nmi

2024 spring : lecture 02 : error

quality when theres quantity

the rational number ⅓ exists but it does not exist in the set known as FP. instead, it is approximated by the nearest FPN. ie, hello errors.

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

notation, fps

operators for use with FPS.

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

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

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

absolute error, $\Delta x = |x - x_c|$ and

relative error, $\delta x = \Delta x / |x|$.

real vs engineered

eg, x_c is FPN approximation of x, $y_c = y + \Delta y$, usw.

an engineered problem (eg, rounding) has two kinds of error:

1. forward, wrt how well the engineered problem approximates the real problem; and
2. backward, wrt how the desired results relate back to the expected input.

error analysis

hold that thought…

consider $y = \varphi(x) \Rightarrow y_c = \varphi_c(x)$.

$\Rightarrow$ forward error, $\Delta y = y - y_c$,
$\Rightarrow$ forward absolute, $|\Delta y| = |y - y_c|$,
$\Rightarrow$ **forward relative**, $\delta y = |\Delta y|/|y| = \eta$.

forward stability. engineered problem is forward stable if there exists $\eta > 0$ such that $||y - y_c|| \leq \eta * ||y||$.

backward stability. engineered problem is backward stable if there exists $\varepsilon > 0$ such that $y_c = \varphi(x+\Delta x)$ where $||\Delta x|| \leq \varepsilon * ||x||$. ie, y_c exactly solves nearby real problem y with backward relative error ε.

mixed forward-backward stability $\Leftrightarrow$ numerically stable. engineered problem is numerically stable if there exists $\eta > 0$, $\varepsilon > 0$ such that $y_c + \Delta y = \varphi(x+\Delta x)$, where $||\Delta y|| \leq \eta * ||y||$, $||\Delta x|| \leq \varepsilon * ||x||$.

condition number

this is a basic understanding of κ,ɣ. it gets more complicated, so

hold this thought, too.

error magnification, ɣ, and condition number, κ, relate forward and backward error.

ɣ = forward error / backward error = $|\Delta y| / |\Delta x|$.

κ = relative fwd error / relative bwd error = $|\delta y| / |\delta x|$.

large κ ~ ill-conditioned; small κ ~ well-conditioned.

wrt limits, as $\Delta x \to 0$, so

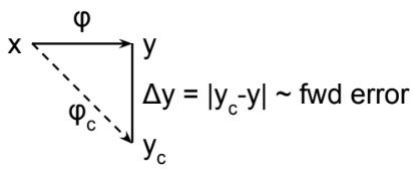
$$\gamma = \lim_{\Delta x \to 0} \sup_{\Delta x \leq \varepsilon} |\Delta y| / |\Delta x| \text{ and}$$

$$\kappa = \lim_{\Delta x \to 0} \sup_{\Delta x \leq \varepsilon} [\, |\Delta y| / |y| \,] / [\, |\Delta x| / x| \,].$$

forward stability analysis

reference problem (RP) is "true" problem. eg, $y = \varphi(x)$, where x is input, φ is operation and y is output. eg, $\varphi(x) = x^3 + x - 1$.

engineered problem (EP) is approximation. eg, $y_c = \varphi_c(x) \Rightarrow y_c \approx \varphi(x)$, where φ_c is approximate operation and y_c is output.



forward stability analysis: define conditions for stability by putting an upper bound on $||\Delta y|| = ||\varphi(x) - \varphi_c(x)||$. ie, $||\Delta y|| \leq \eta||y||$ for some η.

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

ie, if youre going to implement a complex algorithm with a computer, youve already parted ways with fws at its abstract level.

note: this is why i said you could use pythons native and/or library functions when you need to calculate / approximate K, etc, later on.

note: $\sqrt{3} \in$ (computational) symbolic systems. both python and mathematica have this to some degree. mathematica is far more advanced but also has big caching issues - ie, it does what it wants lots.

backward stability analysis

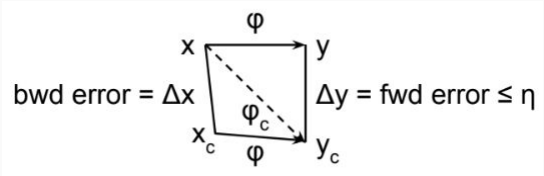
x_c approximation of x

Δx, absolute error (rounding, etc)

δx, relative error

ie, Δx, δx are perturbations $\Rightarrow$ x_c is perturbed value of x.

$x_c = x_c \Rightarrow x_c = x_c - x + x = \Delta x + x \Rightarrow x(1+\Delta x/x) = x(1 + \delta x)$.



backward stability analysis: how does perturbation in x effect φ, φ_c?

ie, $\varphi(x + \Delta x) = \varphi_c(x)$, where smallest Δx is backward error. this is reflecting forward error back into backward error.

ie, "when we modify RP φ to get EP φ_c , for what set of data have we actually solved the problem?"

ie, solving $\varphi_c(x) = \varphi(\Delta x+x)$ for $\Delta x \leq \varepsilon$ then thats backward stable.

perturbations, condition

eg, $p(x) = 17x^3 + 11x^2 + 2 \Rightarrow \Delta y$

$= p(x+\Delta x) - p(x)$
$= (17(x+\Delta x)^3 + 11(x+\Delta x)^2 + 2) - (17x^3 + 11x^2 + 2)$
$= 51x^2\Delta x + 51x(\Delta x)^2 + 17(\Delta x)^3 + 22x\Delta x + 11(\Delta x)^2.$

let $|\Delta x| \ll 1 \Rightarrow$ disregard higher orders of Δx.

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

consider $x = 1 \pm 0.1$

$\Rightarrow \Delta y \approx 30 \pm 7.3$ vs $p(1) = 30$.

specifically, the ±7.3 that results from ±0.1 is inherent to this p(x). ie, $\varphi(x) = p(x)$. however, $\varphi_c(x) \approx \varphi(x)$ is also an operation subject to inherent conditions.

consider condition number where ~1 is ideal.

$K_{REL} = |\delta y|/|\delta x| = |7.3/30| \,/\, |0.1/1| = 2.4333\ldots$ ~ not great!

$K_{ABS} = |\Delta y|/|\Delta x| = |7.3|/|0.1| = 73$ ~ godawful.

theorem 01 wrt bound, aggregate error

suppose $i = 1,\ldots,n$ and $0 < \delta_i \le \mu_M$ and $e_i \in \{-1,+1\}$. additionally suppose $n\mu_M < 1$. then

$$\prod^n(1 + \delta_i)^{e\_i} = 1 + \Theta_n,$$

where $|\Theta_n| \le \gamma_n \coloneqq n\mu_M / (1 - n\mu_M)$. **ie, Θ_n aggregates error and γ_n is its bound.**

note: μ_M is rounding error. in FPS, $\mu_M = ½\ \varepsilon_M$, machine error.

proof-lite. (its just a sketch.)

$\prod^n(1 + \delta_i)^{e\_i} \le \prod^n(1 + \delta_i) \le \prod^n(1 + n\mu_M) = (1 + n\mu_M)^n$.

by binomial theorem,

$(1 + n\mu_M)^n \le n\mu_M / (1 - n\mu_M)$.

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

proof.

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

RP $\varphi(x,y) = x \cdot y$; EP $\varphi_c(x_c,y_c) = x_c \odot y_c$
$= (x_{c\_1} \otimes y_{c\_1}) + (x_{c\_2} \otimes y_{c\_2}) + (x_{c\_3} \otimes y_{c\_3})$.

$x_{c\_j} = x_j(1+\delta_{j\_x})$, $y_{c\_j} = y_j(1+\delta_{j\_y})$, representation error

$\Rightarrow \varphi_c(x_c,y_c) = \sum^3 x_j(1+\delta_{j\_x}) \otimes y_j(1+\delta_{j\_y})$

$= \sum^3 x_j(1+\delta_{j\_x})y_j(1+\delta_{j\_y})(1+\delta_{j\_\otimes})$, $\otimes$ error
$= \sum^3 x_jy_j(1+\delta_{j\_x})(1+\delta_{j\_y})(1+\delta_{j\_\otimes})$

$= \sum^3 x_jy_j(1+\Theta_{3\_j})$, order 3 per j ~ th 01
$= x_1y_1(1+\Theta_{3\_1}) + x_2y_2(1+\Theta_{3\_2}) + x_3y_3(1+\Theta_{3\_3})$
$= x_1y_1 + x_2y_2 + x_3y_3 + x_1y_1\Theta_{3\_1} + x_2y_2\Theta_{3\_2} + x_3y_3\Theta_{3\_3}$

let $\Delta x_j = x_j\,\Theta_{3\_j}$

$= \varphi(x,y) + y_1\Delta x_1 + y_2\Delta x_2 + y_3\Delta x_3 = \varphi(x,y) + \varphi(\Delta x,y)$

bc its the dot product bt y and Δx; let $\Delta y = 0$

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

$||(\Delta x,\Delta y)|| \le \gamma_3||(x,y)||$ bc $|\Theta_{3\_ej}| \le \gamma_3$ th 01

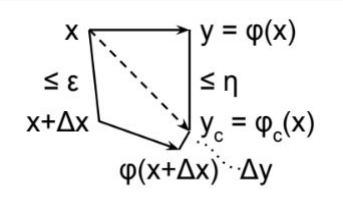
$\Rightarrow$ bounded $\Rightarrow$ bws. ∎

numerical stability ~ mixed fwd-bwd

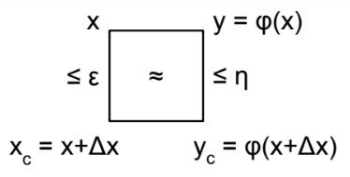
φ_c is numerically stable if $\exists\ \eta > 0, \varepsilon > 0$ st $y_c + \Delta y = \varphi(x+\Delta x)$ where $||\Delta y|| \le \eta||y||$, $||\Delta x|| \le \varepsilon||x||$.

...

ie, a small perturbation in x results in a small perturbation of y.



≈



vocabulary

precision @wiki

significant digits @wiki

*In numerical analysis, **accuracy** is also the nearness of a calculation to the true value; while **precision** is the resolution of the representation, typically defined by the number of decimal or binary digits. (wiki)*

*Significant figures, also referred to as **significant digits**, are specific digits within a number written in positional notation that carry both reliability and necessity in conveying a particular quantity. When presenting the outcome of a measurement (such as length, pressure, volume, or mass), if the number of digits exceeds what the measurement instrument can resolve, only the number of digits within the resolution's capability are dependable and therefore considered significant. (wiki)*

ie & eg, given "0.012345", it is precise to the 6th decimal and of 5 significant digits.

resources, lecture

additional resources:

recursion @ibm @geeksforgeeks
comparison with iteration @khan

error analysis @wiki

numerical stability @wiki

condition number @wiki

primary resources:

numerical analysis by tim sauer; and

math 685, hunter college, spring 2023 with vincent martinez.

for later:

matrix condition number by nick higham.

next time

bisection

fixed point iteration

newtons method

secant method

homework 01

due tuesday, january 6, noon

submit via blackboard

1. code conversion from decimal to binary.

note: check your work with pythons native conversion but you must code the algorithm.