00.02 basics : error

01 quality when theres quantity

the rational number $\frac{1}{3}$ exists but it does not exist in the set known as FPS. instead, it is approximated by the nearest FPN. ie, hello errors.

if mathematical operations happen to this number, then hello more errors.

02 notation, floating-point operators

$\circ$, "round to nearest FPN". eg, $x, y \in \mathbb{R} \mapsto \circ(x + y) = \circ(x) + \circ(y)$.

also, $\oplus \ \ominus \ \otimes \oslash$ such that $x \oplus y = \circ(x + y)$. usw.

03 error

suppose $x \in \mathbb{R}$ and $\hat{x} \in \mathbb{FP}$ is its FPN approximation. ie, $\hat{x} = \circ x$. then

- **absolute error**, $\Delta x = |x - \hat{x}|$ and
- **relative error**, $\delta x = \frac{\Delta x}{|x|}$.

rounding is an algorithm and has two kinds of error:

1. **forward**, wrt how well the algorithm approximates the true output; and
2. **backward**, wrt how desired results relate back to expected input.

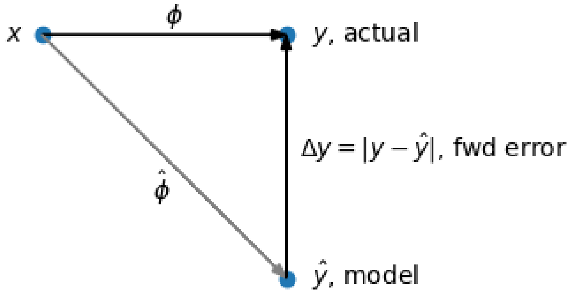
i) forward error

suppose x, y such that $\phi(x) = y$ and $\hat{\phi}(x) = \hat{y}$, where $\hat{\phi}$ is the numerical approximation to real problem ϕ. then

- **forward error**, $\Delta y = y - \hat{y}$;
- **absolute forward error**, $|\Delta y| = |y - \hat{y}|$; and
- **relative forward error**, $\delta y = \frac{|\Delta y|}{|y|}$.

code, visual: forward error

```
if __name__ == "__main__":

```



ii) backward error

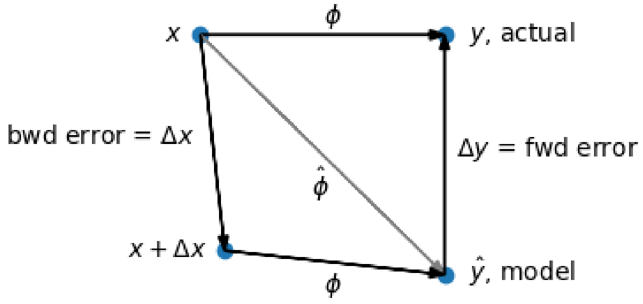
suppose $x, \Delta x$ such that $\hat{\phi}(x) = \hat{y} = \phi(x + \Delta x)$. then

- **backward error**, $\Delta x = \Delta x_{\text{min}}$ where $\hat{\phi}(x) = \hat{y} = \phi(x + \Delta x)$;
- **absolute backward error**, $|\Delta x|$; and
- **relative backward error**, $\delta x = \frac{|\Delta x|}{|x|}$.

ie, backward error identifies the ("nearby") problem the algorithm actually solved.

code, visual: backward error

```
if __name__ == "__main__":…
```



why perturb x there?

why is Δx applied to $\phi(x)$ and not $\hat{\phi}(x)$? bc perturbing the latter changes the computed approximation itself rather than identifying a nearby problem for which the original approximation is the exact solution.

backward error analysis is based on the idea that numerical methods introduce errors and the goal is to understand how these errors relate to the original problem. determining the smallest perturbation Δx that makes $\hat{\phi}(x)$ the exact solution to $\phi(x + \Delta x)$, allows quantification of the computation to see how far it has strayed from the original problem and to determine if it is still meaningful.

04 terminology

fyi, this course favors (albeit imperfectly) "actual problem" and "model solution".

code, visual: words words words

```
if __name__ == "__main__":

```

05 stability

i) forward stability

an engineered solution is **forward stable** if there exists $\eta > 0$ such that $||y - \hat{y}|| \leq \eta \times ||y||$.

---

however, forward error analysis is not prevalent bc "true" ϕ is not always readily available. eg, $\sqrt{3} \in \mathbb{R}$ but $\sqrt{3} \notin \mathbb{FP}$.

ie, to implement an algorithm computationally, is to part ways with forward stability analysis at its abstract level. however, computationally, **use pythons native and/or standard libary functions to stand in for "true"** ϕ.

note: $\sqrt{3} \in \mathbb{CAS}$, where $\mathbb{CAS}$ refers to "computer algebra systems" but that is not the same as a set of numbers.

ii) backward stability

an engineered solution is **backward stable** if there exists $\epsilon > 0$ such that $||\Delta x|| \leq \epsilon \times ||x||$ where $\hat{\phi}(x) = \hat{y} = \phi(x + \Delta x)$.

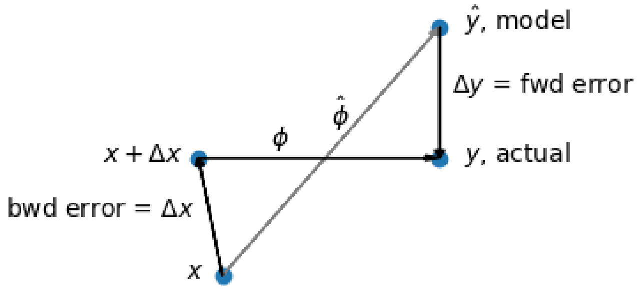
iii) numerical stability

an algorithm is of mixed stability iif there exists a Δx such that both Δx is small and $\phi(x + \Delta x) - \hat{y}$ is small. ie,

an engineered solution is **numerically stable** iif there exists $\eta > 0, \epsilon > 0$ such that $\frac{||y - \hat{y}||}{||y||} \leq \eta$ and $\frac{||\Delta x||}{||x||} \leq \epsilon$, where $y = \phi(x + \Delta x)$ and $\hat{y} = \hat{\phi}(x)$.

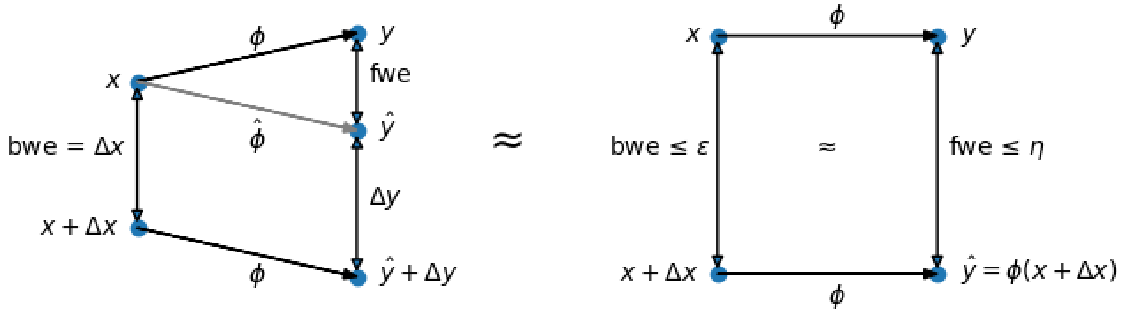
code, visual: numerical stability à la wiki

```
if __name__ == "__main__":

```



code, visual: numerical stability à la higham, corliss

```
if __name__ == "__main__":

```



06 propagation

i) error magnification

error magnification, frequently γ, relates forward and backward error wrt amplification. ie, it quantifies how small errors in input can increase in the final solution.

$$\gamma = \text{forward error}/\text{backward error} = |\Delta y|/|\Delta x|$$

$$\Downarrow \quad \text{or more formally}$$

$$= \lim_{\Delta x \to 0} \sup\nolimits_{\Delta x \leq \epsilon} |\Delta y|/|\Delta x|.$$

ii) condition number

condition number, frequently κ, relates forward and backward error wrt likelihood. ie, it quantifies how small perturbations in input affect the final solution.

$$\kappa = \text{relative forward error}/\text{relative backward error} = |\delta y|/|\delta x|$$

$$\Downarrow \quad \text{or more formally}$$

$$= \lim_{\Delta x \to 0} \sup\nolimits_{\Delta x \leq \epsilon} [\,|\Delta y|/|y|\,]/[\,|\Delta x|/|x|\,].$$

a small condition number indicates a "well-conditioned" system and a large condition number indicates an "ill-conditioned" system wrt stability.

⌄ example 01

consider perturbations wrt $p(x) = 17x^3 + 11x^2 + 2 \Rightarrow \Delta y$.

$$\begin{aligned}\Delta y &= p(x + \Delta x) - p(x) \\ &= [17(x + \Delta x)^3 + 11(x + \Delta x)^2 + 2] - [17x^3 + 11x^2 + 2] \\ &= 51x^2 \Delta x + 51x(\Delta x)^2 + 17(\Delta x)^3 + 22x(\Delta x) + 11(\Delta x)^2\end{aligned}$$

$$\Downarrow \quad |\Delta x| \ll 1, \text{disregard higher orders of } \Delta x$$

$$\Delta y \approx 51x^2 \Delta x + 22x\Delta x.$$

$$\Downarrow \quad \text{consider } x = 1 \pm 0.1$$

$$\Delta y \approx 30 \pm 7.3.$$

$p(1) = 30$, so the ± 7.1 that results from $\Delta x = \pm 0.1$ is inherent to this $p(x)$.

consider 1 as an ideal upper bound for κ.

$$\kappa_{\text{REL}} = |\delta y| / |\delta x| = |7.3/30| / |0.1/1| = 2.4\overline{3} \sim \text{ not great!}$$

$$\kappa_{\text{ABS}} = |\delta y| / |\delta x| = |7.3| / |0.1| = 73 \sim \text{ godawful.}$$

theorem 01. rounding error limit

suppose $i = 1, \ldots, n$ and $0 < \delta_i \leq \mu_{\text{mach}}$ and $\epsilon_i \in \{-1, +1\}$. additionally, suppose $n\mu_{\text{mach}} < 1$. then

$$\prod^{n} (1 + \delta_i)^{e_i} = 1 + \Theta_n$$

where $|\Theta_n| \leq \gamma_n \equiv n\mu_{\text{mach}}/(1 - n\mu_{\text{mach}})$. ie, Θ_n **aggregates error and** γ_n **is its bound.**[1]

note: rounding error, $\mu_{\text{mach}} = \frac{1}{2}\epsilon_{\text{mach}}$, where ϵ_{mach} is machine error.

proof-lite

$$\prod^{n} (1 + \delta_i)^{e_i} \leq \prod^{n} (1 + \delta_i) \leq \prod^{n} (1 + \mu_{\text{mach}}) = (1 + \mu_{\text{mach}})^n.$$

$$\leq (1 + n\mu_{\text{mach}})^n$$

$$\leq \frac{1}{1 - n\mu_{\text{mach}}} \quad \text{bc binomial theorem, } n\mu_{\text{mach}} < 1$$

$$\Rightarrow \text{choose } \gamma_n = \frac{1}{1 - n\mu_{\text{mach}}} - 1 = \frac{n\mu_{\text{mach}}}{1 - n\mu_{\text{mach}}}$$

$$\Rightarrow \Theta_n + 1 \leq \gamma_n + 1. \;\; \checkmark$$

theorem 02. dot product in $\mathbb{R}^3$

dot product in $\mathbb{R}^3$ is backward stable in $\mathbb{FP}^3$.

proof

determine backward stability for the computational approximation of the dot product.

note: for simplicity, this proof only considers $\otimes$, which is more expensive then $\oplus$.

$$
\begin{aligned}
&\phi(x,y) = x \cdot y \quad x, y \in \mathbb{R}^3 \quad \text{and} \\
&\hat{\phi}(\hat{x},\hat{y}) = \hat{x} \odot \hat{y} = (\hat{x_1} \otimes \hat{y_1}) + (\hat{x_2} \otimes \hat{y_2}) + (\hat{x_3} \otimes \hat{y_3})
\end{aligned}
$$

$$\Downarrow \quad \text{where } \hat{x}_j = x_j(1+\delta_{x_j}), \hat{y}_j = y_j(1+\delta_{y_j}), \quad \text{representation error}$$

$$
\begin{aligned}
&= \sum_{j=1}^{3} x_j(1+\delta_{x_j}) \otimes y_j(1+\delta_{y_j}) \\
&= \sum_{j=1}^{3} x_j(1+\delta_{x_j}) y_j(1+\delta_{y_j})(1+\delta_{\otimes_j}), \quad \otimes \text{ operation error} \\
&= \sum_{j=1}^{3} x_j y_j(1+\delta_{x_j})(1+\delta_{y_j})(1+\delta_{\otimes_j}) \\
&= \sum_{j=1}^{3} x_j y_j(1+\Theta_{3,j}) \quad \text{order } 3 \text{ per } j, \text{ theorem 01} \\
&= x_1 y_1(1+\Theta_{3,1}) + x_2 y_2(1+\Theta_{3,2}) + x_3 y_3(1+\Theta_{3,3})
\end{aligned}
$$

$$\Downarrow \quad \text{let } \Delta x_j = x_j \Theta_{3,j}$$

$$= \phi(x,y) + \underbrace{y_1 \Delta x_1 + y_2 \Delta x_2 + y_3 \Delta x_3}_{\text{dot product bt } y, \Delta x} = \phi(x,y) + \phi(\Delta x, y)$$

$$\Downarrow \quad \text{let } \Delta y = 0$$

$$= \phi(x + \Delta x, y + \Delta y)$$

$$\Rightarrow \text{choose } \gamma_3 \text{ such that } ||(\Delta x, \Delta y)|| \leq \gamma_3 ||(x,y)|| \text{ and } |\Theta_{3,j}| \leq \gamma_3 = \frac{3\mu_{\text{mach}}}{1-3\mu_{\text{mach}}}, \text{ theorem 01}$$

$$\Rightarrow \therefore \text{ bounded and backward stable. } \blacksquare$$

condition number vs correlation

however, distinguish between stability and correlation. eg, $\kappa \ll 1$ indicates high stability wrt to a system between its modeled inputs to output but the inputs may not correlate (ie, relate linearly) to the output.

matrix condition number

matrix condition number, $cond(A)$, is a specific type of condition number that relates to matrices. for matrix $A = \{a_{ii}\}$,

$$cond(A) = \kappa(A) \geq \frac{\max_i(|a_{ii}|)}{\min_i(|a_{ii}|)}.$$

iii) error magnification vs condition number

error magnification indicates how much an error is amplified; condition number indicates how likely it is for an error to be amplified.

resources

- perturbation theory @wiki
- stability analysis @wiki (this one is way too much tbh, so for fun, yes?)
- condition number @wiki

additional reading for theorem 01.

- mcclure, david. "ch 09, computer rounding errors", *computational physics guide*, editura, 2009.[3]
- mcclure, david. "ch 10, computer rounding errors: applications", *computational physics guide*, editura, 2009.[4]

additional reading for stability. or for anything, really.

- higham, nick. *backward error*. ★
- higham, nick. *numerical stability*. ★

references

1. corliss, richard.
2. *ibid.*
3. may be searched at source: portland state university.
4. may be searched at source: portland state university.