Tema 1

Errors

Exemple d'errors

Càlcul aproximat de la massa de la Terra

Usant la llei de la gravitació universal de Newton i la llei de la caiguda lliure dels cossos de Galileu, s'obté la fórmula:

$$M = \frac{gR^2}{G}, \tag{1}$$

on g és l'acceleració de la gravetat, R el radi de la Terra, i G la constant de la gravitació universal. Es disposa dels valors experimentals següents:

$$\overline{g} = 9.80665 \text{ m s}^{-2}, \quad \overline{G} = 6.67428 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}, \quad \overline{R} = 6371.0 \text{ km}.$$

Aplicant la fórmula anterior, resulta l'aproximació $\overline{M} = 5.9639 \cdot 10^{24}$ kg.

Nota $M = 5.9736 \cdot 10^{24}$ *kg* (Wikipedia, NASA).

$M = 5.9742 \cdot 10^{24}$ *kg* (J.M.A. Danby, *Fundamentals of Celestial Mechanics*, Willmann-Bell, Inc., 1992).

Caldria estudiar els errors comesos atenent a les aproximacions donades dels valors experimentals.

Fonts d'error

- Errors de modelització: els models matemàtics són aproximacions de la realitat.
- Errors de truncament: els mètodes numèrics fan aproximacions del model matemàtic.
- Errors experimentals: les mesures de les dades del problema no són exactes i porten errors de diversos tipus:
  - errors aleatoris: les mesures estan afectades per factors "aleatoris" que no podem controlar;
  - errors sistemàtics: les mesures provenen, per exemple, d'una cal·libració incorrecta de l'aparell de mesura;
  - errors aberrants: deguts a errors humans, a canvis sobtats en les condicions de l'experiment, etc.
- Errors d'arrodoniment: les operacions es realitzen amb un nombre finit de xifres (amb l'ajuda d'una calculadora o d'un ordinador).

Error absolut i error relatiu

Definicions i notacions

Sigui x el valor exacte d'una quantitat i $\overline{x}$ un valor aproximat.

- Error absolut en x:

$$e_a(x) := e_a(\overline{x}, x) = x - \overline{x} \ ,$$

- Error relatiu en x:

$$e_r(x) := e_r(\overline{x}, x) = \frac{e_a(x)}{x} \simeq \frac{e_a(x)}{\overline{x}} = \frac{x - \overline{x}}{\overline{x}} \ .$$

Fites d'error:

- $\varepsilon_a(x) := \varepsilon_a(\overline{x}, x))$ és una fita de l'error absolut en x si $|e_a(x)| \leq \varepsilon_a(x)$,
- $\varepsilon_r(x) := \varepsilon_r(\overline{x}, x))$ és una fita de l'error relatiu en x si $|e_r(x)| \leq \varepsilon_r(\mathrm{x})$.

Notació usual:

- $x = \overline{x} \pm \varepsilon_a(x) \iff x \in [\overline{x} - \varepsilon_a(x), \overline{x} + \varepsilon_a(x)]$,
- $x = \overline{x}\,(1 \pm \varepsilon_r(x)) \iff x \in [\overline{x} - \varepsilon_r(x)|\overline{x}|, \overline{x} + \varepsilon_r(x)|\overline{x}|]$.

Error absolut i error relatiu

Exemples

Sigui $a = \sqrt{20000} = 141.4213562...$ i $\overline{a} = 141.4$.

$$e_a(\overline{a}, a) = e_a(141.4, \sqrt{20000}) = 0.02135... < 0.022 \equiv \varepsilon_a(a)\ ,$$

$$e_r(\overline{a}, a) = e_r(141.4, \sqrt{20000}) \simeq \frac{0.02135}{141.4} = 0.00015099... < 0.00016 \equiv \varepsilon_r(a)\ .$$

La primera fita indica que l'error no afecta el primer dígit fraccionari i la segona, que l'error no afecta el tercer dígit significatiu (tot i que tampoc el quart).

Sigui $b = \sqrt{800000} = 894.42719...$ i $\overline{b} = 894.4$.

$$e_a(\overline{b}, b) = e_a(894.4, \sqrt{800000}) = 0.02719.... < 0.028 \equiv \varepsilon_a(b)\ ,$$

$$e_r(\overline{b}, b) = e_r(894.4, \sqrt{800000}) \simeq \frac{0.02719}{894.4} = 0.000030399... < 0.000031 \equiv \varepsilon_r(b)\ .$$

La primera fita indica que l'error no afecta el primer dígit fraccionari i la segona fita, que l'error no afecta el quart dígit significatiu.

Representació de nombres en una base

Definició i exemples

La representació en base $b \geq 2$ d'un nombre real $x \neq 0$ és

$$\begin{aligned} x &= \pm a_{q-1}a_{q-2}\ldots a_0.a_{-1}a_{-2}\ldots_{b)} \quad \text{[}q \text{ xifres abans del punt]} \\ &= \pm(a_{q-1}b^{q-1}+a_{q-2}b^{q-2}+\ldots+a_0+a_{-1}b^{-1}+\ldots)\,, \quad |x| \geq 1 \ (q \geq 1) \\ x &= \pm 0.0\ldots 0a_{q-1}a_{q-2}\ldots_{b)} \quad \text{[}-q \text{ zeros després del punt]} \\ &= \pm(a_{q-1}b^{q-1}+a_{q-2}b^{q-2}+\ldots)\,, \quad |x| < 1 \ (q < 1) \end{aligned}$$

on els a_j $(j < q)$ són xifres en la base b: $0 \leq a_j < b$ amb la primera xifra significativa $a_{q-1} \neq 0$. Quan $b = 10$, no cal indicar la base.

$$\begin{aligned} 107.125 &= 1\cdot 10^2 + 0\cdot 10^1 + 7\cdot 10^0 + 1\cdot 10^{-1} + 2\cdot 10^{-2} + 5\cdot 10^{-3} \\ 0.00333\ldots &= 3\cdot 10^{-3} + 3\cdot 10^{-4} + 3\cdot 10^{-5} + \ldots \\ \pi &= 3\cdot 10^0 + 1\cdot 10^{-1} + 4\cdot 10^{-2} + 1\cdot 10^{-3} + 5\cdot 10^{-4} + \ldots \\ 0.1_{2)} &= 0.5 \\ 0.1_{10)} &= 0.0\overline{0011}_{2)} \end{aligned}$$

Representació de nombres en una base

Exemple de representació en base 2

Representació de $125.1_{10)}$ en base 2.

$$125.1_{10)} = a_{q-1}a_{q-2}\dots a_0.a_{-1}a_{-2}a_{-3}\dots{}_{2)}.$$

amb els bits $a_j \in \{0,1\}$, $(j < q = 3)$.
Representació de la part entera $125_{10)}$ en base 2.

$$
\begin{aligned}
125/2 &= 62 \quad \text{(prenem residu } \mathbf{1}\text{)}\\
62/2 &= 31 \quad \text{(prenem residu } \mathbf{0}\text{)}\\
31/2 &= 15 \quad \text{(prenem residu } \mathbf{1}\text{)}\\
15/2 &= 7 \quad \text{(prenem residu } \mathbf{1}\text{)}\\
7/2 &= 3 \quad \text{(prenem residu } \mathbf{1}\text{)}\\
3/2 &= 1 \text{(pernem quocient } \mathbf{1}\text{)} \quad \text{(prenem residu } \mathbf{1}\text{)}
\end{aligned}
$$

$$125_{10)} = 1111101_{2)}.$$

Representació de nombres

Exemple de representació en base 2

Representació de la part fraccionària $0.1_{10)}$ en base 2.

$$
\begin{array}{ll}
0.1 \times 2 = 0.2 & \text{(prenem part entera } \mathbf{0}\text{)} \\
0.2 \times 2 = 0.4 & \text{(prenem part entera } \mathbf{0}\text{)} \\
0.4 \times 2 = 0.8 & \text{(prenem part entera } \mathbf{0}\text{)} \\
0.8 \times 2 = 1.6 & \text{(prenem part entera } \mathbf{1}\text{)} \\
0.6 \times 2 = 1.2 & \text{(prenem part entera } \mathbf{1}\text{)} \\
0.2 \times 2 = 0.4 & \text{(prenem part entera } \mathbf{0}\text{)} \\
\ldots &
\end{array}
$$

$$0.1_{10)} = 0.0\overline{0011}_{2)}$$

Representació binària completa

$$125.1_{10)} = 1111101.0\overline{0011}_{2)}\ .$$

Representació decimal en punt flotant

Definició

Fent flotar el punt decimal q posicions en la representació decimal d'$x \neq 0$:

$$x = \pm 0.\alpha_1\alpha_2 \ldots \cdot 10^q = \pm m \cdot 10^q \, , \quad \alpha_j = a_{q-j} \in \{0, 1, \ldots 9\} \;\; (j > 0)$$

Aquesta representació decimal en punt flotant d'x ve donada per:

- el nombre enter q, anomenat exponent, i
- el nombre real m, tal que $0.1 \leq m < 1$, anomenat mantissa.

El primer dígit fraccionari $\alpha_1 \neq 0$ d'm és el primer dígit significatiu d'x.

Exemples:

- $g = 9.80665 = 0.980665 \cdot 10^1$, l'exponent és 1 i la mantissa, 0.980665.
- $G = 6.67428 \cdot 10^{-11} = 0.667428 \cdot 10^{-10}$, l'exponent és -10 i la mantissa, 0.667428.

Representació decimal en punt flotant

Representació aproximada en calculadores

Les calculadores poden emmagatzemar una quantitat finita de dígits. Per tant, no es poden emmagatzemar totes les mantisses (ni tots els exponents).

Si sols disposem de t dígits per a la mantissa, la representació

$$x = \pm 0.\alpha_1\alpha_2\ldots\alpha_t\alpha_{t+1}\ldots \cdot 10^q = \pm m\, 10^q, \text{ amb } \alpha_1 \neq 0,$$

pot ser aproximada, arrodonint el darrer dígit t, per $\mathrm{fl}_t(x)$, flotant d'x amb t dígits significatius i arrodoniment:

- $\mathrm{fl}_t(x) = \pm 0.\alpha_1\alpha_2\ldots\alpha_t \cdot 10^q$, si $\alpha_{t+1} < 5$;
- $\mathrm{fl}_t(x) = \pm \mathrm{fl}_t\left((0.\alpha_1\alpha_2\ldots\alpha_t + 10^{-t})\cdot 10^q\right)$, si $\alpha_{t+1} \geq 5$.

Exemple: Sigui $x = 0.999527 \cdot 10^1$. Llavors:

$$\mathrm{fl}_5(x) = 0.99953 \cdot 10^1, \quad \mathrm{fl}_4(x) = 0.9995 \cdot 10^1, \quad \mathrm{fl}_3(x) = 0.100 \cdot 10^2$$

Representació decimal en punt flotant

Errors d'arrodoniment

Observem que, atenent a les definicions:

$$|e_a(\mathrm{fl}_t(x), x)| = |x - \mathrm{fl}_t(x)| \leq \frac{1}{2}10^{-t}10^q = \frac{1}{2}10^{q-t} =: \varepsilon_a(\mathrm{fl}_t(x), x)\ .$$

$\frac{1}{2}10^{q-t}$ és una fita de l'error absolut en la representació en punt flotant amb t dígits significatius i arrodoniment de qualsevol nombre real $x \neq 0$ amb exponent q.

Com que $x \neq 0$ i $m \geq 0.1$:

$$|e_r(\mathrm{fl}_t(x), x)| = \frac{|e_a(\mathrm{fl}_t(x), x)|}{|x|} \leq \frac{1}{2}\frac{10^{q-t}}{m\,10^q} \leq \frac{1}{2}10^{1-t} =: \varepsilon_r(\mathrm{fl}_t(x), x)\ .$$

$\frac{1}{2}10^{1-t}$ és una fita de l'error relatiu en la representació en punt flotant amb t dígits significatius i arrodoniment de qualsevol nombre real $x \neq 0$.

Representació decimal en punt flotant

Exemples d'errors d'arrodoniment

- $\mathrm{fl}_6(g) = 0.980665 \cdot 10^1$: $t = 6$, $q = 1$.

$$\varepsilon_a(g) = \frac{1}{2}10^{1-6} = \frac{1}{2}10^{-5}, \qquad \varepsilon_r(g) = \frac{1}{2}10^{1-6} = \frac{1}{2}10^{-5}.$$

- $\mathrm{fl}_6(G) = 0.667428 \cdot 10^{-10}$: $t = 6$, $q = -10$.

$$\varepsilon_a(G) = \frac{1}{2}10^{-10-6} = \frac{1}{2}10^{-16}, \qquad \varepsilon_r(G) = \frac{1}{2}10^{1-6} = \frac{1}{2}10^{-5}.$$

Representació binària en punt flotant

Representació aproximada en ordinadors i errors d'arrodoniment

Els ordinadors poden emmagatzemar una quantitat finita de bits per a les mantisses i exponents.

La representació en punt binari flotant d'un nombre qualsevol $x \neq 0$

$$x = \pm 0.\alpha_1\alpha_2 \ldots \alpha_t\alpha_{t+1} \ldots \cdot 2^q = \pm m\, 2^q, \text{ amb } \alpha_1 = 1,$$

pot ser aproximada, emprant t bits per a la mantisa arrodonint el darrer bit, per $\mathrm{fl}_t(x)$ (flotant d'x amb t bits significatius i arrodoniment):

- $\mathrm{fl}_t(x) = \pm 0.\alpha_1\alpha_2 \ldots \alpha_t \cdot 2^q$, si $\alpha_{t+1} = 0$;
- $\mathrm{fl}_t(x) = \pm \mathrm{fl}_t\left((0.\alpha_1\alpha_2 \ldots \alpha_t + 2^{-t}) \cdot 2^q\right)$, si $\alpha_{t+1} = 1$.

Com que $x \neq 0$ i $m < \frac{1}{2}$,

$$|e_r(\mathrm{fl}_t(x), x)| = \frac{|e_a(\mathrm{fl}_t(x), x)|}{|x|} \leq \frac{1}{2}\frac{2^{q-t}}{m\, 2^q} \leq \frac{1}{2} 2^{1-t} =: \varepsilon_r(\mathrm{fl}_t(x), x)$$

i, per tant, 2^{-t} és una fita de l'error relatiu de la representació en punt flotant amb t bits significatius i arrodoniment de qualsevol nombre real $x \neq 0$.

Representació binària en punt flotant

Formats IEEE de representació en precisió simple i doble

$$x = \pm 1.\alpha_2 \ldots \alpha_t \alpha_{t+1} \ldots\ _{2)} \cdot 2^{q-1} = \pm(1+f)\, 2^{q-1}$$

es representa per:

- $\mathrm{fl}_t(x) = \pm 1.\alpha_2\alpha_3 \ldots \alpha_{t\,2)} \cdot 2^{q-1}$, si $\alpha_{t+1} = 0$;
- $\mathrm{fl}_t(x) = \pm \mathrm{fl}_t\left((1.\alpha_2 \ldots \alpha_{t\,2)} + 2^{-t+1})2^{q-1}\right)$, si $\alpha_{t+1} = 1$.

Format	base (b)	digits (t)	$q_{min}-1$	$q_{max}-1$	bits
IEEE simple	2	24	-126	128	32
IEEE doble	2	53	-1022	1024	64

Taula: Formats IEEE (simple i doble precisió)

IEEE simple	signe (1)	$e = q - 1 + 127$ (8)	mantissa f (23)
IEEE doble	signe (1)	$e = q - 1 + 1023$ (11)	mantissa f (52)

Taula: Distribució de memòria en el format IEEE (simple i doble).

Representació binària en punt flotant

Formats IEEE de representació en precisió simple i doble

- Es guarden els $t-1$ primers bits de la mantissa f, ja que el primer bit de la matissa m és 1.
- Si un nombre real x es pot escriure exactament, es diu que és un nombre de màquina. Altrament tindrà una representació en punt flotant $\mathrm{fl}_t(x) \neq x$ amb error.
- En precisió simple, la fita de l'error relatiu d'arroniment és $2^{-24} \approx 0.6 \cdot 10^{-7}$, es pot garantir gairebé una precisió de 6 dígits significatius amb arrodoniment.
- En precisió doble, la fita d'error relatiu d'arrodoniment és $2^{-53} \approx 1.1 \cdot 10^{-16}$, es pot garantir gairebé una precisió de 16 dígits significatius.
- En IEEE simple, els valors $e = 0$ $(q = -126)$ i $e = 255$ $(q = 127)$ es reserven a NaN (Not a Number) i overflow, respectivament.
- En IEEE doble, els valors $e = 0$ $(q = -1022)$ i $e = 1023$ $(q = 1023)$ es reserven a NaN (Not a Number) i overflow, respectivament.

Representació binària en punt flotant

Formats IEEE de representació en precisió simple i doble

- Com es representa $x = 125.1$ en format IEEE amb precisió simple?

Es té la representació amb punt binari flotant:

$$x = 125.1 = 1111101.0\overline{0011}_{2)} = 1.1111010\overline{0011}_{2)} \cdot 2^6$$

L'exponent desplaçat e es representa en base 2 per
$e = 6 + 127 = 133 = 10000101_{2)}$.
Resulta finalment la representació en memòria usant IEEE simple:

IEEE simple	0	10000101	11110100011001100110011

Representació binària en punt flotant

Èpsilon de la màquina

- $\epsilon = \frac{1}{2}b^{1-t}$ s'anomena èpsilon de la màquina. Coincideix amb el nombre positiu més petit que sumat a 1 dóna diferent de 1, és a dir

$$\epsilon = \min\{\varepsilon : \mathrm{fl}_t(1+\varepsilon) \neq 1\}.$$

  Com més petit és, més precisa és la màquina: la precisió indica el nombre t de xifres significatives correctament representades amb arrodoniment en base b.

- Notem que $\mathrm{fl}_t(1+\varepsilon) = 1$ no vol dir ε sigui igual a 0 sinó que és més petit que l'èpsilon de la màquina.
- Treballant amb $t = 3$ dígits significatius, si

$$x = 0.1 \cdot 10^1 \quad \text{i} \quad y = 0.456 \cdot 10^{-4}$$

  llavors $x + y = x$, però $y \neq 0$.

Problemes numèrics

Operacions aritmètiques usant representació flotant

Les calculadores i els ordinadors, degut a la representació dels nombres en punt flotant, fan els càlculs de manera aproximada.
Això té implicacions importants: l'ordre de les operacions pot afectar el resultat final!.

Exemple: Treballant amb $t = 4$ dígits, si es calcula $a + b + c$, on

$$a = 0.5317 \cdot 10^{-2},\ b = 0.3387 \cdot 10^{2},\ c = -0.3381 \cdot 10^{2},$$

emprant ordenacions diferents, resulta:

$$\mathrm{fl}_4(a + \mathrm{fl}_4(b + c)) = \mathrm{fl}_4(0.5317 \cdot 10^{-2} + 0.6000 \cdot 10^{-1}) = 0.6532 \cdot 10^{-1}$$

$$\mathrm{fl}_4(\mathrm{fl}_4(a + b) + c) = \mathrm{fl}_4(0.3388 \cdot 10^{2} - 0.3381 \cdot 10^{2}) = 0.7000 \cdot 10^{-1}\ .$$

El resultat exacte és $a + b + c = 0.65317 \cdot 10^{-1}$ ens diu que és millor la primera ordenació.

Problemes numèrics

Exemple de cancel·lació

Les dues solucions de l'equació $x^2 - 18x + 1 = 0$ són

$$x_{1,2} = 9 \pm \sqrt{80} = \begin{cases} x_1 = 0.17944271909999916 \cdot 10^2 \\ x_2 = 0.5572809000084121 \cdot 10^{-1} \end{cases}$$

Si prenem $\sqrt{80} = 8.9443$ (és a dir $t = 5$) s'obté

$$x_1 = 9 + 8.9443 = 17.9443 = 0.179443 \cdot 10^2 \quad \text{(6 xifres)},$$

$$x_2 = 9 - 8.9443 = 0.0557 = 0.557 \cdot 10^{-1} \quad \text{(3 xifres!)}).$$

En calcular x_2 hi ha una cancel·lació de dígits, perquè restem dues quantitats que són properes i dóna un resultat significativament erroni.

Propagació d'errors

Causes

Hi ha dues raons (o almenys així es pot pensar) responsables de la propagació de l'error en un procès de càlcul:

- Errors en les dades. Si les dades tenen error, aquest error es propaga al resultat de les operacions.
- Errors en les operacions Encara que se sumin dos nombres de màquina x, y, el resultat representat $\mathrm{fl}_t(x+y)$ pot ser diferent de $x+y$. Les funcions f internes que s'apliquen tenen també errors que es poden considerar errors en les operacions.
- Les dues alhora... Efectivament, els errors de les dades i de les operacions s'acumulen en els resultats intermedis i en el resultat final.

Per simplificar se suposa que les operacions no tenen errors i que, per tant, tots els errors es deuen a la propagació dels errors de les dades.

Fórmula de propagació d'errors

Funcions d'una variable

Teorema del valor mig: Sigui $f : [a, b] \to \mathbb{R}$ una funció contínua, derivable a $]a, b[$. Aleshores existeix un punt $\xi \in (a, b)$, tal que

$$f(b) - f(a) = f'(\xi)(b - a).$$

Aplicació: Propagació d'errors en funcions d'una variable.
Sigui $x \in \mathbb{R}$ i sigui $\overline{x} \approx x$.

Del teorema anterior, es té que

$$e_a\,(f(\overline{x}), f(x)) := f(x) - f(\overline{x}) = f'(\xi)(x - \overline{x}), \quad \xi \in < \overline{x}, x > .$$

Usant que la funció és contínua i suposant que els errors són petits, es té una fórmula aproximada de propagació de l'error:

$$|e_a\,(f(\overline{x}), f(x))\,| \approx |f'(\overline{x})||e_a(\overline{x}, x)|,$$
$$\varepsilon_a\,(f(\overline{x}), f(x)) := M\varepsilon_a(\overline{x}, x), \quad \text{on} \quad M = \max_{\xi \in [\overline{x}-\varepsilon_a, \overline{x}+\varepsilon_a]} |f'(\xi)|.$$

Fórmula de propagació d'errors

Error relatiu: Coeficient de propagació

De les darreres expressions, resulta

$$|e_r\left(f(\overline{x}), f(x)\right)| \approx |\overline{x}|\frac{|f'(\overline{x})|}{|f(\overline{x})|}|e_r\left(\overline{x}, x\right)|, \ (f(x) \neq 0).$$

De fet, el terme $\varphi(x) = |x|\dfrac{|f'(x)|}{|f(x)|}$ s'anomena coeficient de propagació (de l'error relatiu) i caldria controlar-lo en un procès de càlcul. Si podem fitar-lo, és a dir, si podem dir que

$$|x|\frac{|f'(x)|}{|f(x)|} \leq M$$

per alguna $M > 0$, en un entorn de $\overline{x}$, llavors

$$\varepsilon_r\left(f(\overline{x}), f(x)\right) := M\varepsilon_r\left(\overline{x}, x\right).$$

Fórmula de propagació d'errors

Exemples d'aplicació

Càlcul de les arrels de l'equació $x^2 - 18x + 1 = 0$: $x_{1,2} = 9 \pm \sqrt{80}$, ara calculant x_1 amb 4 dígits fraccionaris amb arrodoniment:

$$x_1 = 17.9443 \pm \frac{1}{2} \cdot 10^{-4}$$

i després calculant x_2 així:

$$x_2 = f(x_1) = \frac{1}{x_1} \quad \mapsto \quad \overline{x_2} = f(\overline{x_1}) = \frac{1}{\overline{x_1}} = 0.05572800\ldots$$

Estimació de les fites dels errors

$$\varepsilon_a(\overline{x_2}, x_2) = \varepsilon_a\left(\frac{1}{\overline{x_1}}, \frac{1}{x_1}\right) \simeq \left|\frac{-1}{x_1^2}\right| \varepsilon_a(\overline{x_1}, x_1) \simeq \frac{1}{17.9443^2}\,\frac{1}{2} \cdot 10^{-4} \simeq 0.16 \cdot 10^{-6}$$

$$\varepsilon_r(\overline{x_2}, x_2) = \frac{\varepsilon_a(\overline{x_2}, x_2)}{|x_2|} \simeq 17.9443 \cdot 0.16 \cdot 10^{-6} \simeq 0.29 \cdot 10^{-5}$$

Així, $\varepsilon_a(\overline{x_2}, x_2) \leq 0.5 \cdot 10^{-6}$ indica que $\overline{x_2}$ té 6 xifres decimals correctes i $\varepsilon_r(\overline{x_2}, x_2) \leq 0.5 \cdot 10^{-5}$ indica que $\overline{x_2}$ té 5 xifres significatives correctes.
Hi ha una millora respecte al càlcul anterior de x_2: s'han guanyat dues xifres correctes en el resultat evitant els efectes de cancel·lació.

Fórmula de propagació d'errors

Exemples d'aplicació

Fita de l'error comès en avaluar $f(x) = \ln \cos^2(x)$ en un punt x del qual només coneixem tres dígits correctes $\overline{x} = 0.735$.

- Valor aproximat: $\ln \cos^2(\overline{x}) = -0.5972683\ldots$.
- Fita de l'error en x: $\varepsilon_a(x) = \frac{1}{2}10^{-3}$.
- Derivada de la funció: $f'(x) = -2\tan(x)$.
- Fita del valor absolut de la derivada: $|f'(\xi)|$ per a $\xi \in [0.7345, 0.7355]$ (i.e, $[\overline{x} - \varepsilon_a, \overline{x} + \varepsilon_a]$): com que la funció tangent és creixent i positiva a tot l'interval $]0, \pi/2 \simeq 1.5708[$, llavors $\tan(\xi) \leq \tan(0.7355) \lesssim 0.905$, i

$$|f'(\xi)| \leq 2 \cdot 0.905 = 1.810\ .$$

- Fita de l'error absolut en $f(x)$: aplicant la fórmula de propagació de l'error:

$$\varepsilon_a(f(\overline{x} = 0.735), f(x)) = 1.810\frac{1}{2}10^{-3} = 0.905 \cdot 10^{-3}.$$

$$f(x) = -0.5972683 \pm 0.000905\ .$$

Fórmula de propagació d'errors

Exemple 2

El coeficient de propagació de l'error relatiu φ resulta ser

$$\varphi(x) = x\frac{f'(x)}{f(x)} = x\frac{-2\tan(x)}{\log\cos^2(x)}.$$

Es té que $|\varphi(x)| \leq 2.2$ si $x \approx 0.735$ i per tant

$$\varepsilon_r(f(0.735)) \leq 2.2\varepsilon_r(\overline{x}, x) \leq 2.2\frac{0.5 \cdot 10^{-3}}{0.735} \approx 1.5 \cdot 10^{-3}.$$

$$f(x) = -0.5972683(1 \pm 0.0015)\ .$$

Fórmula de propagació d'errors

Funcions de diverses variable

Teorema del valor mig en diverses variables.

Sigui G un obert de $\mathbb{R}^n$, i $f : G \to \mathbb{R}$ una funció diferenciable sobre G. Siguin $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ dos punts de $G \subset \mathbb{R}^n$ tals que el segment que els uneix està contingut a G. Aleshores existeix un punt ξ d'aquest segment tal que

$$f(y) - f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\xi)(y_i - x_i).$$

Aplicació: Propagació d'errors en funcions de diverses variables

Sigui $x \in \mathbb{G} \subset \mathbb{R}^n$, sigui $\overline{x} \approx x$.

$$e_a(f(\overline{x}), f(x)) \approx \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\overline{x}) e_a(\overline{x_i}, x_i)$$

$$\varepsilon_a(f(\overline{x}), f(x)) := \sum_{i=1}^{n} M_i \varepsilon_a(\overline{x_i}, x_i), \quad \text{on} \quad M_i = \max_{\xi \in [\overline{x} - \varepsilon_a, \overline{x} + \varepsilon_a]^n} \left| \frac{\partial f}{\partial x_i}(\xi) \right|$$

Propagació d'errors en diverses variables

Casos especials

$$f(x,y) = x + y$$

$$\varepsilon_a(\overline{x}+\overline{y}, x+y) = \varepsilon_a(\overline{x},x) + \varepsilon_a(\overline{y},y)$$

$$\varepsilon_r(\overline{x}+\overline{y}, x+y) = \left|\frac{x}{x+y}\right| \varepsilon_r(\overline{x},x) + \left|\frac{y}{x+y}\right| \varepsilon_r(\overline{y},y)$$

$$f(x,y) = xy$$

$$\varepsilon_a(\overline{xy}, xy) \approx |y|\ \varepsilon_a(\overline{x},x) + |x|\ \varepsilon_a(\overline{y},y)$$

$$\varepsilon_r(\overline{xy}, xy) \approx \varepsilon_r(\overline{x},x) + \varepsilon_r(\overline{y},y)$$

$$f(x,y) = x/y$$

$$\varepsilon_a(\overline{x}/\overline{y}, x/y) \approx 1/|y|\ \varepsilon_a(\overline{x},x) + |\frac{x}{y^2}|\ \varepsilon_a(\overline{y},y)$$

$$\varepsilon_r(\overline{x}/\overline{y}, x/y) \approx \varepsilon_r(\overline{x},x) + \varepsilon_r(\overline{y},y)$$

Propagació d'errors en diverses variables

Exemple de càlcul de la massa de la Terra

Error propagat en $M(g, G, R) = \dfrac{gR^2}{G}$, a partir de les aproximacions de les magnituds:

$$\overline{g} = 9.80665\ , \quad \overline{G} = 6.67428 \cdot 10^{-11}\ , \quad \overline{R} = 6371.0 \cdot 10^{3}\ .$$

- Errors en les magnituds suposant que són correctes fins a l'última xifra amb arrodoniment:

$$\varepsilon_a(g) = \frac{1}{2}10^{-5}\ , \quad \varepsilon_a(G) = \frac{1}{2}10^{-16}\ , \quad \varepsilon_a(R) = \frac{1}{2}10^{2}\ .$$

- Fórmula de propagació d'errors en diverses variables:

$$\varepsilon_a(M) \approx \left|\frac{\partial M}{\partial g}\right|(\overline{g}, \overline{G}, \overline{R})\varepsilon_a(g) + \left|\frac{\partial M}{\partial G}\right|(\overline{g}, \overline{G}, \overline{R})\varepsilon_a(G) + \left|\frac{\partial M}{\partial R}\right|(\overline{g}, \overline{G}, \overline{R})\varepsilon_a(R),$$

$$\frac{\partial M}{\partial g} = \frac{R^2}{G}\ , \quad \frac{\partial M}{\partial G} = -\frac{gR^2}{G^2}\ , \quad \frac{\partial M}{\partial R} = \frac{2gR}{G}\ .$$

Propagació d'errors en diverses variables

Exemple de càlcul de la massa de la Terra

- Derivades parcials de $M(g, G, R)$ en les aproximacions:

$$\begin{aligned}
\frac{\partial M}{\partial g}(\overline{g}, \overline{G}, \overline{R}) &\approx 6.0815 \cdot 10^{23} \, , \\
\frac{\partial M}{\partial G}(\overline{g}, \overline{G}, \overline{R}) &\approx -8.9357 \cdot 10^{34} \, , \\
\frac{\partial M}{\partial R}(\overline{g}, \overline{G}, \overline{R}) &\approx 1.8722 \cdot 10^{18} \, .
\end{aligned}$$

- Fita aproximada de l'error en M:

$$\varepsilon_a(M) \approx 1.0112 \cdot 10^{20}.$$

$$M \approx 5.96391525 \cdot 10^{24} \pm 1.0112 \cdot 10^{20}$$

$$\Longleftrightarrow M \in [5.96381413 \cdot 10^{24}, 5.96401637 \cdot 10^{24}].$$

Propagació dels errors amb aritmètica intervalar

Exemple de càlcul de la massa de la Terra

L'aritmètica intervalar té una visió diferent a la del teorema del valor mig.

- $g = 9.80665$, $\varepsilon_a(g) = \frac{1}{2}10^{-5}$: $g \in [9.806645, 9.806655] = [g_m, g_M]$;
- $G = 6.67428 \cdot 10^{-11}$,
  $\varepsilon_a(G) = \frac{1}{2}10^{-16}$: $G \in [6.674275 \cdot 10^{-11}, 6.674285 \cdot 10^{-11}] = [G_m, G_M]$;
- $R = 6371.0 \cdot 10^3$,
  $\varepsilon_a(R) = \frac{1}{2}10^{2}$: $R \in [6.37095 \cdot 10^6, 6.37105 \cdot 10^6] = [R_m, R_M]$.

$$M \in [\frac{g_m R_m^2}{G_M}, \frac{g_M R_M^2}{G_m}] \to M \in [5.9638141 \cdot 10^{24}, 5.9640164 \cdot 10^{24}]$$

coincideix pràcticament amb l'anterior ja que els errors són petits

$$M \in [5.9638143 \cdot 10^{24}, 5.96401637 \cdot 10^{24}].$$

Mètodes estables i inestables

Exemple de recurrència inestable

Es volen calcular les integrals

$$R_n = \int_0^1 x^n e^{x-1}\,\mathrm{dx}.$$

S'observa que:

- $R_0 = 1 - \exp(-1)$.
- $0 < R_n < \frac{1}{n+1}$ per a tot $n \geq 0$.
- $R_n = 1 - nR_{n-1}$ (integració per parts).

Es proposa el mètode de càlcul recurrent següent per al càlcul d'un R_N:

- $R_0 = 1 - \exp(-1)$.
- $R_n = 1 - nR_{n-1}$, $n = 0, ..., N$.

Mètodes estables i inestables

Exemple de recurrència inestable

n	R_n
1	3.6787944117144233340e-01
2	2.6424111765711153320e-01
3	2.0727664702865540041e-01
4	1.7089341188538339834e-01
5	1.4553294057308800829e-01
10	8.3877070058292702020e-02
15	5.9033793641901866607e-02
18	-2.9453670751536265020e-02

Taula: $R_{18} < 0$ no té cap xifra significativa correcta!. Tots els càlculs s'han fet usant format de dades double en llenguatge *C*.

Mètodes estables i inestables

Exemple de recurrència inestable

Anàlisi dels errors

Si $e_0 = R_0 - \overline{R_0}$ l'error absolut inicial en la dada R_0,

$$e_n = R_n \overline{R}_n - R_n = 1 - nR_{n-1} - 1 + n\overline{R}_{n-1} = -n(R_{n-1} - \overline{R}_{n-1}) = -ne_{n-1},$$

L'error de R_N,

$$e_N = (-1)^N N! e_0,$$

es fa molt gran quan N augmenta, independentment d'e_0.
El mètode recurrent és inestable i no permet el càlcul de les integrals per a N gran.

Algorismes estables i inestables

Exemple de recurrència estable

Capgirant la recurrència, es té la recurrència inversa:

$$R_n = 1 - nR_{n-1} \to R_{n-1} = \frac{1 - R_n}{n}$$

Per calcular un R_N, es proposa utilitzar la recurrència inversa a partir d'un R_M apropiat:

- $R_M = 0.$
- $R_{n-1} = \frac{1-R_n}{n}$, $n = M, ..., N + 1.$

Anàlisi de l'error propagat des de R_M fins a R_N.

$$e_{n-1} = -\frac{1}{n}e_n \Rightarrow e_N = (-1)^{M-N}\frac{1}{M(M-1)\cdots(N+1)}e_M.$$

Com que l'error inicial és $e_M < \frac{1}{M+1}$, la recurrència inversa troba R_N amb un error de propagació, que es fa més petit a cada pas, i que es pot fitar per:

$$|e_N| < \frac{1}{(M+1)M(M-1)\cdots(N+1)} .$$

Mètodes estables i inestables

Exemple de recurrència estable

n	R_n	$\lvert e_n \rvert$
40	0	2.3e-2
35	2.7046289710763393721e-02	2.9e-10
30	3.1279673932168072796e-02	7.4e-18
25	3.7086214423739237436e-02	4.3e-25
20	4.5544884075818053986e-02	6.8e-32
18	5.0119854958094255126e-02	1.83-34
15	5.9017540879297773766e-02	3.6e-38
10	8.3877070103394166256e-02	1.02e-43

Taula: Els càlculs fets amb un mètode estable (recurrència inversa).

Problemes mal condicionats

Són problemes on la solució depèn de manera molt sensible de les dades.

Exemple:
El sistema d'equacions

$$\begin{aligned} 2.0000x + 0.6667y &= 2.6667, \\ 1.0000x + 0.3333y &= 1.3333, \end{aligned}$$

té solució $x = 1.0000$, $y = 1.0000$, mentre que el sistema

$$\begin{aligned} 2.0000x + 0.6665y &= 2.6667, \\ 1.0000x + 0.3333y &= 1.3333, \end{aligned}$$

té solució $x = 1.6666$, $y = -1.0000$.