

LUCRAREA 1

ERORI DE CALCUL NUMERIC

1.1. Obiectivele lucrării

În cadrul lucrării se va evidenția modul în care erorile numerice pot fi caracterizate, motivele apariției acestora, precum și maniera în care acestea se propagă. Se vor studia erorile inerente (care provin din simplificarea modelului fizic, pentru a putea fi descris printr-un model matematic), erorile de rotunjire (datorate faptului că în calcule, numerele cu un număr infinit de zecimale se aproximează prin numere cu un număr finit de zecimale) și erorile de trunchiere (datorate aproximării finite a unor procese teoretic infinite).

1.2. Aspecte teoretice

1.2.1. Moduri de exprimare a erorii

Prin valoare aproximativă se înțelege aproximarea numerică a unei mărimi numerice reale. Fie A valoarea exactă a unei mărimi numerice reale și a valoarea aproximativă corespunzătoare aceleiași mărimi, rezultată în urma unui calcul numeric sau determinată pe cale experimentală.

Diferența dintre valoarea exactă a unei mărimi și valoarea aproximativă corespunzătoare aceleiași mărimi se numește **eroare absolută**:

$$E_a = \text{valoarea exactă} - \text{valoarea aproximativă} \quad (1.1)$$

Dacă $E_a > 0$, se spune că *valoarea aproximativă* va aproxima *valoarea exactă* prin

lipsă, iar dacă $E_a < 0$, atunci se spune că *valoarea aproximativă* va aproxima *valoarea exactă* prin adaos.

Exemplu numeric

Dacă se consideră pentru $E = 2,71828182845905$ valoarea exactă, iar pentru valoarea estimată $e = 2,718$, atunci eroarea absolută este:

$$E_a = E - e = 2,71828182845905 - 2,718 = 2,818 \times 10^{-4}.$$

Eroarea relativă este definită ca fiind modulul raportului dintre eroarea absolută și valoarea exactă a mărimii considerate:

$$E_r = \frac{\text{valoarea exactă} - \text{valoarea aproximativă}}{\text{valoarea exactă}} \quad (1.2)$$

În general, se utilizează și exprimarea procentuală a erorii relative, definită de relația:

$$E_r^{\%} = \frac{\text{valoarea exactă} - \text{valoarea aproximativă}}{\text{valoarea exactă}} \cdot 100 \quad [\%] \quad (1.3)$$

Exemplu numeric

Dacă se consideră pentru $E = 2,71828182845905$ valoarea exactă, iar pentru valoarea estimată $e = 2,718$, atunci eroarea relativă este:

$$E_r = \frac{E_a}{e} = 1,037 \times 10^{-4}.$$

Pentru metodele numerice, valoarea exactă va fi cunoscută în situația în care se lucrează cu funcții ce pot fi rezolvate pe cale analitică. Un astfel de caz îl reprezintă analiza teoretică a stării în care se găsește un sistem simplu. Dar, în aplicațiile practice nu se cunoaște a priori soluția problemei. În aceste situații, o alternativă este să se folosească cea mai bună estimare disponibilă a valorii exacte. În cadrul metodelor numerice, erorile de aproximare sunt determinate în absența unor informații privind valoarea exactă. De exemplu, anumite metode numerice folosesc un proces iterativ

pentru a putea determina soluția problemei. În cadrul acestor metode o aproximație curentă este calculată pe baza unei aproximații anterioare. Procesul este iterativ, astfel încât în fiecare iterație vor fi obținute aproximații ale căror valori sunt apropiate de valoarea exactă. Eroarea va fi estimată pe baza diferenței dintre valoarea curentă și cea anterioară a aproximației. Astfel, eroarea relativă procentuală va fi calculată cu relația:

$$E_r = \frac{\text{valoarea aproximativa curenta} - \text{valoarea aproximativa anterioara}}{\text{valoarea aproximativa curenta}} \cdot 100 [\%] \quad (1.4)$$

Semnul valorilor corespunzătoare erorilor absolută și relativă obținute cu ajutorul relațiilor (1.1) și (1.2) poate fi pozitiv sau negativ. Dacă valoarea aproximativă este mai mare decât valoarea exactă, sau valoarea aproximației anterioare este mai mare decât valoarea aproximației curente, eroarea este negativă. În caz contrar eroarea va avea valoarea pozitivă. Totodată, în relațiile (1.2) – (1.4), numitorul poate avea o valoare mai mică decât zero, situație în care eroarea va avea din nou semnul negativ. În cadrul metodelor numerice ne interesează dacă valoarea absolută a erorii relative procentuale este mai mică decât o valoare prestabilită, e_s . Astfel, calculul va fi repetat până în momentul în care este satisfăcută condiția:

$$|E_r| < e_s \quad (1.5)$$

1.2.2. Tipuri de erori

Se pot evidenția următoarele tipuri de erori:

Erorile de metodă (trunchiere) sunt legate exclusiv de metodele numerice utilizate pentru soluționarea modelului matematic. Acestea se datorează faptului că formulele și ecuațiile exacte se înlocuiesc cu formule și ecuații aproximative, pentru a permite calculul printr-un număr finit de operații aritmetice. Astfel, eroarea de trunchiere nu se poate calcula exact, dar poate fi estimată.

O categorie importantă de procese numerice infinite o reprezintă seriile Taylor. Acestea, fiind serii de puteri, pot fi utilizate la evaluarea funcțiilor elementare (sinus,

cosinus, exponențială, logaritmică etc.) și a celor speciale (de exemplu Bessel) prin reducerea la operații aritmetice elementare de tipul adunării și înmulțirii.

De exemplu, dezvoltările în serie Taylor pentru diverse funcții sunt următoarele:

$$\begin{aligned}
 e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \\
 \sin(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \\
 \cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \\
 \ln(x) &= (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3!} - \frac{(x-1)^4}{4!} + \dots
 \end{aligned} \tag{1.6}$$

Erorile de rotunjire se datorează faptului că în calculele numerele cu un număr infinit de zecimale se aproximează prin numere cu un număr finit de zecimale. Pe de altă parte, orice calculator operează doar cu un număr finit de numere reale (evident cu un număr finit de zecimale), în timp ce mulțimea numerelor reale este infinită. Apare astfel o aproximare inevitabilă a numerelor reale care intervin în calcule cu numerele reprezentate în calculator.

Modul uzual de reprezentare a unui număr real într-un sistem de calcul este de forma:

$$A = m \cdot 10^e \tag{1.7}$$

unde: m se numește mantisă, iar e exponent.

Cu excepția numărului nul, la care $m = 0$, în rest mantisa satisface relația:

$$0,1 \leq |m| \leq 1 \tag{1.8}$$

Dacă presupunem că numărul exact este reprezentat conform relației (1.7)

$$A = \pm 0, j j j j j i i i \dots \cdot 10^{\pm ij}$$

unde j și i sunt cifre (1, ...9), iar modul uzual de reprezentare într-un sistem de calcul este de forma:

$$A = \pm 0, j j j j j \dots \cdot 10^{\pm ij}$$

se poate spune că cifrele i sunt pierdute prin rotunjire.

Astfel, eroarea relativă de rotunjire depinde de numărul de cifre semnificative folosite în reprezentarea numărului real și nu de valoarea numărului.

Erorile inerente sunt cele care provin din simplificarea modelului fizic, pentru a putea fi descris printr-un model matematic. Aceste erori sunt legate de cunoașterea aproximativă a unor valori provenite din măsurători (aparatele de măsură lucrează inevitabil cu anumite abateri, citirea greșită a valorilor de pe scala aparatului de înregistrare, necunoașterea scalei aparatului de măsură, transmiterea eronată a datelor, înregistrarea greșită a datelor în tabelul de măsurători etc.) sau din faptul că avem de-a face cu numere iraționale (algebrice sau transcendente: numerele π , e , $\sqrt{2}$ etc.). Astfel, numărul π poate fi înregistrat ca dată inițială cu valorile: 3,14; 3,141592; 3,1415926535, ceea ce înseamnă că din start se pornește cu o eroare inerentă.

Evident, rezultatul oricăror calcule depinde și de precizia datelor introduse inițial. Deoarece în sistemele numerice de calcul cele mai complicate proceduri se reduc în final la operații aritmetice, trebuie analizată propagarea erorilor în cazul celor patru operații aritmetice elementare. Ca erori inerente pot fi considerate și erorile de conversie făcute la trecerea în baza 2 a unor numere care se introduc în memoria calculatoarelor.

1.2.3. Propagarea erorilor

Propagarea erorilor reprezintă un aspect deosebit de important pentru calculul numeric. O eroare care poate apărea la efectuarea unei operații aritmetice se propagă spre următoarele operații, într-o măsură mai mare sau mai mică, în funcție de felul și ordinea acestor operații.

Fie două valori numerice a și f , care aproximează mărimile exacte A și F . Erorile absolute și relative ale celor două mărimi numerice au expresiile:

$$E_A = A - a, \quad E_F = F - f \quad (1.9)$$

$$e_A = \left| \frac{E_A}{a} \right|, \quad e_F = \left| \frac{E_F}{f} \right| \quad (1.10)$$

În cele ce urmează se va analiza propagarea erorilor pentru cele patru operații aritmetice de bază.

- **Adunarea.**

Erorile absolută și relativă ale sumei sunt date de relațiile:

$$E_{A+F} = A + F - (a + f) = E_A + E_F \quad (1.11)$$

$$e_{A+F} = \left| \frac{E_{A+F}}{a+f} \right| = \left| \frac{a}{a+f} \frac{E_A}{a} + \frac{f}{a+f} \frac{E_F}{f} \right| = \left| e_A \frac{a}{a+f} + e_F \frac{a}{a+f} \right| \quad (1.12)$$

Eroarea relativă pentru adunare este o sumă ponderată a erorilor introduse la reprezentarea în calculator a cantității sumate. În acest caz se introduce o eroare suplimentară e_s la reprezentarea sumei $a + f$:

$$e_{A+F}^T = \left| e_A \frac{a}{a+f} + e_F \frac{f}{a+f} \right| + e_s \quad (1.13)$$

- **Înmulțirea.**

Erorile absolută și relativă ale produsului sunt date de relațiile:

$$E_{A \cdot F} = A \cdot F - a \cdot f = (a + E_A)(f + E_F) - a \cdot f \approx a \cdot E_F + f \cdot E_A \quad (1.14)$$

$$e_{A \cdot F} = \left| \frac{E_{A \cdot F}}{a \cdot f} \right| = \left| \frac{E_A}{a} + \frac{E_F}{f} \right| = |e_A + e_F| \quad (1.15)$$

În relația (1.14) s-a neglijat termenul $E_A \cdot E_F$, deoarece are o valoare cu mult mai mică în raport cu restul termenilor.

Se poate observa faptul că erorile relative introduse inițial se adună. În afară de aceste erori pot apărea însă noi erori, deoarece produsul $A \cdot F$ poate avea, la rândul său, un număr de cifre semnificative mai mare decât cel admis, necesitând o nouă rotunjire

(simetrică). Notând cu e_p această nouă eroare, eroarea relativă totală e_{A-F}^T la înmulțirea a două numere va fi:

$$e_{A-F}^T = |e_A + e_F + e_p| \quad (1.16)$$

- **Împărțirea.**

Erorile absolută și relativă ale câtului sunt date de relațiile:

$$E_{\frac{A}{F}} = \frac{A}{F} - \frac{a}{f} = \frac{a + E_A}{f + E_F} - \frac{a}{f} = \frac{f \cdot E_A - a \cdot E_F}{f^2 + fE_F} \approx \frac{fE_A - aE_F}{f^2} \quad (1.17)$$

$$e_{\frac{A}{F}} = \left| \frac{E_{\frac{A}{F}}}{\frac{A}{F}} \right| = \left| \frac{E_A}{A} - \frac{E_F}{F} \right| = |e_A - e_F| \quad (1.18)$$

În cadrul relației (1.17) s-a considerat faptul că termenul f^2 are o valoare cu mult mai mare decât $f \cdot E_F$.

Din relația (1.18) rezultă că eroarea relativă la împărțire, datorată erorilor de reprezentare în calculatoare, este egală cu diferența erorilor relative introduse inițial în valorile numărătorului și numitorului. La acestea se adaugă eroarea de rotunjire corespunzătoare reprezentării raportului a/f , pe care o vom nota cu e_d , astfel încât eroarea relativă totală la împărțirea a două numere A și F este:

$$e_{\frac{A}{F}}^T = |e_A - e_F + e_d| \quad (1.19)$$

- **Scăderea.**

Erorile absolută și relativă ale diferenței sunt date de relațiile:

$$E_{A-F} = A - F - (a - f) = E_A - E_F \quad (1.20)$$

$$e_{A-F} = \left| \frac{E_{A-F}}{a - f} \right| = \left| \frac{a}{a - f} \frac{E_A}{a} - \frac{f}{a - f} \frac{E_F}{f} \right| = \left| e_A \frac{a}{a - f} - e_F \frac{f}{a - f} \right| \quad (1.21)$$

Concluziile referitoare la erorile absolută și relativă sunt asemănătoare cu cele de la adunare. Expresia erorii relative indică o alterare pronunțată a preciziei în situațiile când cei doi operanzi au același semn și sunt de valori apropiate.

1.3. Desfășurarea lucrării

1. Se vor studia următoarele aspecte teoretice: determinarea erorii relative de rotunjire a sistemului de calcul, propagarea erorilor inerente și analiza erorii de trunchiere.
2. Să se calculeze erorile absolută, relativă și procentuală pentru aproximarea numerelor π și e cu două cifre semnificative exacte.
3. Fie $a = 4,31 \pm 1\%$, $b = 4,30 \pm 2\%$ și $c = 3,2 \pm 1\%$. Să se calculeze $(a + b) \cdot c$ și să se analizeze rezultatul obținut.
4. Să se studieze propagarea erorilor la adunarea și înmulțirea a două numere oarecare a și b . Erorile absolute ale celor două numere vor fi introduse de la tastatură.
5. Să se evalueze valoarea funcțiilor $\sin(x)$, $\cos(x)$, $\ln(x)$, e^x , pentru care x este introdus de la tastatură, folosind formulele dezvoltării în serie Taylor prezentate în Paragraful 1.2.2, în care se vor considera un număr variabil de termeni. Se vor analiza rezultatele obținute.