Trunc. errors (math)
Round-off errors ($fl(x)$)

An example error analysis

- $u(x) = e^{2x}$
- We will use $\frac{u_{i+1} - 2u_i + u_i}{h^2}$ to approximate the true second der. u_i'' at a point x_i
- <u>Question</u>: How does the relative error depend on our choice of stepsize h?
- We know the exact answer: $u_i'' = 4e^{x_i}$
- Absolute error:

$$\Delta(h) = |\text{approx} - \text{true}| = \left| \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} - u_i'' \right|$$

- Relative error:

$$\mathcal{E}(h) = \left| \frac{\text{approx} - \text{true}}{\text{true}} \right| = \left| \frac{\Delta(h)}{u_i''} \right|$$

- Let's <u>model</u> the absolute error as a sum of two contributions: truncation error and round-off error

$$\Delta(h) = \Delta_{trunc}(h) + \Delta_{RO}(h)$$

- First look at $\Delta_{trunc}(h)$

$$\Delta_{trunc}(h) = \left| \left(\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} \right) - \underbrace{\left(\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + \mathcal{O}(u_i^{(4)} h^2) \right)}_{\text{exact } u_i''} \right|$$

first lectures

$$\boxed{\Delta_{trunc}(h) = \left| \mathcal{O}(u_i^{(4)} h^2) \right|}$$

- Now look at $\Delta_{RO}(h)$:

  - what we want to compute : $\dfrac{u_{i+1} - 2u_i + u_{i-1}}{h^2}$, or $\dfrac{(u_{i+1} - u_i) - (u_i - u_{i-1})}{h \times h}$

  - what we actually compute is something like this ...

$$fl\left[\frac{fl\left[fl\left(fl(u_{i+1}) - fl(u_i)\right) - fl\left(fl(u_i) - fl(u_{i-1})\right)\right]}{fl\left(fl(h) \times fl(h)\right)}\right]$$

  - Consider limit of small h and focus on the subtractions of near identical numbers

$$fl(u_{i+1}) - fl(u_i) \qquad \left(\text{and similar for } fl(u_i) - fl(u_{i-1})\right)$$

  - Recall : $x(1-\delta_m) < fl(x) < x(1+\delta_m)$

  - We can estimate an upper bound

$$\begin{aligned} fl(u_{i+1}) - fl(u_i) &\leq u_{i+1}(1+\delta_m) - u_i(1-\delta_m) \\ &= (u_{i+1} - u_i) + (u_{i+1} + u_i)\delta_m \end{aligned}$$

  - In the limit $h \to 0$, i.e. when $u_{i+1} \to u_i$

$$fl(u_{i+1}) - fl(u_i) = \mathcal{O}(u_i \delta_m)$$

↳ does not go to zero when $h \to 0$

(similar contr. from $fl(u_i) - fl(u_{i-1})$)

  - Estimate round-off error in $\dfrac{u_{i+1} - 2u_i + u_{i-1}}{h^2}$ to be

$$\boxed{\Delta_{RO}(h) = \mathcal{O}\left(\frac{u_i \delta_m}{h^2}\right)}$$

- Putting it together :

$$\Delta(h) = \left| \Delta_{trunc}(h) + \Delta_{RO}(h) \right|$$

$$= \left| \mathcal{O}(u_i^{(4)} h^2) + \mathcal{O}\left(\frac{u_i \delta_m}{h^2}\right) \right|$$

- Relative error :

$$\mathcal{E}(h) = \left| \frac{\Delta(h)}{u_i''} \right| = \left| \mathcal{O}\left(\frac{u_i^{(4)}}{u_i''} h^2 \right) + \mathcal{O}\left(\frac{u_i \delta_m}{u_i''} \frac{1}{h^2} \right) \right|$$

↳ grows when $h \to 0$

↳ grows when $h \to \infty$

[In our particular case, with $u(x) = e^{2x}$, we also have that $\mathcal{O}(u_i) \approx \mathcal{O}(u_i'') \approx \mathcal{O}(u_i^{(4)})$

- Log_{10} of relative error :

$$\log_{10}(\mathcal{E}(h)) = \log_{10} \left| C_1 h^2 + C_2 h^{-2} \right|$$

- Look at the limits $h \to 0$ and $h \to \infty$:

$$\log_{10}(\mathcal{E}(h)) \approx \begin{cases} -2\log_{10}(h) + \log_{10}(C_2) & \text{for } h \to 0 \quad (\log_{10}(h) \to -\infty) \\ 2\log_{10}(h) + \log_{10}(C_1) & \text{for } h \to \infty \quad (\log_{10}(h) \to \infty) \end{cases}$$