CONŢINUTUL CURSULUI:

1. Erori: definiţie, surse, exemple. Erori în date şi erori computaţionale. Erori de trunchiere şi erori de rotunjire. Stabilitate şi condiţionare. Analiza propagării erorilor.
2. Reprezentarea numerelor în virgulă mobilă (forma ştiinţifică standard): numere în virgulă mobilă; normalizare; proprietăţi; eroare de rotunjire; precizia maşinii. Exemple.
3. Sisteme de ecuaţii liniare inferior/superior triunghiulare: metoda substituţiei ascendente/descendente.
4. Metoda de eliminare Gauss (MEG):
   (i) MEG fără pivotare: descriere; algoritm; limitări; condiţii necesare şi suficiente;
   (ii) MEG cu pivotare parţială: descriere; algoritm; limitări;
   (iii) MEG cu pivotare parţială scalată: descriere; algoritm;
   (iv) MEG cu pivotare totală: descriere; algoritm.
5. Metoda Gauss-Jordan: descriere; algoritm.
6. Factorizarea LU fără pivotare: teorema de existenţă şi unicitate; algoritm; legătura cu MEG fără pivotare; condiţii necesare şi suficiente.
7. Factorizările LDU şi LDL^{T}: teorema de existenţă şi unicitate; algoritm.
8. Factorizarea LU cu pivotare (PLU): teorema de existenţă; algoritm; legătura cu MEG cu pivotare parţială (scalată).
9. Factorizarea Cholesky: proprietăţi ale matricelor simetrice şi pozitiv definite (SPD); algoritm; condiţii necesare şi suficiente.
10. Sisteme de ecuaţii liniare supradeterminate (problema celor mai mici pătrate):
    (i) definiţia soluţiei în sensul celor mai mici pătrate;
    (ii) sitemul de ecuaţii normale asociat;
    (iii) rezolvare folosind factorizarea Cholesky; algoritm;
    (iv) aplicaţie: problema regresiei liniare; limitare.
11. Factorizarea QR: metoda Gram-Schmidt clasică; metoda Gram-Schmidt modificată.
12. Metoda reflexiilor (Householder). Aplicaţii: factorizarea QR prin metoda reflexiilor (Hauseholder); factorizarea QR cu pivotare prin metoda reflexiilor (Hauseholder).
13. Metoda rotaţiilor (Givens). Aplicaţie: factorizarea QR prin metoda rotaţiilor (Givens).
14. Descompunerea valorilor singulare (DSV): existenţa DSV; interpretare geometrică; algoritm DSV (naiv); instabilitatea algoritmului; proprietăţi; matrice de rang nemaxim şi DSV.
15. Analiza senzitivităţii/stabilităţii sistemelor de ecuaţii liniare pătratice: perturbări şi inverse ale matricelor; număr de condiţionare a unei matrice; estimare riguroasă; estimare riguroasă pe componente.

BIBLIOGRAFIE:

1. Kendall E. Atkinson, *An Introduction to Numerical Analysis*, Second Edition. John Wiley & Sons, London, 1989.
2. Gene H. Golub, Charles F. Van Loan, *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, 1996.
3. Michael T. Heath, *Scientific Computing. An Introductory Survey.* SIAM, Philadelphia, 2018.
4. Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser, *Data Structures and Algorithms in Python.* John Wiley & Sons, USA, 2013.
5. Robert Johansson, *Numerical Python*, Second Edition. APress, USA, 2019.
6. Hans Petter Langtangen, *A Primer on Scientific Programming with Python*, Fourth Edition. Springer Verlag, Heidelberg, 2014.

NOTARE:

VERIFICARE SCRISĂ: săptămâna#14; **120 minute**

$$\boxed{\begin{aligned}\textbf{NOTA} = \text{Rotunjire/Trunchiere}\Big[\, & 0,60\times\ (\textbf{Nota Verificare Scrisă}) \\ & + 0,40\times\ (\textbf{Nota Laborator}) + \textbf{Bonus}\,\Big]\end{aligned}}$$

- **Nota Verificare Scrisă** $\in [1, 10]$ – obţinută la verificarea scrisă;
- **Nota Laborator** $\in [1, 10]$ – cf. rezultatelor de la testul/testele de laborator, activităţii de la laborator şi a temelor primite şi rezolvate;
- **Bonus** $\in [0, 1]$ – cf. răspunsurilor date la curs, de rezolvarea unor teme/probleme suplimentare etc.;
- **Condiţie necesară de promovare: NOTA** ≥ 5;
- **Nota Laborator**, obţinută în timpul semestrului, se ia considerare atât în sesiunea mai-iunie 2024, cât şi în sesiunile de restanţe şi/sau reexaminări din 2024.

CURS#1

1. Erori: definiţie, surse, exemple. Erori în date şi erori computaţionale. Erori de trunchiere şi erori de rotunjire. Stabilitate şi condiţionare. Analiza propagării erorilor.
2. Reprezentarea numerelor în virgulă mobilă (forma ştiinţifică standard): numere în virgulă mobilă; normalizare; proprietăţi; eroare de rotunjire; precizia maşinii. Exemple.

ERORI: DEFINIȚIE, SURSE, EXEMPLE

DEFINIȚIE:

Fie $x^* \in \mathbb{R}$ o valoare exactă și $\tilde{x} \in \mathbb{R}$ o aproximare a lui $x^* \in \mathbb{R}$.

(i) Sn eroare absolută a aproximării $\tilde{x}$ a lui x^*

$$e_a(\tilde{x}) := |x^* - \tilde{x}|$$

(ii) Sn eroare relativă a aproximării $\tilde{x}$ a lui $x^* \in \mathbb{R} \setminus \{0\}$

$$e_r(\tilde{x}) := \frac{|x^* - \tilde{x}|}{|x^*|} = \frac{e_a(\tilde{x})}{|x^*|}$$

(iii) Spunem că aproximarea $\tilde{x}$ a lui $x^* \in \mathbb{R} \setminus \{0\}$ are $m \in \mathbb{N}$ cifre semnificative în baza 10 dacă

$$e_r(\tilde{x}) \leq 0,5 \times 10^{-m}$$

SURSE ALE ERORII

1) Modelarea matematică a problemei: de regulă, se fac ipoteze simplificatoare care modifică problema originală → ERORI DE MODELARE (MODELLING ERRORS)

2) Inacurateţea datelor problemei: de regulă, se fac erori umane, inerente, de măsurare a datelor problemei, i.e. DATE PERTURBATE (NOISY DATA)

3) Reprezentarea numerelor în virgulă mobilă (floating point / v.m.) → ERORI DE ROTUNJIRE (ROUNDING ERRORS)

4) <u>Aproximarea numerică a problemei</u>

(ie folosirea unor algoritmi) ⟶

ERORI DE TRUNCHIERE
(TRUNCATION ERRORS)

5) <u>Programarea metodei numerice</u>

ERORI DE PROGRAMARE
(BLUNDERS)

<u>EXEMPLUL #1:</u>

Aria suprafeței Pământului: $4\pi r^2$

- Pământul este modelat ca o sferă de rază r ⟶ 1) modelare
- $r \simeq 6.370$ km ⟶ 2) măsurători
- $\pi \simeq 3,14$ ⟶ 3) reprezentare în v.m.
- Calculul formulei ariei unei sfere ⟶ 3) calcul prin reprezentare în v.m.

ERORI ÎN DATE. ERORI COMPUTAȚIONALE

O problemă clasică este calculul valorii unei funcții, e.g. $f: \mathbb{R} \to \mathbb{R}$ (se poate generaliza $f: \mathbb{R}^n \to \mathbb{R}^m$):

$x \in \mathbb{R}$ – data exactă;

$f(x) \in \mathbb{R}$ – rezultatul exact;

$\hat{x} \in \mathbb{R}$ – dată perturbată (aproximare);

$\hat{f}: \mathbb{R} \to \mathbb{R}$ – aproximare a rezultatului exact (algoritm + reprezentare în v.m.)

$$\text{EROAREA TOTALĂ} := \hat{f}(\hat{x}) - f(x)$$

$$= \underbrace{\left[\hat{f}(\hat{x}) - f(\hat{x})\right]}_{\text{EROARE COMPUTAȚIONALĂ}} + \underbrace{\left[f(\hat{x}) - f(x)\right]}_{\text{EROARE ÎN DATE}}$$

EROARE COMPUTAȚIONALĂ – depinde de algoritm și de reprezentare în v.m.

EROARE ÎN DATE – depinde de date, nu de algoritm

ERORI DE TRUNCHIERE, ERORI DE ROTUNJIRE

Fie $\tilde{f}: \mathbb{R} \to \mathbb{R}$ funcția care definește aproximarea teoretică (algoritmul) pt calculul lui $f: \mathbb{R} \to \mathbb{R}$.

În acest caz, $\hat{f}(\hat{x})$ este reprezentarea în v.m. a aproximării teoretice $\tilde{f}(\hat{x})$

$$\boxed{\hat{f}(\hat{x}) = fl(\tilde{f}(\hat{x}))}.$$

În consecință, obținem:

EROAREA COMPUTAȚIONALĂ =

$$\hat{f}(\hat{x}) - f(\hat{x}) =$$

$$fl(\tilde{f}(\hat{x})) - f(\hat{x}) =$$

$$\underbrace{[fl(\tilde{f}(\hat{x})) - \tilde{f}(\hat{x})]}_{\text{EROARE DE ROTUNJIRE}} + \underbrace{[\tilde{f}(\hat{x}) - f(\hat{x})]}_{\text{EROARE DE TRUNCHIERE}}$$

EROARE DE ROTUNJIRE — depinde de reprezentarea în v.m. a numerelor

EROARE DE TRUNCHIERE — depinde de algoritm

STABILITATE ȘI CONDIȚIONARE

Dificultățile de rezolvare a unei probleme pot fi cauzate de problema însăși și nu de formula de aproximare a sa (algoritm).

<u>DEFINIȚIE:</u>

O problemă sn <u>stabilă / bine condiționată</u> dacă o perturbare relativă a datelor problemei generează o perturbare relativă a soluției rezonabilă (de ordin apropiat).

În caz contrar, problema sn <u>instabilă / rău condiționată</u>.

<u>DEFINIȚIE:</u>

Sn <u>număr de condiționare</u> pentru problema $f: \mathbb{R} \to \mathbb{R}$ în $x \in \mathbb{R}^*$

$$\boxed{\operatorname{cond}(f) := \frac{|(f(\hat{x}) - f(x)) / f(x)|}{|(\hat{x} - x) / x|}}$$

unde $\hat{x}$ este în vecinătatea lui x.

OBSERVAȚIE:

Problema $f:\mathbb{R}\to\mathbb{R}$ este instabilă / / rău condiționată dacă $\mathrm{cond}(f) >> 1$ și stabilă / bine condiționată în caz contrar.

EXEMPLUL #2 (evaluarea unei funcții):

$f:\mathbb{R}\to\mathbb{R}$ funcție

$x\in\mathbb{R}$ dată exactă

$\hat{x} = x+h \in\mathbb{R}$ dată perturbată

$$e_a(f(\hat{x})) := f(\hat{x}) - f(x)$$
$$= f(x+h) - f(x) \simeq h f'(x)$$

$$e_r(f(\hat{x})) := \frac{f(x+h) - f(x)}{f(x)} \simeq h\frac{f'(x)}{f(x)}$$

$$\mathrm{cond}(f) := \frac{|e_r(f(\hat{x}))|}{|e_r(\hat{x})|} \simeq \frac{|h f'(x)/f(x)|}{|h/x|}$$
$$= \left|\frac{x f'(x)}{f(x)}\right|$$

$f : \mathbb{R} \to \mathbb{R}, \ f(x) = e^x \implies$

$\operatorname{cond}(f) \simeq |x|$

<u>EXEMPLUL #3</u> (instabilitate):

Fie problema calculului lui $\cos \hat{x}$ cu $\hat{x}$ în vecinătatea lui $x = \frac{\pi}{2}$.

$\hat{x} = x + h$, h mic

$$\cos(x+h) - \cos x \simeq -h \sin x$$

$$= -h \sin \frac{\pi}{2} = -h$$

$$\frac{\cos(x+h) - \cos x}{\cos x} \simeq -h \operatorname{tg} x \simeq \infty$$

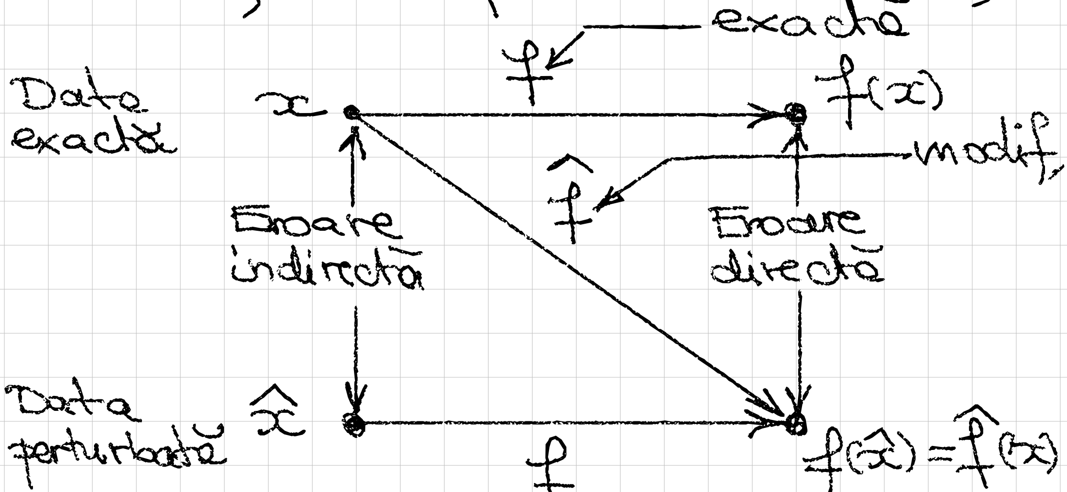
$\implies$ problemă <u>instabilă</u>

$$\cos(1{,}57079) = 0{,}63267949 \times 10^{-5}$$

$$\cos(1{,}57078) = 1{,}63267949 \times 10^{-5}$$

ANALIZA PROPAGĂRII ERORII

Analiza directă a erorii (*forward error analysis*) este adeseori dificilă și poate duce la estimări ale erorii pesimiste. Alternativ, se poate utiliza *analiza indirectă a erorii* (*backward error analysis*) în care se consideră soluția aproximativă obținută ca soluție exactă pentru o problemă modificată și se analizează influența modificării problemei inițiale asupra rezultatului obținut.



EXEMPLUL #4 (analiza indirectă a erorii):

$$f(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad \text{exactă}$$

$$\hat{f}(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \quad \text{modificată}$$

$x = 1$ dată exactă

Analiza directă a erorii:

$$\hat{f}(x) - f(x)$$

Analiza indirectă a erorii:

$$\hat{x} \in \mathbb{R} : f(\hat{x}) = \hat{f}(x)\Big|_{x=1}$$

$$e^{\hat{x}} = \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6}\right)\Big|_{x=1}$$

$$\hat{x} = \ln\left(1 + x + \frac{x^2}{2} + \frac{x^3}{6}\right)\Big|_{x=1}$$

$$\hat{x} = \ln 2{,}666667 = 0{,}980829$$

Eroarea directă $= f(\hat{x}) - f(x)$

$= -0{,}051615$ ✓

Eroarea indirectă $= \hat{x} - x$

$= -0{,}019171$ ✓

$\boxed{\text{același mesaj}}$

REPREZENTAREA NUMERELOR ÎN VIRGULĂ MOBILĂ (V.M. / FLOATING POINT)

1) NUMERE ÎN V.M.

În computer, numerele reale sunt reprezentate în <u>virgulă mobilă</u> (v.m.).

Formal, un număr real $x \in \mathbb{R}$ este <u>reprezentat în v.m. în baza β</u> ca

$$\boxed{fl(x) = \pm\left(d_0 + \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \dots + \frac{d_{t-1}}{\beta^{t-1}}\right)\beta^e}$$

$\beta \in \mathbb{N}$, $\beta \geq 2$ – <u>baza</u>

$t \in \mathbb{N}$ – <u>precizia</u>

$e \in [L, U]$ – <u>exponentul</u>, $L, U \in \mathbb{Z}$

$0 \leq d_i \leq \beta - 1$, $i = \overline{0, t-1}$

(cifre în baza β)

Succesiunea de cifre în baza β

$m := (d_0, d_1 d_2 \dots d_t)_\beta$ sn <u>mantisa</u>

În computer, semnul, exponentul și mantisa sunt memorate în câmpuri separate alocate reprezentării în v.m. a lui $x \in \mathbb{R}$, fiecare având o lungime fixă.

Numărul $0 \in \mathbb{R}$ este reprezentat în mod unic prin $(d_0, d_1 d_2 \dots d_{t-1})_\beta = (0,0 0 \dots 0)_\beta$ și $e = 0$.

Majoritatea computerelor folosesc $\beta=2$, dar au fost folosite în trecut $\beta=16$ (IBM) și $\beta=3$ (Rusia).

SISTEM	β	t	L	U
~~IEGE~~ SP	2	24	−126	127
~~IEEE~~ DP	2	53	−1.022	1.023
Cray	2	48	−16.383	16.384

2) NORMALIZARE

- Un număr real $x \in \mathbb{R}$ este <u>reprezentat în v.m. normalizat în baza β</u> dacă $d_0 \neq 0$ cu excepția lui $x = 0 \in \mathbb{R}$.
Astfel, mantisa lui $x \in \mathbb{R}$ satisface

$$1 \leq m < \beta$$

- O convenție alternativă pt <u>reprezentarea în v.m. normalizată în baza β</u> a numărului $x \in \mathbb{R}$ este $d_0 = 0$ și $d_1 \neq 0$.
În acest caz, mantisa lui $x \in \mathbb{R}$ satisface relația:

$$\beta^{-1} \leq m < 1$$

- Numerele reprezentate în v.m. se normalizează pt:
  - a avea o reprezentare unică;
  - a mări precizia (prin memorare de 0).

3) PROPRIETĂȚI ALE NUMERELOR ÎN V.M.

- Mulțimea numerelor în v.m. normalizate în baza β este <u>finită și discretă</u>.

$$\boxed{\#\text{numere în v.m. normalizate în baza } \beta = 2(\beta-1)\beta^{t-1}(U-L+1)+1}$$

<u>EX</u>: Justificați!

- Cel mai mic număr în v.m. normalizat în baza β <u>pozitiv</u> este

  Underflow level $= UFL = \beta^{L}$

  și definește așa-numitul <u>nivel de nedeterminare inferioară</u> (<u>underflow level</u>) $\Rightarrow$ $x \in \mathbb{R}$, $|x| < UFL$, <u>nu</u> poate fi reprezentat în v.m.!

- Cel mai mare număr în v.m. normalizat în baza β este

  Overflow level $= OFL = \beta^{U+1}(1-\beta^{-t})$

și definește așa-numitul <u>nivel de nedeterminare superioară</u> (<u>overflow level</u>). $\Rightarrow$

$x \in \mathbb{R}$, $|x| > OFL$, <u>nu</u> poate fi reprezentat în v.m!

- Nu toate numerele reale, $x \in \mathbb{R}^*$, admit o reprezentare <u>exactă</u> în mulțimea numerelor reprezentate în v.m. normalizate în baza β (floating point nr.), $fl(x) \in \mathbb{M} \subset \mathbb{R}$.

4) EROARE DE ROTUNJIRE
(ROUNDING / ROUNDOFF ERROR)

Oricărui număr real, $x \in \mathbb{R}^*$, care nu este reprezentat exact în mulțimea numerelor reprezentate în v.m. normalizate în baza β, notată cu $\mathbb{M} \subset \mathbb{R}$, i se asociază cel mai apropiat element $fl(x) \in \mathbb{M}$:

$$x \in \mathbb{R}^* \longmapsto fl(x) \in \mathbb{M}.$$

- Acest procedeu s.n. <u>rotunjire</u> (<u>rounding</u>) și induce <u>eroarea de rotunjire</u> (<u>rounding / roundoff error</u>):

$$\boxed{e_r(x) := \frac{|x - fl(x)|}{|x|}}$$

- $fl(x) \in \mathbb{M}$, reprezentarea în v.m. normalizată în baza β cu t cifre semnificative a lui $x \in \mathbb{R}$ s.n. și <u>forma științifică standard cu baza β cu t cifre semnificative</u>.

- Fie $x \in \mathbb{R}^*$ scris în formă științifică standard normalizată în baza β cu infinit de multe cifre semnificative

$$x = \pm (d_0, d_1 d_2 \ldots d_{t-1} d_t \ldots)_\beta \times \beta^e,$$

$d_0 \neq 0$.

<u>Reprezentarea în v.m. normalizată în baza β cu t cifre semnificative în mantisă</u> a lui $x \in \mathbb{R}^*$ se obține prin:

(a) prin <u>TRUNCHIERE (CHOPPING)</u>:

$$\boxed{fl(x) = \pm (d_0, d_1 d_2 \ldots d_{t-1})_\beta \times \beta^e}$$

(b) prin <u>ROTUNJIRE (ROUNDING-UP)</u>:

Se trunchiază numărul

$$x + \beta^{-t+e} = \pm (c_0, c_1 c_2 \ldots c_{t-1} c_t \ldots)_\beta \times \beta^{-e}$$

și se obține

$$\boxed{fl(x) = \pm (c_0, c_1 c_2 \ldots c_{t-1})_\beta \times \beta^{-e}}$$

5) PRECIZIA MAȘINII

Eroarea de rotunjire se poate estima astfel:

$$e_r(x) := \frac{|x - fl(x)|}{|x|} \leq$$

$$\leq \begin{cases} \beta^{1-t} & \text{prin trunchiere} \\ \frac{1}{2}\beta^{1-t} & \text{prin rotunjire} \end{cases} \leq \boxed{\beta^{1-t} =: \epsilon_M}$$

$\epsilon_M := \beta^{1-t}$ sn. precizia mașinii

(machine precision epsilon)

Prin urmare, în ambele situații obținem estimarea erorii de rotunjire:

$$\boxed{e_r(x) := \frac{|x - fl(x)|}{|x|} \leq \epsilon_M, \quad \forall x \in \mathbb{R}^*}$$

CONSECINȚĂ:

$$\boxed{\forall x \in \mathbb{R}, \quad fl(x) = (1+\delta)x, \quad |\delta| \leq \epsilon_M}$$

Dem: EX!

OBSERVAȚIE:

Precizia mașinii ϵ_M se mai poate defini ca cel mai mic număr pozitiv ε pentru care $1+\varepsilon > 1$ în v.m.

Cum determinați numeric (printr-un cod) valoarea lui ϵ_M? **Ex!**

EXEMPLUL #5 (analiza erorilor: eroare de trunchiere, eroare de rotunjire):

Fie $f \in C^2[a,b]$, $h>0$ suficient de mic fixat și $x \in (a,b)$ fixat.

(i) Vrem o aproximare pentru $f'(x)$ SIMPLĂ și ACURATĂ.

—//—

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

$\xi_x \in [x, x+h]$ $\Longrightarrow$

$$f'(x) = \underbrace{\frac{f(x+h)-f(x)}{h}}_{\text{formulă de aproximare}} + \underbrace{\left(-\frac{f''(\xi_x)}{2}h\right)}_{=: e_t(x)}$$

$=: e_t(x)$ eroare de trunchiere

$$f'(x) \simeq \frac{f(x+h)-f(x)}{h}, \quad h>0 \text{ fixat } \underline{\underline{\text{mic}}}$$

(ii) <u>Estimarea erorii de trunchiere</u> :

— 〃 —

$$|e_t(x)| = \left| f'(x) - \frac{f(x+h)-f(x)}{h} \right|$$

$$= \left| -\frac{f''(\xi_x)}{2} h \right| = |f''(\xi_x)| \frac{h}{2} \underset{f\in C^2[a,b]}{\le}$$

$$\le \underbrace{\max_{t\in[a,b]} |f''(t)|}_{=:M} \cdot \frac{h}{2} =: \frac{M}{2} h \Rightarrow$$

$$\boxed{|e_t(x)| \le \frac{M}{2} h}$$

(iii) Evaluarea funcției f se realizează prin folosirea reprezentării sub formă științifică standard normalizată, deci conține o <u>eroare de rotunjire</u> $\tilde{f}(x) = f(x) + e_r(x)$, unde $|e_r(x)| \le \epsilon$, $\forall x \in [a,b]$

Astfel, aproximarea lui $f'(x)$ prin formula de aproximare obținută la (i), ie

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

se realizează, în fapt, prin:

$$f'(x) \simeq \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h} \Rightarrow$$

$$\left| f'(x) - \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h} \right| =$$

$$= \left| f'(x) - \frac{f(x+h) - f(x)}{h} + \frac{f(x+h) - f(x)}{h} - \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h} \right|$$

$$\leq \left| f'(x) - \frac{f(x+h) - f(x)}{h} \right| + \frac{1}{h}\left[\left| f(x+h) - \tilde{f}(x+h) \right| + \left| \tilde{f}(x) - f(x) \right| \right]$$

$$= |e_t(x)| + \frac{1}{h}\left(|e_r(x+h)| + |e_r(x)|\right)$$

$$\leq \frac{Mh}{2} + \frac{1}{h}(\epsilon + \epsilon) = \frac{Mh}{2} + \frac{2\epsilon}{h}$$

$$\Rightarrow \boxed{\left|f'(x) - \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h}\right| \leq \frac{Mh}{2} + \frac{2\epsilon}{h}}$$

(iv) Determinați valoarea optimă a lui $h>0$, h_{opt}, care să minimizeze eroarea totală de la (iii)

$$F : (0, \infty) \longrightarrow \mathbb{R}$$

$$F(h) = \frac{Mh}{2} + \frac{2\epsilon}{h}$$

$$F'(h) = \frac{M}{2} - \frac{2\epsilon}{h^2} \Rightarrow$$

$$F'(h) = 0 \Leftrightarrow h = \pm 2\sqrt{\epsilon / M} \Leftrightarrow$$

$$\boxed{h^* = 2\sqrt{\epsilon / M}}$$ punct de minim pt F

$$F''(h^*) = \frac{4\epsilon}{h^3}\Big|_{h=h^*} = \frac{4\epsilon}{(h^*)^3} > 0 \quad \square$$

EXEMPLUL #6:

Fie $x > y \geq 0$ două numere reale reprezentate sub formă științifică standard normalizată în baza 2 și

$$2^{-q} \leq 1 - \frac{y}{x} \leq 2^{-p}, \quad p, q \in \mathbb{N}.$$

Atunci, prin operația $(x - y)$ se pierd cel mult q și cel puțin p cifre semnificative în baza 2.

Ex!

EXEMPLUL #7 (anularea erorilor / pierderea cifrelor semnificative):

Fie $f : (0, \infty) \to \mathbb{R}$, $f(x) = (\sqrt{x+1} - \sqrt{x})\, x$

a) Calculați $f(500)$ folosind reprezentarea științifică standard normalizată în baza 10 cu 6 cifre semnificative prin rotunjire.

b) Propuneți o alternativă de calcul al lui $f(500)$ mai bună dpdv al acurateței! (raționalizare)

---

a) $\sqrt{500} = 22,3606797 = 0,223607 \times 10^2$

$\sqrt{501} = 22,383029 = 0,223830 \times 10^2$

$500 = 0,5 \times 10^3$

$\boxed{f(500)} = (\sqrt{501} - \sqrt{500})\, 500$

$= (0,223830 \times 10^2 - 0,223607 \times 10^2)$

$\times\ 0,5 \times 10^3$

$= 0,000223 \times 10^2 \times 0,5 \times 10^3$

$= 0,1115 \times 10^2 = \boxed{11,15}$

$\boxed{f^{(exact)}(500)} = 0,111748 \times 10^2 = \boxed{11,1748}$

$$e_r(f(500)) = \frac{|11,1748 - 11,15|}{|11,1748|} =$$

b) $f(x) = \dfrac{x}{\sqrt{x+1} + \sqrt{x}}\Big|_{x=500} = \ldots = \boxed{11,1748}$

EXEMPLUL #8 (propagarea erorilor):

Fie $x, y \in \mathbb{R}^*$ și $\tilde{x}, \tilde{y}$ aproximările acestora aî

$$\begin{cases} x = \tilde{x} + \varepsilon_x, & \left|\frac{\varepsilon_x}{x}\right| << 1 \\ y = \tilde{y} + \varepsilon_y, & \left|\frac{\varepsilon_y}{y}\right| << 1 \end{cases}$$

(ie $\varepsilon_x/x \pm 1 \simeq \pm 1$)

Determinați $e_r(\tilde{x}\tilde{y})$.

//

$$e_a(\tilde{x}\tilde{y}) = |xy - \tilde{x}\tilde{y}|$$
$$= |(\tilde{x} + \varepsilon_x)(\tilde{y} + \varepsilon_y) - \tilde{x}\tilde{y}|$$
$$= |\underline{\tilde{x}\tilde{y}} + \tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x + \varepsilon_x\varepsilon_y - \underline{\tilde{x}\tilde{y}}|$$
$$= |\tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x + \varepsilon_x\varepsilon_y|$$
$$= |(x - \varepsilon_x)\varepsilon_y + (y - \varepsilon_y)\varepsilon_x + \varepsilon_x\varepsilon_y|$$
$$= |x\varepsilon_y + y\varepsilon_x - \varepsilon_x\varepsilon_y| \Longrightarrow$$

$$e_r(\tilde{x}\tilde{y}) = \frac{|x\varepsilon_y + y\varepsilon_x - \varepsilon_x\varepsilon_y|}{|xy|} \leq$$

$$\leq \frac{|x|\cdot|\varepsilon_y| + |y|\cdot|\varepsilon_x| + |\varepsilon_x|\cdot|\varepsilon_y|}{|x|\cdot|y|}$$

$$= \frac{|\varepsilon_y|}{|y|} + \frac{|\varepsilon_x|}{|x|} + \frac{|\varepsilon_x|}{|x|}\cdot\frac{|\varepsilon_y|}{|y|}$$

$$= \frac{|y-\tilde{y}|}{|y|} + \frac{|x-\tilde{x}|}{|x|} + \frac{|x-\tilde{x}|}{|x|}\,\frac{|y-\tilde{y}|}{|y|}$$

$$\simeq \frac{|y-\tilde{y}|}{|y|} + \frac{|x-\tilde{x}|}{|x|} = e_r(\tilde{y}) + e_r(\tilde{x})$$

Pt că $\left|\frac{\varepsilon_x}{x}\right| + \left|\frac{\varepsilon_x}{x}\right|\cdot\left|\frac{\varepsilon_y}{y}\right| =$

$$= \left|\frac{\varepsilon_x}{x}\right|\left(1 + \left|\frac{\varepsilon_y}{y}\right|\right) \simeq \left|\frac{\varepsilon_x}{x}\right|\cdot 1 = \frac{|\varepsilon_x|}{|x|}$$

Am obținut:

$$\boxed{e_r(\tilde{x}\tilde{y}) \leq e_r(\tilde{x}) + e_r(\tilde{y})}$$

<u>CONCLUZIE</u>: $e_r(\tilde{x}\tilde{y})$ se poate controla prin intermediul $e_r(\tilde{x})$ și $e_r(\tilde{y})$, în sensul în care aceste erori <u>nu se propagă foarte rapid</u>.

EXEMPLUL #9 (calcul îmbricat / nested computation):

Evaluați $f(x) = x^3 - 6{,}1x^2 + 3{,}2x + 1{,}5$ în punctul $x = 4{,}71$ folosind operațiile cu numere reprezentate sub formă științifică standard normalizată în baza 10 cu 3 cifre semnificative (prin trunchiere, respectiv rotunjire). Folosiți afișare cu 20 de zecimale!

$$f(x) = x^2(x - 6{,}1) + 3{,}2x + 1{,}5$$
$$= x[x(x-6{,}1) + 3{,}2] + 1{,}5$$

$\#$op. elementare $= 5$

$$f(x) = x \cdot x \cdot x - 6{,}1 \cdot x \cdot x + 3{,}2 \cdot x + 1{,}5$$

$\#$op elementare $= 8$

$\square$