# Esercizi di Calcolo Scientifico

## Indice

# 1 Prima esercitazione

## 1.1 Esercizio 1

Testo dell'esercizio e primi 3 punti.

**d)** Dato $g = \frac{7}{5}$, verifica $y \neq \mathcal{F}$ e determina $\hat{y} = fl(y) \in \mathcal{F}$.

$$y = \frac{7}{5} = 1.4$$

$$\begin{array}{r|c|c|l}
\text{moltiplicato per} & 1.4 & 1 & \\
\text{la base (2)} & 0.8 & 0 & \\
 & 1.6 & 1 & \text{periodo} \\
 & 1.2 & 1 & \\
 & 0.4 & 0 & \\
 & 0.8 & &
\end{array}$$

$$y = (1.\overline{0110})_2$$

Ha rappresentazione infinita, quindi $\notin \mathcal{F}$.

$$\begin{aligned} y &= 2^1 \cdot (0.1\overline{011\underline{0}})_2 \quad (t+1)\text{esima cifra} \\ f(y) &= 2^1 \cdot (0.1011)_2 \end{aligned}$$

**e)** Calcola $z = \tilde{x} + 2\tilde{y}$ e $\tilde{z} = \tilde{x} fl(t) 2\tilde{y}$.

$$\tilde{x} = 2^{-2} \cdot (0.1101)_2, \qquad \tilde{y} = 2^1 \cdot (0.1011)_2$$

$$\begin{aligned} \tilde{x} + 2\tilde{y} &= 2^{-2} \cdot (0.1101)_2 + 2^2 \cdot (0.1011)_2 \\ &= 2^2 \cdot (0.00001101)_2 + 2^2 \cdot (0.1011)_2 \\ &= 2^2 \cdot (0.1011\underline{1}101)_2 \quad \text{numero in aritmetica esatta} \end{aligned}$$

Sappiamo che $\tilde{z} = fl(z)$

$$0.1011 + 0.0001 = 0.1100$$

Quindi $\tilde{z} = fl(z) = 2^2 \cdot (0.1100)_2$

### Analisi degli errori

**Errore inerente** Relativo al condizionamento. Quando l'errore è sensibile al problema che abbiamo sui dati, quanto il problema amplifica gli errori in input.

**Errore algoritmico** Relativo alla stabilità. Differenza tra il risultato in aritmetica esatta e in aritmetica di macchina. Un algoritmo è stabile quando l'errore algoritmico è molto minore di quello della macchina.

## 1.2 Esercizio 2

Si vuole calcolare $y = f(x)$.

**a)** Definsci l'errore inerente e il condizionamento.

$$\text{condizionamento} = \frac{|\text{err}_{relativo}(\text{output})|}{|\text{err}_{relativo}(\text{intput})|}$$

**b)** Studia il condizionamento di $f(x) = \dfrac{2+x^2}{x^2-2}$.

Utilizzando Taylor si può scrivere

$$\text{cond}_f(x) = \frac{|x| \cdot |f'(x)|}{f|x|}$$

$$f'(x) = \frac{2x(x^2-2) - 2x(2+x^2)}{(x^2-2)^2} = \frac{-8x}{(x^2-2)^2}$$

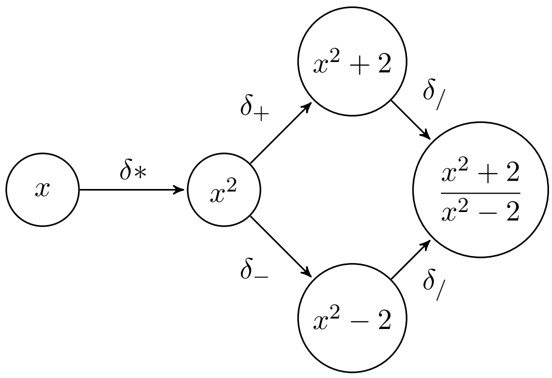
$$\begin{aligned}\text{cond}_f(x) &= \frac{|x| \cdot \left|\dfrac{8x}{(x^2-2)^2}\right|}{\left|\dfrac{2+x^2}{x^2-2}\right|} \\ &= \frac{|x| \cdot |8x|}{(x^2-2)^2} \cdot \frac{|x^2-2|}{2+x^2} \\ &= \frac{8x^2}{(x^2-2)^2|x^2-2|}\end{aligned}$$

A questo punto bisogna dire se è ben condizionato o mal condizionato. Il problema è mal condizionato se $x^2 \simeq 2$, ovvero $x \simeq \sqrt{2}$ (e quindi tende a $+\infty$), ben condizionato altrimenti.

**c)** Definsci l'errore algoritmico e il concetto di stabilità.

**d)** Studia la stabilità dell'algoritmo che valuta $f(x)$ nel punto $x$ (ovvero per quali valori di $x$ l'algoritmo è stabile o meno).

Si intendono le seguenti operazioni elementari:




Il delta di un'operazione generica ($\delta_{op}$) si calcola

$$\delta_{op} = \frac{|y(op)z - y(fl(op))z|}{|y(op)z|}$$

Ipotesi $|\delta_{op}| \leq u$. Si fa il primo addendo fratto il risultato dell'addizione:

$$c(x \to x^2) = \frac{x \cdot 2x}{x^2} = 2$$

$$c(y \to y+2) = \frac{y \cdot 1}{y+2} = \frac{y}{y+2}$$

$$c\Big(y \to \frac{y}{z}\Big) = \frac{y \cdot \frac{1}{z}}{\frac{y}{z}} = \frac{\frac{y}{z}}{\frac{y}{z}} = 1$$

$$c\Big(z \to \frac{y}{z}\Big) = \frac{z \cdot \Big(-\frac{y}{z^2}\Big)}{\frac{y}{z}} = \frac{-y \cdot z}{z^2} = \frac{z}{y} = -1$$

$$\begin{tikzcd}
 & & x^2+2 \arrow[dr, "1"] & \\
x \arrow[r, "2"] & x^2 \arrow[ur, "\frac{x^2}{x^2+2}"] \arrow[dr, "\frac{x^2}{x^2-2}"'] & & \frac{x^2+2}{x^2-2} \\
 & & x^2-2 \arrow[ur, "-1"'] &
\end{tikzcd}$$

$$err_{alg} = \delta_/ + 1\Big(\delta_+ + \frac{x^2}{x^2+2} \cdot (\delta_* + 2\delta)\Big)$$

$$\delta_/ - 1\Big(\delta_- + \frac{x^2}{x^2-2} \cdot (\delta_* + 2\delta)\Big)$$

$$\Rightarrow |err_{alg}| = |\delta_/| + |\delta_+| + \frac{x^2}{x^2+2} \cdot (|\delta_*| + 2|\delta|)$$

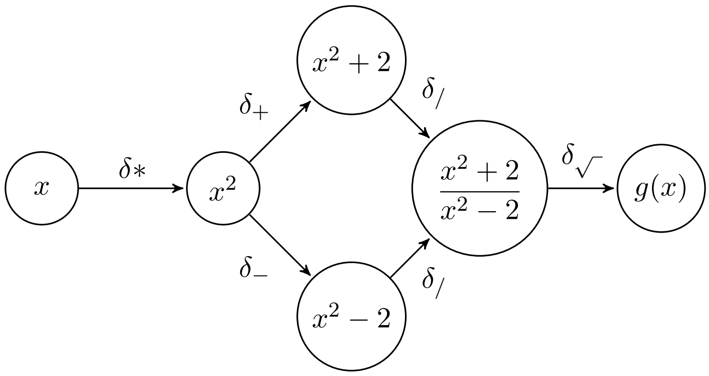
$$|\delta_/| + |\delta_-| + \frac{x^2}{x^2-2} \cdot (|\delta_*| + 2|\delta|)$$

$$|err_{alg}| \leq u + u + u + 2u + u + \frac{x^2}{|x^2-2|} \cdot 3u = 6u + \frac{x^2}{|x^2-2|} \cdot 3u$$

Vogliamo evitare che il denominatore ($|x^2 - 2|$) si annulli. Quindi l'algoritmo è instabile se $x^2 \simeq 2$, cioè $x \simeq \sqrt{2}$ o $x \simeq -\sqrt{2}$, stabile altrimenti.

**e)** Studia il condizionamento di $g(x) = \sqrt{f(x)}$ e la stabilità dell'algoritmo che valuta $g(x)$ in $x$.

Sappiamo che $g(x) = \sqrt{\dfrac{2+x^2}{x^2-2}}$.




$$cond_g(x) = \frac{|x| \cdot |g'(x)|}{|g(x)|}$$

$$g'(x) = \frac{f'(x)}{2g(x)} = \frac{f'(x)}{2\sqrt{f(x)}}$$

$$\Rightarrow cond_g(x) = \frac{|x| \cdot \dfrac{|f'(x)|}{2\sqrt{f(x)}}}{\sqrt{f(x)}} = \frac{|x| \cdot |f'(x)|}{2 \cdot |f(x)|} = \frac{1}{2}$$

Se questo è piccolo, anche $g$ lo è, quindi $g$ è condizionato anche quando questo lo è. La funzione $g(x)$ è ben condizionata quando lo è $f(x)$.

$$c(y \to \sqrt{y}) = \frac{y \cdot \dfrac{1}{2\sqrt{y}}}{\sqrt{y}} = \frac{y}{2|y|}$$

$$|c(y \to \sqrt{y})| = \frac{1}{2} \cdot \frac{|y|}{|y|} = \frac{1}{2}$$

$$err_{alg}(g) = \delta_{\sqrt{\ }} + \frac{f(x)}{2|f(x)|} \cdot err_{alg}(f)$$

$$\Rightarrow |err_{alg}(g)| \leq u + \frac{1}{2}|err_{alg}(f)|$$

L'algoritmo di calcolo $g$ è stabile quando lo è quello che calcola $f$, cioè quando $x^2 \not\approx 2$.

### 1.3 Esercizio 3

Sia $\mathcal{F} = \mathcal{F}\ (2,\, t,\, p_{min},\, p_{max})$ con arrotondamento.

**a)** Determina i parametri sapendo che $p_{max} = t$, $\mathcal{F}$ contiene 64 elementi positivi, e $real_{max} = 15$.

$$\mathcal{F} = \{0\} \cup \{\pm(d_1 B^{-1} + \ldots + d_t B^{-t}) B^p \;\left|\; \begin{array}{l} 1 \leq d_1 \leq B-1 \\ 0 \leq d_2, \ldots, d_t \leq B-1 \\ -p_{min} \leq p \leq p_{max} \end{array}\right.$$

$$(B-1)B^{t-1}(p_{min} + p_{max} + 1) = 64 \Rightarrow 2^{t-1}(p_{min} + t + 1) = 64$$

$$\begin{aligned} real_{max} &= B^{p_{max}}(B-1)(B^{-1}+\ldots+B^{-t}) \\ &= B^{p_{max}}(1-B^{-t}) \\ &= 2^t(1-2^{-t}) \\ &= 2^t-1 \\ &\Rightarrow 2^t = 16 \\ &\Rightarrow t=4,\ p_{max}=t=4 \end{aligned}$$

$$\begin{aligned} 2^3(p_{min}+5) &= 64 \\ p_{min}+5 &= \frac{64}{8} = 8 \\ p_{min} &= 3 \end{aligned}$$

**b)** Dopo aver definito la precisione di macchina, determina quella di $\mathcal{F}$.

Nel caso di arrotondamento, $u = \dfrac{B^{1-t}}{2}$. Nel caso di $\mathcal{F}$, $u = \dfrac{2^{1-4}}{2} = \dfrac{2^{-3}}{2} = 2^{-4}$.

**c)** Dato $x = \dfrac{1}{5}$, verifica $x \neq \mathcal{F}$ e trova $fl(x) \in \mathcal{F}$.

$$\tilde{x} = fl(x) = 2^{-2}(0.1101)_2$$

Come prima, perché abbiamo lo stesso numero di cifre della mantissa, abbiamo la stessa rappresentazione.

**d)** Dato $y = \dfrac{1}{3}$, verifica $y \notin \mathcal{F}$ e determina $fl(y)$.

Abbiamo $y = 0.\overline{3}$

$$\begin{array}{r|l|} 0.\overline{3} & 0 \\ 0.\overline{6} & 0 \\ 1.\overline{3} & 1 \\ 0.\overline{6} & 0 \end{array}$$

$$\begin{aligned} \Rightarrow y &= (0.\overline{01})_2 \notin \mathcal{F} \\ &= (0.010101\ldots)_2 \end{aligned}$$

Scrivere con 0 prima della virgola ma con 1 come prima cifra decimale $\Rightarrow$ spostare tutto di una cifra a sinistra $\Rightarrow$

$$\begin{aligned} y &= 2^{-1}(0.101010\ldots)_2 \\ &= 2^{-1}(0.\overline{10})_2 \end{aligned}$$

$$fl(y) = 2^{-1}(0.1011)_2$$

**e)** Calcola $x + y = z$ e determina $\tilde{z} = fl(z)$.

$$z = x + y = \frac{1}{3} + \frac{1}{5} = \frac{8}{15}$$

$$\begin{array}{r|l}
{}^{8}\!/_{15} & 0 \\
{}^{16}\!/_{15} & 1 \\
{}^{2}\!/_{15} & 0 \\
{}^{4}\!/_{15} & 0 \\
{}^{8}\!/_{15} & 0 \\
{}^{16}\!/_{15} & 1
\end{array}$$

$$\begin{aligned}
z &= (0.\overline{1000})_2 \\
&= (0.10001000\ldots)_2 \\
\tilde{z} &= (0.1001)_2
\end{aligned}$$

**e)** Calcola $\tilde{\omega} = \tilde{x} fl(\bigoplus)\tilde{y}$ e trova la relazione tra $\tilde{\omega}$ e $\tilde{z}$.

**f)** Definisci i numeri denormalizzati. Quali sono quelli per $\mathcal{F}$?

I numeri denormalizati sono numeri che non fanno parte dell'insieme originario.

$$\mathcal{G} = \{\pm(d_2 \cdot 2^{-2} + d_3 \cdot 2^{-3} + d_4 \cdot 2^{-4}) \cdot 2^{-3} \mid d_2, d_3, d_4 \in \{0, 1\}\}$$

# 2 Esercitazione 19 Marzo 2019

## 2.1 Esame 13 Feb 2017

**Esercizio 1)** $B = 2$, numero di elementi positivi $= 64$.

$$\begin{aligned} 64 &= (B-1)B^{t-1}(1 + p_{min} + p_{max}) \\ 64 &= 2^{t-1}(1 + p_{min} + p_{max}) \end{aligned}$$