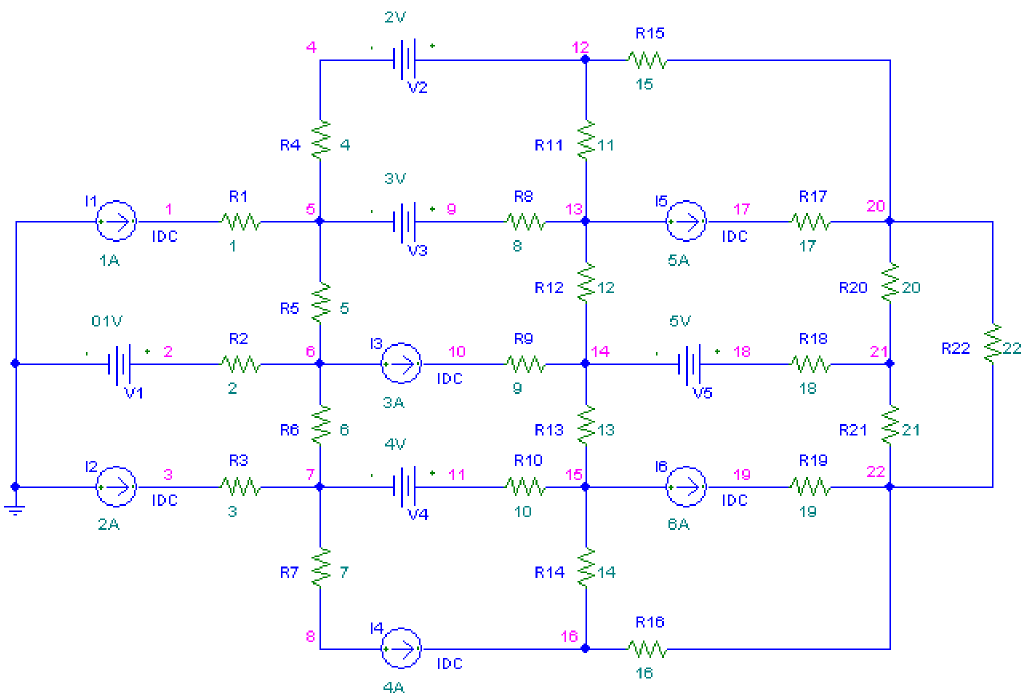
- Comparison of (1) and (2): h adaptivity
- Comparison of (1) and (3): p adaptivity

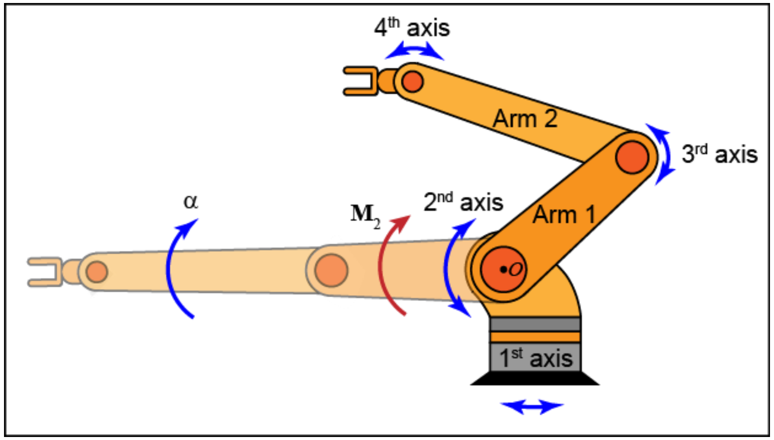
hp Adaptivity

- *h* adaptivity: Improvement in approximation by using small *h* (before precision error dominates)
- *p* adaptivity: Improvement in approximation by using higher order functions with errors $\propto$ O(h^p)
- For simple functions like $f(x) = x^2$, we will have O(h) improvement with smaller h (before precision error dominates), but EXACT solution when second-order approximation is used: an example where *p* adpativity is much better than *h* adaptivity.
- Most often the analytical forms of $f(x)$ and $f'(x)$ are unknown, although we can evaluate $f(x)$ with given x (when x is a multi-variate vector, $f(x)$ evaluation can be computationally expensive)

Realistic Examples for *f(x)*

- $f(V_1, V_2, \ldots)$ can be the transient current to the load, where V_i is the nodal voltage of circuit node i.
- $f(\varphi_1, \theta_1, \varphi_2, \theta_2, \ldots)$ is the distance (or vector) from the robotic palm tip to the object to be fetched, where (φ_i, θ_i) is the solid angle of the i-th robotic joints in a pseudo-rigid-body robotic arm.





How can we verify if adaptivity by either h or p is good enough to represent the realistic physical world?

Caveat: The ground truth may be unknown, and sampling may be expensive or limited!

Richardson Extrapolation Coefficient

- When the ground truth is known or can be estimated by another method (say at the asymptotic trend), we can estimate the error $E(h)$ and $E(2h)$ of each approximation by $x + h$ and by $x + 2h$.
- The Richardson extrapolation coefficient η is defined as:

$$R(h) \equiv \frac{E(2h)}{E(h)} \cong \eta$$

- η will be close to 2 for first-order approximation, and 2^p for p-th order approximation.

Richardson Coefficient Without Known Truth

- If the ground truth is unknown, we can alternatively estimate:

$$R(h) \cong \frac{\hat{A}(4h) - \hat{A}(2h)}{\hat{A}(2h) - \hat{A}(h)} \cong \eta$$

where $\hat{A}(h)$ represents the local approximation function with sampling using *h*.

Ex.: $\hat{A}(h) = f'(x)\big|_h = \frac{f(x+h) - f(x)}{h}$

- Still the same: η will be close to 2 for first-order approximation, and 2^p for p-th order approximation.

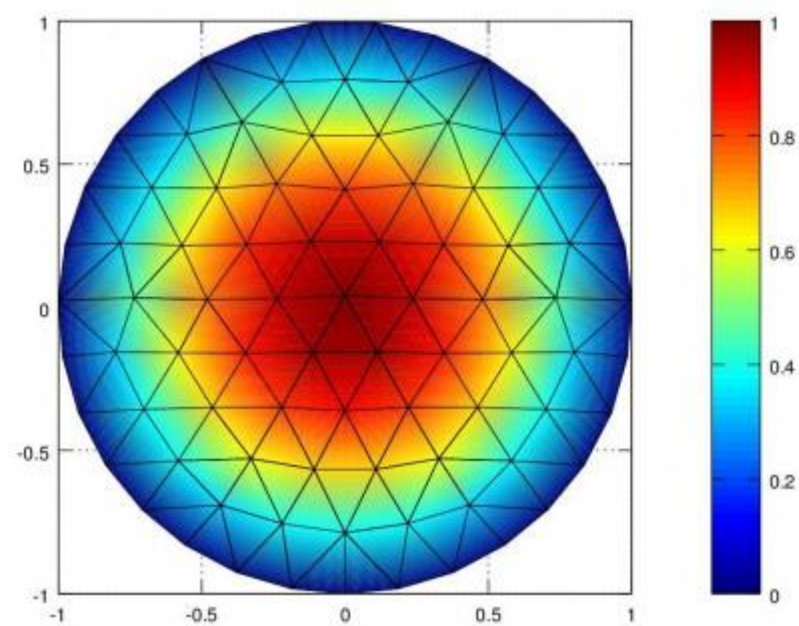
Extension of Richardson Extrapolation

- We have seen the *h* adaptivity in Richardson extrapoltion. We can apply the similar principle to compare the *p* adaptivity, which we will do in the later treatment of ordinary differential equation (ODE) when we can give more realistic examples.

- Can we use the *p* error estimate to guide the choice of *h*?

Importance in Choice of *h*

- In a 3D geometry where we hope to know the temperature equation by solving the Laplace equation
- If we change *h* to *h/2* in all directions
- No. of grid points: 8X
- No. of variables: 8X
- Computational cost: 64X to 512X



Hacker Practice

- For $f(x) = x^3$, we know the exact $f'(x=1) = 3$.
- Estimate $f'(x=1)$, varying the value of h from 2^{-4} to 2^{-10}

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

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

$$f'(x) = \frac{-1}{2h} f(x+2h) - \frac{3}{2h} f(x) + \frac{2}{h} f(x+h)$$

- Tabulate the relative error in calculating $f'(x)$.
- Estimate η for each choice of h by:

$$\eta = \frac{E(2h)}{E(h)} \qquad \eta = \frac{\hat{A}(4h) - \hat{A}(2h)}{\hat{A}(2h) - \hat{A}(h)}$$

After Thoughts

- If you get a correct estimate of η by:

$$\eta = \frac{\hat{A}(4h) - \hat{A}(2h)}{\hat{A}(2h) - \hat{A}(h)}$$

- Is it likely that you have a coding error such as:

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

- Is it likely that you have a wrong implementation of the local analysis (when x is a multi-variable vector)?
- Did you check whether the implementation of $f(x)$ is correct?

- How then should you plan your modular programming and regression test???