ANALISI IN AVANTI DEL I ORDINE DEGLI ERRORI DI ARROTONDAMENTO

Si basa sul teorema fondamentale sulle operazioni tra numeri finiti: se $a, b \in F(\beta, t, L, U)$,

$$fl(a \bullet b) = (a \bullet b)(1 + \epsilon) \qquad |\epsilon| \leq u$$

ove u è la precisione di macchina. Si assume che i dati siano esatti, ossia che appartengano a F.

TECNICA IN AVANTI. Si calcola l'errore relativo sul risultato finale in termine degli errori introdotti dalle singole operazioni, trascurando i termini in cui compaiono prodotti di errori (ANALISI DEL I ORDINE).

ESEMPIO. $\varphi(a,b,c) = a+b+c$

- **ALGORITMO 1.**

$$\begin{aligned}
fl(a+b) &= (a+b)(1+\epsilon_1) = y \qquad |\epsilon_1| \leq u \\
fl(y+c) &= (y+c)(1+\epsilon_2) = z \qquad |\epsilon_2| \leq u \\
fl(a+b+c) &= z = (y+c)(1+\epsilon_2) = \\
&= ((a+b)(1+\epsilon_1)+c)(1+\epsilon_2) = \\
&= (a+b)(1+\epsilon_1)(1+\epsilon_2) + c(1+\epsilon_2) = \\
&\simeq (a+b)(1+\epsilon_1+\epsilon_2) + c(1+\epsilon_2) = \\
&= a+b+c+(a+b)\epsilon_1 + (a+b+c)\epsilon_2
\end{aligned}$$

$(a+b)\epsilon_1\epsilon_2$ **trascurabile**

$$\boldsymbol{\epsilon_{alg1} = \frac{fl(a+b+c)-(a+b+c)}{a+b+c} \simeq \frac{a+b}{a+b+c}\epsilon_1 + \epsilon_2}$$

Fattori di amplificazione dell'errore: $\left\{ \begin{array}{ll} \frac{a+b}{a+b+c} & \textbf{per } \epsilon_1 \\ 1 & \textbf{per } \epsilon_2 \end{array} \right.$

Si definisce INDICE ALGORITMICO I_{alg} **la somma dei valori assoluti dei fattori di**

amplificazione dei singoli errori introdotti da ciascuna operazione:

$$\boldsymbol{I_{alg1} = \left| \frac{a+b}{a+b+c} \right| + 1}$$

Il fattore di amplificazione dell'ultima operazione eseguita è sempre 1.

- **ALGORITMO 2.** $a + (b + c)$

$$fl(b + c) \quad = \quad (b + c)(1 + \epsilon_3) = w \qquad |\epsilon_3| \leq u$$

- **ALGORITMO 2.** $a + (b + c)$

$$\begin{aligned} fl(b+c) &= (b+c)(1+\epsilon_3) = w \qquad |\epsilon_3| \leq u \\ fl(a+w) &= (a+w)(1+\epsilon_4) = v \qquad |\epsilon_4| \leq u \end{aligned}$$

- **ALGORITMO 2.** $\boldsymbol{a + (b + c)}$

$$\begin{aligned}
\boldsymbol{fl(b+c)} &= \boldsymbol{(b+c)(1+\epsilon_3) = w} \qquad \boldsymbol{|\epsilon_3| \le u} \\
\boldsymbol{fl(a+w)} &= \boldsymbol{(a+w)(1+\epsilon_4) = v} \qquad \boldsymbol{|\epsilon_4| \le u} \\
\boldsymbol{fl(a+b+c)} &= \boldsymbol{v} \\
&= \boldsymbol{(a+w)(1+\epsilon_4)}
\end{aligned}$$

- **ALGORITMO 2.** $a + (b + c)$

$$\begin{aligned}
fl(b+c) &= (b+c)(1+\epsilon_3) = w \qquad |\epsilon_3| \leq u \\
fl(a+w) &= (a+w)(1+\epsilon_4) = v \qquad |\epsilon_4| \leq u \\
fl(a+b+c) &= v \\
&= (a+w)(1+\epsilon_4) \\
&= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4)
\end{aligned}$$

- **ALGORITMO 2.** $a + (b + c)$

$$
\begin{aligned}
fl(b+c) &= (b+c)(1+\epsilon_3) = w \quad |\epsilon_3| \leq u \\
fl(a+w) &= (a+w)(1+\epsilon_4) = v \quad |\epsilon_4| \leq u \\
fl(a+b+c) &= v \\
&= (a+w)(1+\epsilon_4) \\
&= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3)(1+\epsilon_4) + a(1+\epsilon_4)
\end{aligned}
$$

- **ALGORITMO 2.** $a + (b + c)$

$$\begin{aligned} fl(b+c) &= (b+c)(1+\epsilon_3) = w \quad |\epsilon_3| \leq u \\ fl(a+w) &= (a+w)(1+\epsilon_4) = v \quad |\epsilon_4| \leq u \\ fl(a+b+c) &= v \\ &= (a+w)(1+\epsilon_4) \\ &= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4) \\ &= (b+c)(1+\epsilon_3)(1+\epsilon_4) + a(1+\epsilon_4) \\ &= (b+c)(1+\epsilon_3+\epsilon_4+\epsilon_3\epsilon_4) + a(1+\epsilon_4) \end{aligned}$$

- **ALGORITMO 2.** $a + (b + c)$

$$
\begin{aligned}
fl(b+c) &= (b+c)(1+\epsilon_3) = w \quad |\epsilon_3| \le u \\
fl(a+w) &= (a+w)(1+\epsilon_4) = v \quad |\epsilon_4| \le u \\
fl(a+b+c) &= v \\
&= (a+w)(1+\epsilon_4) \\
&= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3)(1+\epsilon_4) + a(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3+\epsilon_4+\epsilon_3\epsilon_4) + a(1+\epsilon_4) \\
&\simeq (b+c)(1+\epsilon_3+\epsilon_4) + a(1+\epsilon_4)
\end{aligned}
$$

- **ALGORITMO 2.** $a + (b + c)$

$$
\begin{aligned}
fl(b+c) &= (b+c)(1+\epsilon_3) = w \qquad |\epsilon_3| \le u \\
fl(a+w) &= (a+w)(1+\epsilon_4) = v \qquad |\epsilon_4| \le u \\
fl(a+b+c) &= v \\
&= (a+w)(1+\epsilon_4) \\
&= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3)(1+\epsilon_4) + a(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3+\epsilon_4+\epsilon_3\epsilon_4) + a(1+\epsilon_4) \\
&\simeq (b+c)(1+\epsilon_3+\epsilon_4) + a(1+\epsilon_4) \\
&= a+b+c+(b+c)\epsilon_3 + (a+b+c)\epsilon_4
\end{aligned}
$$

- **ALGORITMO 2.** $a + (b + c)$

$$\begin{aligned}
fl(b+c) &= (b+c)(1+\epsilon_3) = w \qquad |\epsilon_3| \le u \\
fl(a+w) &= (a+w)(1+\epsilon_4) = v \qquad |\epsilon_4| \le u \\
fl(a+b+c) &= v \\
&= (a+w)(1+\epsilon_4) \\
&= (a+(b+c)(1+\epsilon_3))(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3)(1+\epsilon_4) + a(1+\epsilon_4) \\
&= (b+c)(1+\epsilon_3+\epsilon_4+\epsilon_3\epsilon_4) + a(1+\epsilon_4) \\
&\simeq (b+c)(1+\epsilon_3+\epsilon_4) + a(1+\epsilon_4) \\
&= a+b+c+(b+c)\epsilon_3 + (a+b+c)\epsilon_4
\end{aligned}$$

$$\epsilon_{alg2} = \frac{fl(a+b+c)-(a+b+c)}{a+b+c} \simeq \frac{b+c}{a+b+c}\epsilon_3 + \epsilon_4$$

Fattori di amplificazione dell'errore: $\left\{ \begin{array}{ll} \frac{b+c}{a+b+c} & \text{per } \epsilon_3 \\ 1 & \text{per } \epsilon_4 \end{array} \right.$

$$I_{alg2} = \left|\frac{b+c}{a+b+c}\right| + 1$$

Se $I_{alg1} < I_{alg2}$ allora l'algoritmo 1 è più stabile dell'algoritmo 2.

$$a = .2337126\ 10^{-4}, b = .3367843\ 10^{2}, c = -0.3367781\ 10^{2}:$$

$$I_{alg1} = |\frac{a+b}{a+b+c}| + 1 = 0.510^{5} + 1$$

$$I_{alg2} = |\frac{b+c}{a+b+c}| + 1 = .96 + 1$$

Il secondo è più stabile per i valori assunti dai dati.

$$\Downarrow$$

Un algoritmo può essere più stabile di un altro per i valori assunti dai dati.

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$
\begin{aligned}
fl(a^2) &= a * a(1+\epsilon_1) = x \qquad |\epsilon_1| \le u \\
fl(b^2) &= b * b(1+\epsilon_2) = y \qquad |\epsilon_2| \le u
\end{aligned}
$$

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$\begin{array}{lcl} fl(a^2) & = & a * a(1+\epsilon_1) = x \quad |\epsilon_1| \leq u \\ fl(b^2) & = & b * b(1+\epsilon_2) = y \quad |\epsilon_2| \leq u \\ fl(x-y) & = & (x-y)(1+\epsilon_3) \quad |\epsilon_3| \leq u \end{array}$$

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$\begin{aligned}
fl(a^2) &= a * a(1+\epsilon_1) = x \quad |\epsilon_1| \le u \\
fl(b^2) &= b * b(1+\epsilon_2) = y \quad |\epsilon_2| \le u \\
fl(x-y) &= (x-y)(1+\epsilon_3) \quad |\epsilon_3| \le u \\
fl(a^2-b^2) &= (a^2(1+\epsilon_1) - b^2(1+\epsilon_2))(1+\epsilon_3)
\end{aligned}$$

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$\begin{array}{lcl} fl(a^2) & = & a * a(1+\epsilon_1) = x \quad |\epsilon_1| \leq u \\ fl(b^2) & = & b * b(1+\epsilon_2) = y \quad |\epsilon_2| \leq u \\ fl(x-y) & = & (x-y)(1+\epsilon_3) \quad |\epsilon_3| \leq u \\ fl(a^2-b^2) & = & (a^2(1+\epsilon_1) - b^2(1+\epsilon_2))(1+\epsilon_3) \\ & = & a^2(1+\epsilon_1)(1+\epsilon_3) - b^2(1+\epsilon_2)(1+\epsilon_3) \end{array}$$

ESEMPIO. $\varphi(a, b) = a^2 - b^2$

- **ALGORITMO 1.**

$$
\begin{aligned}
fl(a^2) &= a * a(1+\epsilon_1) = x \quad |\epsilon_1| \le u \\
fl(b^2) &= b * b(1+\epsilon_2) = y \quad |\epsilon_2| \le u \\
fl(x-y) &= (x-y)(1+\epsilon_3) \quad |\epsilon_3| \le u \\
fl(a^2-b^2) &= (a^2(1+\epsilon_1) - b^2(1+\epsilon_2))(1+\epsilon_3) \\
&= a^2(1+\epsilon_1)(1+\epsilon_3) - b^2(1+\epsilon_2)(1+\epsilon_3) \\
&= a^2(1+\epsilon_1+\epsilon_3+\epsilon_1\epsilon_3) - b^2(1+\epsilon_2+\epsilon_3+\epsilon_2\epsilon_3)
\end{aligned}
$$

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$
\begin{aligned}
fl(a^2) &= a * a(1+\epsilon_1) = x \qquad |\epsilon_1| \le u \\
fl(b^2) &= b * b(1+\epsilon_2) = y \qquad |\epsilon_2| \le u \\
fl(x-y) &= (x-y)(1+\epsilon_3) \qquad |\epsilon_3| \le u \\
fl(a^2-b^2) &= (a^2(1+\epsilon_1) - b^2(1+\epsilon_2))(1+\epsilon_3) \\
&= a^2(1+\epsilon_1)(1+\epsilon_3) - b^2(1+\epsilon_2)(1+\epsilon_3) \\
&= a^2(1+\epsilon_1+\epsilon_3+\epsilon_1\epsilon_3) - b^2(1+\epsilon_2+\epsilon_3+\epsilon_2\epsilon_3) \\
&\simeq a^2 - b^2 + a^2\epsilon_1 - b^2\epsilon_2 + (a^2-b^2)\epsilon_3
\end{aligned}
$$

ESEMPIO. $\varphi(a,b) = a^2 - b^2$

- **ALGORITMO 1.**

$$\begin{aligned}
fl(a^2) &= a * a(1+\epsilon_1) = x \quad |\epsilon_1| \le u \\
fl(b^2) &= b * b(1+\epsilon_2) = y \quad |\epsilon_2| \le u \\
fl(x-y) &= (x-y)(1+\epsilon_3) \quad |\epsilon_3| \le u \\
fl(a^2-b^2) &= (a^2(1+\epsilon_1) - b^2(1+\epsilon_2))(1+\epsilon_3) \\
&= a^2(1+\epsilon_1)(1+\epsilon_3) - b^2(1+\epsilon_2)(1+\epsilon_3) \\
&= a^2(1+\epsilon_1+\epsilon_3+\epsilon_1\epsilon_3) - b^2(1+\epsilon_2+\epsilon_3+\epsilon_2\epsilon_3) \\
&\simeq a^2 - b^2 + a^2\epsilon_1 - b^2\epsilon_2 + (a^2-b^2)\epsilon_3
\end{aligned}$$

$$\begin{aligned}
\epsilon_{alg1} &= \frac{fl(a^2-b^2)-(a^2-b^2)}{a^2-b^2} \\
&\simeq \frac{a^2}{a^2-b^2}\epsilon_1 - \frac{b^2}{a^2-b^2}\epsilon_2 + \epsilon_3
\end{aligned}$$

$$I_{alg1} = \left| \frac{a^2}{a^2-b^2} \right| + \left| \frac{b^2}{a^2-b^2} \right| + 1$$

- **ALGORITMO 2.**$\varphi(a,b) = (a+b)(a-b)$

$$\begin{aligned} fl(a+b) &= (a+b)(1+\epsilon_4) = z \qquad |\epsilon_4| \leq u \\ fl(a-b) &= (a-b)(1+\epsilon_5) = v \qquad |\epsilon_5| \leq u \end{aligned}$$

- **ALGORITMO 2.** $\varphi(a,b) = (a+b)(a-b)$

$$\begin{aligned} fl(a+b) &= (a+b)(1+\epsilon_4) = z \quad |\epsilon_4| \leq u \\ fl(a-b) &= (a-b)(1+\epsilon_5) = v \quad |\epsilon_5| \leq u \\ fl(zv) &= zv(1+\epsilon_6) \quad |\epsilon_6| \leq u \end{aligned}$$

- **ALGORITMO 2.** $\varphi(a,b) = (a+b)(a-b)$

$$\begin{array}{lcl}
fl(a+b) & = & (a+b)(1+\epsilon_4) = z \quad |\epsilon_4| \leq u \\
fl(a-b) & = & (a-b)(1+\epsilon_5) = v \quad |\epsilon_5| \leq u \\
fl(zv) & = & zv(1+\epsilon_6) \quad |\epsilon_6| \leq u \\
fl(a^2-b^2) & = & (a+b)(1+\epsilon_4)(a-b)(1+\epsilon_5)(1+\epsilon_6)
\end{array}$$

- **ALGORITMO 2.** $\varphi(a,b)=(a+b)(a-b)$

$$\begin{aligned}
fl(a+b) &= (a+b)(1+\epsilon_4)=z \quad |\epsilon_4|\le u\\
fl(a-b) &= (a-b)(1+\epsilon_5)=v \quad |\epsilon_5|\le u\\
fl(zv) &= zv(1+\epsilon_6) \quad |\epsilon_6|\le u\\
fl(a^2-b^2) &= (a+b)(1+\epsilon_4)(a-b)(1+\epsilon_5)(1+\epsilon_6)\\
&= (a+b)(a-b)(1+\epsilon_4+\epsilon_5+\epsilon_4\epsilon_5+\epsilon_6+\epsilon_4\epsilon_6+\epsilon_5\epsilon_6++\epsilon_4\epsilon_5\epsilon_6
\end{aligned}$$

- **ALGORITMO 2.** $\varphi(a,b) = (a+b)(a-b)$

$$\begin{aligned}
fl(a+b) &= (a+b)(1+\epsilon_4) = z \quad |\epsilon_4| \le u \\
fl(a-b) &= (a-b)(1+\epsilon_5) = v \quad |\epsilon_5| \le u \\
fl(zv) &= zv(1+\epsilon_6) \quad |\epsilon_6| \le u \\
fl(a^2-b^2) &= (a+b)(1+\epsilon_4)(a-b)(1+\epsilon_5)(1+\epsilon_6) \\
&= (a+b)(a-b)(1+\epsilon_4+\epsilon_5+\epsilon_4\epsilon_5+\epsilon_6+\epsilon_4\epsilon_6+\epsilon_5\epsilon_6++\epsilon_4\epsilon_5\epsilon_6 \\
&\simeq (a^2-b^2)(1+\epsilon_4+\epsilon_5+\epsilon_6)
\end{aligned}$$

$$\epsilon_{alg2} = \frac{fl(a^2-b^2)-(a^2-b^2)}{a^2-b^2} \simeq \epsilon_4+\epsilon_5+\epsilon_6$$

$$I_{alg2} = 1+1+1 = 3$$

Si analizza per quali valori di a e b l'Algoritmo 2 è numericamente più stabile dell'Algoritmo 1.

$$I_{alg1} \geq I_{alg2}$$
$$\frac{a^2}{|a^2 - b^2|} + \frac{b^2}{|a^2 - b^2|} + 1 \geq 3$$
$$\Downarrow$$
$$|a^2 - b^2| \leq \frac{1}{2}(a^2 + b^2)$$

Ciò si verifica se:

1. $a^2 - b^2 \leq \frac{a^2}{2} + \frac{b^2}{2} \Rightarrow \frac{a^2}{b^2} \leq 3$
2. $-(a^2 + b^2) \leq 2(a^2 - b^2) \Rightarrow \frac{1}{3} \leq \frac{a^2}{b^2}$

Dunque Algoritmo 2 è numericamente più stabile di algoritmo 1 se

$$\frac{1}{3} \leq \frac{a^2}{b^2} \leq 3$$

Somma di n numeri finiti

Siano $x_1, x_2, \ldots, x_n \in F$,

$$\varphi(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n = S$$

$$\boxed{\begin{array}{l} s \leftarrow x_1; \\ \textbf{for} \quad i = 2, \ldots, n \\ \qquad \lfloor\ s \leftarrow s + x_i; \end{array}}$$

$$s = fl(x_1 + x_2 + \ldots + x_n)$$

$$
\begin{aligned}
fl(S) &= (...(((x_1 + x_2)(1+\epsilon_2) + x_3)(1+\epsilon_3) + x_4)(1+\epsilon_4)+\\
&\quad ... + x_n)(1+\epsilon_n) =\\
&= x_1(1+\epsilon_2)(1+\epsilon_3)...(1+\epsilon_n) + x_2(1+\epsilon_2)...(1+\epsilon_n)+\\
&\quad +x_3(1+\epsilon_3)...(1+\epsilon_n) + ... + x_n(1+\epsilon_n)\\
&\simeq x_1(1+\epsilon_2+\epsilon_3+...+\epsilon_n) + x_2(1+\epsilon_2+...+\epsilon_n)+\\
&\quad +x_3(1+\epsilon_3+...+\epsilon_n) + ... + x_n(1+\epsilon_n) =\\
&= x_1 + x_2 + ... + x_n + (x_1 + x_2)\epsilon_2+\\
&\quad +(x_1 + x_2 + x_3)\epsilon_3 + ... + (x_1 + x_2 + x_{n-1})\epsilon_{n-1}+\\
&\quad +(x_1 + x_2 + ... + x_n)\epsilon_n
\end{aligned}
$$

$$
\begin{aligned}
\frac{fl(S)-S}{S} &\simeq \frac{x_1+x_2}{x_1+...+x_n}\epsilon_2 + \frac{x_1+x_2+x_3}{x_1+...+x_n}\epsilon_3+\\
&\quad +... + \frac{x_1+...+x_{n-1}}{x_1+...+x_n}\epsilon_{n-1} + \epsilon_n
\end{aligned}
$$

$$
I_{alg} = |\frac{x_1 + x_2}{S}| + |\frac{x_1 + x_2 + x_3}{S}| + ... + |\frac{x_1 + ... + x_{n-1}}{S}| + 1
$$

Supponiamo $|\epsilon_i| \leq u \quad i = 2, ..., n$.

$$I_{alg} = |\frac{x_1 + x_2}{S}| + |\frac{x_1 + x_2 + x_3}{S}| + ... + |\frac{x_1 + ... + x_{n-1}}{S}| + 1$$

- **Se gli x_i $i = 1, ..., n$ sono di segno concorde, $|\epsilon_{alg}| \leq (n-1)u$.**
  Tuttavia l'algoritmo è più stabile numericamente se si sommano i numeri dal più piccolo al più grande.
- **Se gli x_i sono di segno discorde, S può essere piccolo e si ha amplificazione degli errori.**
  Se occorre avere la somma in semplice precisione, è conveniente fare la somma di tutti i positivi e poi di tutti i negativi separatamente in semplice precisione e poi sommare le due somme parziali in doppia precisione.

ESEMPIO. Sia $\beta = 10, t = 7$, arrotondamento. Sia $x_1 = 1, x_i = .1\ 10^{-6}\ i = 2, ..., 10.$

$$\sum_{i=1}^{10} x_i = 1 + 9\ 10^{-7} = 1.0000009$$

$$\boxed{\begin{array}{l} s \leftarrow x_1; \\ \textbf{for} \quad i = 2, ..., 10 \\ \qquad\quad s \leftarrow s + x_i; \end{array}}$$

$$\begin{array}{lll} i = 2 & s = fl(x_1 + x_2) & = & fl(1 + .1\ 10^{-6}) = fl(1.0000001) = .1\ 10^1 = y \\ i = 3 & s = fl(y_1 + x_3) & = & fl(1 + .1\ 10^{-6}) = fl(1.0000001) = .1\ 10^1 = y \\ & & \vdots & \\ i = 10 & s = fl(y_{n-1} + x_n) & = & fl(1 + .1\ 10^{-6}) = fl(1.0000001) = .1\ 10^1 \end{array}$$

$$\epsilon_r = 8.999..\ 10^{-7} \sim 10^{-6}$$

$$\boxed{\begin{array}{l} s \leftarrow x_{10}; \\ \textbf{for} \quad i = 9, \ldots, 1 \ : -1 \\ \quad \lfloor \ s \leftarrow s + x_i; \end{array}}$$

$$\begin{array}{lllll}
i = 9 & s = fl(x_{10} + x_9) & = & fl(.1\ 10^{-6} + .1\ 10^{-6}) = fl(0.0000002) = .2\ 10^{-6} \\
i = 8 & s = fl(z_1 + x_3) & = & fl(.2\ 10^{-6} + .1\ 10^{-6}) = fl(0.0000003) = .3\ 10^{-6} \\
 & & \vdots & \\
i = 2 & s = fl(z_7 + x_2) & = & fl(.8\ 10^{-6} + .1\ 10^{-6}) = fl(0.0000009) = .9\ 10^{-6} \\
i = 1 & s = fl(z_8 + x_1) & = & fl(.9\ 10^{-6} + 1) = fl(1.0000009) = .100001\ 10^{1}
\end{array}$$

$$\epsilon_r = 9.999..\ 10^{-8} \sim 10^{-7}$$

L’errore è 10 volte più piccolo!

ANALISI DELL'ERRORE SUI DATI INIZIALI

ESEMPIO.
Problema: trovare y tale che

$$y^2 - 4y + x = 0$$

Possiamo esprimere le soluzioni del problema mediante la formula

$$y_{1,2} = 2 \pm \sqrt{4 - x}$$

Per $x = 4$ abbiamo $y_{1,2} = 2$.
Per $x = 4 - 10^{-6}$ abbiamo $y_{1,2} = 2 \pm 10^{-3}$.
Piccole variazioni nei dati iniziali (10^{-6}) comportano grosse variazioni nei risultati finali (10^{-3}).

Questo è un esempio di problema MAL CONDIZIONATO.

Vediamo come possiamo stimare l'errore relativo dovuto all'approssimazione iniziale dei dati.

$$\epsilon_{dati} = \frac{\varphi(\bar{x}) - \varphi(x)}{\varphi(x)}$$

Possiamo scrivere

$$\bar{x} = x + \delta \Longrightarrow \delta = \bar{x} - x \Longrightarrow \frac{\delta}{x} = \frac{\bar{x} - x}{x} = \epsilon_x$$

$$\begin{aligned} \epsilon_{dati} &= \frac{\varphi(x+\delta)-\varphi(x)}{\varphi(x)} \\ &\simeq \frac{\varphi(x)+\varphi'(x)\delta-\varphi(x)}{\varphi(x)} \\ &= \frac{\varphi'(x)\delta}{\varphi(x)} \\ &= \frac{\varphi'(x)x}{\varphi(x)}\epsilon_x \end{aligned}$$

$\frac{\varphi'(x)x}{\varphi(x)}$ **è il fattore di amplificazione dell'errore sui dati iniziali e viene detto INDICE DI CONDIZIONAMENTO.**

$$I_{cond} = K(x,\varphi) = \left|\frac{\varphi'(x)}{\varphi(x)}x\right|$$

Se $K(x,\varphi) >> 1$, il problema è mal condizionato; se $K(x,\varphi)$ è piccolo, il problema è ben condizionato.

ESEMPIO di prima...

$$y^2 - 4y + x = 0$$

$$y_1 = 2 + \sqrt{4-x}$$

$$\varphi(x) = 2 + \sqrt{4-x}$$

$$\varphi'(x) = -\frac{1}{2\sqrt{4-x}} \longrightarrow^{x\to 4} \infty$$

ESEMPIO. $y = \varphi(x) = \log(x)$.

$$\epsilon_{dati} = \frac{1}{\log(x)}\epsilon_x$$

$$I_{cond} = |\frac{1}{\log(x)}|$$

Se $x \simeq 1$, il problema è mal condizionato.

OSSERVAZIONE Un problema può essere ben condizionato per certi valori e mal condizionato per altri.

Consideriamo il caso generale:

$$\boldsymbol{y = \varphi(x_1, ..., x_n)}$$

Se $\boldsymbol{x_i}$, $\boldsymbol{i = 1, ..., n}$ è affetto da un errore $\boldsymbol{\Delta x_i}$, con $\boldsymbol{\epsilon_{x_i} = \frac{\Delta x_i}{x_i}}$, allora si calcola:

$$\boldsymbol{y + \Delta y = \varphi(x_1 + \Delta x_1, ..., x_n + \Delta x_n)}$$

Pertanto se φ è sufficientemente regolare,

$$\begin{aligned} E_{dati} &= \Delta y = \varphi(x_1 + \Delta x_1, ..., x_n + \Delta x_n) - \varphi(x_1, ..., x_n) \\ &\simeq \sum_{i=1}^{n} \frac{\partial \varphi(x_1, ..., x_n)}{\partial x_i} \Delta x_i \end{aligned}$$

Inoltre se $\varphi(x_1, ..., x_n) \neq 0$ **e** $x_i \neq 0$**,**

$$\begin{aligned}
\epsilon_{dati} \quad &= \quad \frac{\Delta y}{y} = \frac{\varphi(x_1 + \Delta x_1, ..., x_n + \Delta x_n) - \varphi(x_1, ..., x_n)}{\varphi(x_1, ..., x_n)} \\
&\simeq \quad \sum_{i=1}^{n} \frac{\partial(\varphi(x_1, ..., x_n))}{\partial x_i} \frac{x_i}{\varphi(x_1, ..., x_n)} \epsilon_{x_i}
\end{aligned}$$

In tal caso si dice indice di condizionamento del problema la somma dei valori assoluti dei coefficienti dei singoli errori sui dati iniziali:

$$\boldsymbol{I_{cond} = \sum_{i=1}^{n} K(x_i, \varphi) = \sum_{i=1}^{n} \left| \frac{\partial \varphi(x_1, ..., x_n)}{\partial x_i} \frac{x_i}{\varphi(x_1, ..., x_n)} \right|}$$

Caso delle operazioni elementari

$$
\begin{aligned}
\epsilon_{x\pm y} &= \frac{x}{x\pm y}\epsilon_x \pm \frac{y}{x\pm y}\epsilon_y \\
\epsilon_{xy} &= \epsilon_x + \epsilon_y \\
\epsilon_{x/y} &= \epsilon_x - \epsilon_y \\
\epsilon_{\sqrt{(x)}} &= \frac{1}{2}\epsilon_x \\
\epsilon_{x^\alpha} &= \alpha\epsilon_x
\end{aligned}
$$

Moltiplicazione, divisione, estrazione di radice e potenza, con $|\alpha|$ piccolo, sono ben condizionate. Per l'operazione $x - y$, con $x \simeq y$, si verifica il fenomeno di cancellazione.

ESERCIZI.

1. $\varphi(x) = x - 1$; $I_{cond} = |\frac{x}{x-1}|$.
   Mal condizionamento per $x \simeq 1$.

2. $\varphi(x) = e^x$; $I_{cond} = |x|$.
   Mal condizionamento per $|x| \gg 1$ e buon condizionamento per $|x| \ll 1$.
3. $\sqrt{(x^2+1)} - |x|$; $I_{cond} = |\frac{x}{\sqrt{(x^2+1)}}|$.
   Sempre buon condizionamento.
4. $y = \varphi(a, b, c) = a + b + c$.

$$\epsilon_y = \frac{a}{a+b+c}\epsilon_a + \frac{b}{a+b+c}\epsilon_b + \frac{c}{a+b+c}\epsilon_c$$

$$I_{cond} = |\frac{a}{a+b+c}| + |\frac{b}{a+b+c}| + |\frac{c}{a+b+c}|$$

   Il problema è ben condizionato se a, b, c sono di segno concorde.
5. $y = \varphi(a, b) = a^2 - b^2$.

$$\epsilon_y = \frac{2a^2}{a^2 - b^2}\epsilon_a - \frac{2b^2}{a^2 - b^2}\epsilon_b$$

$$I_{cond} = \frac{2a^2}{|a^2 - b^2|} + \frac{2b^2}{|a^2 - b^2|}$$

   Il problema è mal condizionato se $a^2 \simeq b^2$.

Calcolo di e^x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Se $x < 0$ si sommano e sottraggono termini con ordini di grandezza differenti. Si perdono le cifre che influiscono sul risultato.

ALTERNATIVA

$$e^{-5.5} = \frac{1}{e^{5.5}} \simeq .0040865 \qquad \textbf{errore dello 0.007\%}$$

Per il calcolo di e^x, con $x = [x] + f$, conviene

1. $e^x = e^{[x]}e^f = (e.e....e)(1 + f + f^2/2! + \ldots)$
2. $e^x = (e^{1+\frac{f}{[x]}})^{[x]} = (\sum_{i=0}^{\infty}(1 + \frac{f}{[x]})^i/i!)^{[x]}$, $1 \leq 1 + f/[x] < 2$

Approssimazione della derivata prima

$f \in C^2$

$$f(a+h) = f(a) + f'(a)h + f''(a+th)\frac{h^2}{2}$$

con $t \in (0,1)$.

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

$$E_t = f'(a) - \frac{f(a+h) - f(a)}{h} = f''(a+th)\frac{h}{2}$$

$$|f''(x)| \leq M \text{ per } x \in [a, a+h]$$

$$|E_t| \leq \frac{Mh}{2}$$

$$|fl(f(x)) - f(x)| < tol$$

$$
\begin{aligned}
\left| f'(a) - \tfrac{fl((f(a+h)) - fl(f(a))}{h} \right| &= \left| f'(a) - \tfrac{f(a+h)-f(a)}{h} + \right. \\
&\quad \left. + \tfrac{f(a+h)-f(a)}{h} + - \tfrac{fl((f(a+h)) - fl(f(a))}{h} \right| \\
&\leq \left| f'(a) - \tfrac{f(a+h)-f(a)}{h} \right| + \\
&\quad + \left| \tfrac{f(a+h)-f(a)}{h} - \tfrac{fl((f(a+h)) - fl(f(a))}{h} \right| \\
&= |E_t| + \left| \tfrac{f(a+h) - fl(f(a+h))}{h} + \tfrac{f(a) - fl(f(a))}{h} \right| \\
&\leq \tfrac{Mh}{2} + 2\tfrac{tol}{h}
\end{aligned}
$$

$$\psi(z) = \frac{Mz}{2} + 2\frac{tol}{z}$$

$$\psi'(z) = \frac{M}{2} - \frac{2tol}{z^2}$$

$$\psi'(z) = 0 \Longleftarrow z = 2\sqrt{\frac{tol}{M}}$$