

Esercizio. Scrivere i numeri di macchina positivi dell'insieme di numeri di macchina definito dai parametri:

$$B = 3, m = -2, M = 2, t = 1.$$

I numeri di macchina positivi sono

$$\begin{aligned} &1.0 \cdot 3^{-2}, 1.1 \cdot 3^{-2}, 1.2 \cdot 3^{-2}, 2.0 \cdot 3^{-2}, 2.1 \cdot 3^{-2}, 2.2 \cdot 3^{-2}, p = -2, \\ &1.0 \cdot 3^{-1}, 1.1 \cdot 3^{-1}, 1.2 \cdot 3^{-1}, 2.0 \cdot 3^{-1}, 2.1 \cdot 3^{-1}, 2.2 \cdot 3^{-1}, p = -1, \\ &1.0 \cdot 3^0, 1.1 \cdot 3^0, 1.2 \cdot 3^0, 2.0 \cdot 3^0, 2.1 \cdot 3^0, 2.2 \cdot 3^0, p = 0, \\ &1.0 \cdot 3^1, 1.1 \cdot 3^1, 1.2 \cdot 3^1, 2.0 \cdot 3^1, 2.1 \cdot 3^1, 2.2 \cdot 3^1, p = 1, \\ &1.0 \cdot 3^2, 1.1 \cdot 3^2, 1.2 \cdot 3^2, 2.0 \cdot 3^2, 2.1 \cdot 3^2, 2.2 \cdot 3^2, p = 2. \end{aligned}$$

Esercizio. Scrivere la parola di memoria di 64 bit corrispondente ai numeri 45 e -2^{-100} nello standard IEEE in doppia precisione.

Si ha

$$\begin{aligned} 45 &= 32 + 13 = (101101)_2 = (1.01101)_2 \cdot 2^5 \\ q &= 1023 + p = 1028 = (10000000100)_2 \end{aligned}$$

e quindi la parola di memoria è

$$\underbrace{0}_{\text{segno: 1 bit}} \quad \underbrace{10000000100}_{\text{esponente: 11 bit}} \quad \underbrace{011010 \dots 0}_{\text{mantissa: 52 bit}}.$$

Si ha

$$\begin{aligned} -2^{-100} &= -(1.0 \dots 0)_2 \cdot 2^{-100} \\ q &= 1023 + p = 923 = 512 + 256 + 128 + 16 + 8 + 2 + 1 = (01110011011)_2 \end{aligned}$$

e quindi la parola di memoria è

$$\underbrace{1}_{\text{segno: 1 bit}} \quad \underbrace{01110011011}_{\text{esponente: 11 bit}} \quad \underbrace{0 \dots 0}_{\text{mantissa: 52 bit}}.$$

Esercizio. Se tutti i numeri dello Standard IEEE in doppia precisione fossero memorizzati in qualche supporto, quale sarebbe l'occupazione complessiva di memoria?

Il numero di numeri di macchina nello standard IEEE in doppia precisione è

$$\begin{aligned} 2(B-1)B^t(M-m+1)+1 &= 2 \cdot 1 \cdot 2^{52} \cdot (1023 - (-1022) + 1) + 1 \\ &= 2^{53} \cdot 2046 + 1. \end{aligned}$$

Ciascuno di questi occupa una parola di memoria di 64 bit. Quindi l'occupazione complessiva di memoria è

$$(2^{53} \cdot 2046 + 1) \cdot 64 = 1.1794 \cdot 10^{21} \text{ bit} = 1.4743 \cdot 10^8 \text{ Terabyte.}$$

Esercizio. Quanti sono i numeri macchina nell'intervallo $[B^p, B^{p+1})$, dove $p \in \{m, m+1, \dots, M\}$ è fissato? Esprimere tale numero in termini di ϵ e B .

I possibili numeri di macchina

$$x = (b_0.b_{-1}b_{-2}b_{-3} \cdots b_{-t})_B \cdot B^p,$$

nell'intervallo $[B^p, B^{p+1})$ sono

$$\underbrace{(B-1)}_{\text{scelte di } b_0 \neq 0} \cdot \underbrace{B}_{\text{scelte di } b_{-1}} \cdot \cdots \cdot \underbrace{B}_{\text{scelte di } b_{-t}} = (B-1) B^t = \frac{B-1}{\epsilon}.$$

Esercizio. Si consideri l'insieme di numeri di macchina definito dai parametri

$$B = 16, \quad m = -3, \quad M = 3, \quad t = 2.$$

Trovare $\text{fl}(\frac{1}{13})$ e $\text{fl}(15005)$ sia nel caso del troncamento che in quello dell'arrotondamento.

Determiniamo la rappresentazione di $\frac{1}{13}$ in base 16 utilizzando le cifre A, B, C, D, E, F per 10, 11, 12, 13, 14, 15. Si ha

	$\cdot B$	parte intera	parte frazionaria
$\frac{1}{13}$	$\frac{16}{13}$	1	$\frac{3}{13}$
$\frac{3}{13}$	$\frac{48}{13}$	3	$\frac{9}{13}$
$\frac{9}{13}$	$\frac{144}{13}$	11 = B	$\frac{1}{13}$
$\frac{1}{13}$	.	.	.

Quindi

$$\frac{1}{13} = (0.\overline{13B})_{16} = (1.\overline{3B1})_{16} \cdot 16^{-1}$$

e risulta

$$\text{fl}\left(\frac{1}{13}\right) = (1.3B)_{16} \cdot 16^{-1}$$

sia per il troncamento che per l'arrotondamento.

Determiniamo la rappresentazione di 15005 in base 16. Si ha

	quoziente divisione per B	resto divisione per B
15005	937	13 = D
937	58	9
58	3	10 = A
3	0	3

Quindi

$$15005 = (3A9D)_{16} = (3.A9D)_{16} \cdot 16^3$$

e risulta

$$\text{fl}(15005) = (3.A9)_{16} \cdot 16^3$$

per il troncamento e

$$\text{fl}(15005) = (3.AA)_{16} \cdot 16^3$$

per l'arrotondamento.

Esercizio. Si consideri l'insieme di numeri di macchina definito dai parametri

$$B = 5, \quad m = -2, \quad M = 2, \quad t = 2.$$

Nel caso dell'arrotondamento, trovare $\text{fl}(x)$ per $x = (441.301)_5$. Fare lo stesso ora per $B = 3$ e $x = (11.1b1)_3$ con $b = 0, 1, 2$.

Riguardo $B = 5$, risulta

$$x = (441.301)_5 = (4.41301)_5 \cdot 5^2.$$

Poichè la prima cifra in

$$b_{-t-1}, b_{-t-2}, b_{-t-3}, \dots = 3, 0, 1, \dots$$

diversa da $\frac{B-1}{2} = 2$, vale a dire $b_{-t-1} = 3$, è maggiore di $\frac{B-1}{2} = 2$, si ha

$$\text{fl}(x) = x_2 = (4.42)_5 \cdot 5^2$$

Riguardo $B = 3$, risulta

$$x = (11.1b1)_3 = (1.11b1)_3 \cdot 3.$$

Dobbiamo ora trovare la prima cifra in

$$b_{-t-1}, b_{-t-2}, b_{-t-3}, \dots = b, 1, 0, \dots$$

diversa da $\frac{B-1}{2} = 1$ e vedere se è minore di 1. Se $b = 0$ o $b = 2$, allora tale cifra è $b_{-t-1} = b$ e si ha

$$\begin{aligned}\text{fl}(x) &= x_1 = (1.11)_3 \cdot 3 \text{ se } b = 0 \\ \text{fl}(x) &= x_2 = (1.12)_3 \cdot 3 \text{ se } b = 2.\end{aligned}$$

Se $b = 1$, allora tale cifra è $b_{-t-3} = 0$ e quindi

$$\text{fl}(x) = x_1.$$

Esercizio. Sia nel caso del troncamento che in quello dell'arrotondamento, trovare l'errore relativo ε_r dell'approssimazione $\text{fl}(x)$ di

$$x = B^{p+1} - c\text{eps}B^p,$$

dove $c \in (0, 1)$, e scriverlo nella forma

$$|\varepsilon_r| = C\text{eps} + O(\text{eps}^2)$$

per un'opportuna costante C .

Per un tale x si ha

$$\begin{aligned}x_1 &= B^{p+1} - \text{eps}B^p \\ x_2 &= B^{p+1}.\end{aligned}$$

Nel troncamento, e nell'arrotondamento quando $c > \frac{1}{2}$, si ha

$$\text{fl}(x) = x_1$$

e quindi

$$\begin{aligned}|\varepsilon_r| &= \left| \frac{\text{fl}(x) - x}{x} \right| = \left| \frac{x_1 - x}{x} \right| = \left| \frac{B^{p+1} - \text{eps}B^p - (B^{p+1} - c\text{eps}B^p)}{B^{p+1} - c\text{eps}B^p} \right| \\ &= \left| \frac{-(1-c)\text{eps}B^p}{B^{p+1} - c\text{eps}B^p} \right| = \left| \frac{-(1-c)\text{eps}}{B - c\text{eps}} \right| = \frac{(1-c)\text{eps}}{B - c\text{eps}} \\ &= \frac{(1-c)\text{eps}}{B} \cdot \frac{1}{1 - \frac{c}{B}\text{eps}} = \frac{(1-c)\text{eps}}{B} \cdot \frac{1}{1 + O(\text{eps})} \\ &= \frac{(1-c)\text{eps}}{B} \cdot (1 + O(\text{eps})) \\ &= \frac{1-c}{B}\text{eps} + O(\text{eps}^2).\end{aligned}$$

Invece, nell'arrotondamento quando $c \leq \frac{1}{2}$, si ha

$$\text{fl}(x) = x_2$$

e quindi

$$\begin{aligned}
|\varepsilon_r| &= \left| \frac{\text{fl}(x) - x}{x} \right| = \left| \frac{x_2 - x}{x} \right| = \left| \frac{B^{p+1} - (B^{p+1} - \text{ceps}B^p)}{B^{p+1} - \text{ceps}B^p} \right| \\
&= \left| \frac{\text{ceps}B^p}{B^{p+1} - \text{ceps}B^p} \right| = \left| \frac{\text{ceps}}{B - \text{ceps}} \right| = \frac{\text{ceps}}{B - \text{ceps}} \\
&= \frac{\text{ceps}}{B} \cdot \frac{1}{1 - \frac{\text{ceps}}{B}} \\
&= \frac{\text{ceps}}{B} \cdot (1 + O(\text{eps})) \\
&= \frac{c}{B} \text{eps} + O(\text{eps}^2).
\end{aligned}$$

Esercizio. Qual è l'errore relativo di $\text{fl}(x)$ in caso di overflow nello standard IEEE?

Nell'overflow $\text{fl}(x) = +\infty$ e quindi

$$\varepsilon_r = \frac{\text{fl}(x) - x}{x} = \frac{+\infty - x}{x} = +\infty.$$

Esercizio. Si consideri l'insieme di numeri di macchina

$$B = 2, \quad m = -10, \quad M = 10, \quad t = 3.$$

Calcolare, con un tale insieme di numeri di macchina e con l'arrotondamento, $a \circ b$, per $\circ = +, -, \cdot, /$, nel caso $a = 12$ e $b = 15$ come pure nel caso $a = 25$ e $b = 31$. Ricordare che nel caso in cui a e b non siano numeri di macchina, essi vanno approssimati con numeri di macchina prima di effettuare l'operazione $\circ$.

Consideriamo $a = 12$ e $b = 15$. Si ha

$$\begin{aligned}
a &= 12 = (1100)_2 = (1.100)_2 \cdot 2^3 \\
b &= 15 = (1111)_2 = (1.111)_2 \cdot 2^3.
\end{aligned}$$

Quindi a e b sono numeri di macchina. Si ha

$$\begin{aligned}
a + b &= 12 + 15 = 27 = (11011)_2 = (1.1011)_2 \cdot 2^4 \\
a - b &= 12 - 15 = -3 = -(11)_2 = -(1.1)_2 \cdot 2 \\
a \cdot b &= 12 \cdot 15 = 180 = 128 + 32 + 16 + 4 = (10110100)_2 = (1.0110100)_2 \cdot 2^7 \\
a / b &= 12 / 15 = \frac{4}{5} = 0.8 = (0.\overline{1100})_2 = (1.\overline{1001})_2 \cdot 2^{-1}
\end{aligned}$$

Quindi

$$\begin{aligned} a \uplus b &= (1.110)_2 \cdot 2^4. \\ a \ominus b &= -(1.100)_2 \cdot 2^4 \\ a \oslash b &= (1.011)_2 \cdot 2^7 \\ a \oslash b &= (1.101)_2 \cdot 2^{-1}. \end{aligned}$$

Consideriamo $a = 25$ e $b = 31$. Si ha

$$\begin{aligned} a &= 25 = (11001)_2 = (1.1001)_2 \cdot 2^4 \\ b &= 31 = (11111)_2 = (1.1111)_2 \cdot 2^4. \end{aligned}$$

Quindi a e b non sono numeri di macchina e vanno approssimati con i numeri di macchina

$$\text{fl}(a) = (1.101)_2 \cdot 2^4 = 26 \quad \text{e} \quad \text{fl}(b) = (1.000)_2 \cdot 2^5 = 32.$$

Si ha

$$\begin{aligned} \text{fl}(a) + \text{fl}(b) &= 26 + 32 = 58 = 32 + 16 + 8 + 2 = (111010)_2 = (1.11010)_2 \cdot 2^5 \\ \text{fl}(a) - \text{fl}(b) &= 26 - 32 = -6 = -(110)_2 = -(1.10)_2 \cdot 2^2 \\ \text{fl}(a) \cdot \text{fl}(b) &= 26 \cdot 32 = 832 = 512 + 256 + 64 = (1101000000)_2 = (1.101000000)_2 \cdot 2^9 \\ \text{fl}(a) / \text{fl}(b) &= 26 / 32 = \frac{13}{16} = (0.1101)_2 = (1.101)_2 \cdot 2^{-1} \end{aligned}$$

Quindi

$$\begin{aligned} \text{fl}(a) \uplus \text{fl}(b) &= (1.111)_2 \cdot 2^5. \\ \text{fl}(a) \ominus \text{fl}(b) &= -(1.100)_2 \cdot 2^2 \\ \text{fl}(a) \oslash \text{fl}(b) &= (1.101)_2 \cdot 2^9 \\ \text{fl}(a) \oslash \text{fl}(b) &= (1.101)_2 \cdot 2^{-1}. \end{aligned}$$

Esercizio. Scrivere due algoritmi per il calcolo di

$$f(x) = x_1^2 - x_2^2, \quad x \in \mathbb{R}^2,$$

e tre algoritmi per il calcolo di

$$f(x) = x_1^4 - x_2^4, \quad x \in \mathbb{R}^2.$$

Per la prima funzione, due algoritmi sono basati sullo scrivere $f(x)$ come

$$f(x) = x_1^2 - x_2^2 = (x_1 + x_2)(x_1 - x_2).$$

Questi algoritmi sono

ALGORITMO 1

$$a = x_1 \cdot x_1$$

$$b = x_2 \cdot x_2$$

$$y = a - b$$

ALGORITMO 2

$$a = x_1 + x_2$$

$$b = x_1 - x_2$$

$$y = a \cdot b$$

Per la seconda funzione, tre algoritmi sono basati sullo scrivere $f(x)$ come

$$f(x) = x_1^4 - x_2^4 = (x_1^2 + x_2^2)(x_1^2 - x_2^2) = (x_1^2 + x_2^2)(x_1 + x_2)(x_1 - x_2)$$

Questi algoritmi sono

ALGORITMO 1

$$a = x_1 \cdot x_1$$

$$b = a \cdot a$$

$$c = x_2 \cdot x_2$$

$$d = c \cdot c$$

$$y = b - d$$

ALGORITMO 2

$$a = x_1 \cdot x_1$$

$$b = x_2 \cdot x_2$$

$$c = a + b$$

$$d = a - b$$

$$y = a \cdot b$$

ALGORITMO 3

$$a = x_1 \cdot x_1$$

$$b = x_2 \cdot x_2$$

$$c = a + b$$

$$d = x_1 + x_2$$

$$e = x_1 - x_2$$

$$f = c \cdot d$$

$$y = d \cdot e$$

Esercizio. Mostrare che tale termine in (3) ha modulo minore o uguale a $4 \max\{|\varepsilon_{\text{in}}|, |\varepsilon_{\text{mac}}|, |\varepsilon_{\text{alg}}|\}^2$ per $\max\{|\varepsilon_{\text{in}}|, |\varepsilon_{\text{mac}}|, |\varepsilon_{\text{alg}}|\} \leq 1$.

Si ha

$$\varepsilon_y = \varepsilon_{\text{alg}} + \varepsilon_{\text{mac}} + \varepsilon_{\text{in}} + d$$

con

$$d = \varepsilon_{\text{alg}}\varepsilon_{\text{mac}} + \varepsilon_{\text{alg}}\varepsilon_{\text{in}} + \varepsilon_{\text{mac}}\varepsilon_{\text{in}} + \varepsilon_{\text{alg}}\varepsilon_{\text{mac}}\varepsilon_{\text{in}}.$$

Posto

$$M = \max\{|\varepsilon_{\text{in}}|, |\varepsilon_{\text{mac}}|, |\varepsilon_{\text{alg}}|\}$$

risulta, per $M \leq 1$,

$$\begin{aligned} |d| &= |\varepsilon_{\text{alg}}\varepsilon_{\text{mac}} + \varepsilon_{\text{alg}}\varepsilon_{\text{in}} + \varepsilon_{\text{mac}}\varepsilon_{\text{in}} + \varepsilon_{\text{alg}}\varepsilon_{\text{mac}}\varepsilon_{\text{in}}| \\ &\leq |\varepsilon_{\text{alg}}\varepsilon_{\text{mac}}| + |\varepsilon_{\text{alg}}\varepsilon_{\text{in}}| + |\varepsilon_{\text{mac}}\varepsilon_{\text{in}}| + |\varepsilon_{\text{alg}}\varepsilon_{\text{mac}}\varepsilon_{\text{in}}| \\ &= |\varepsilon_{\text{alg}}| |\varepsilon_{\text{mac}}| + |\varepsilon_{\text{alg}}| |\varepsilon_{\text{in}}| + |\varepsilon_{\text{mac}}| |\varepsilon_{\text{in}}| + |\varepsilon_{\text{alg}}| |\varepsilon_{\text{mac}}| |\varepsilon_{\text{in}}| \\ &\leq M^2 + M^2 + M^2 + \underbrace{M^3}_{=M^2 \cdot M \leq M^2} \\ &\leq M^2 + M^2 + M^2 + M^2 = 4M^2. \end{aligned}$$

Esercizio. Quando x_2 è vicino a -1 si può concludere che ε_{in} ha ordine di grandezza superiore a quello di $\widehat{\varepsilon}_1$ e $\widehat{\varepsilon}_2$, assunti non nulli e dello stesso ordine di grandezza?

Si ha

$$\varepsilon_{\text{in}} \doteq \widehat{\varepsilon}_1 + \frac{x_2}{x_2 + 1} \widehat{\varepsilon}_2.$$

Per x_2 è vicino a -1 si può concludere che

$$\frac{x_2}{x_2 + 1} \widehat{\varepsilon}_2 \gg \widehat{\varepsilon}_2$$

ma anche

$$\frac{x_2}{x_2 + 1} \widehat{\varepsilon}_2 \gg \widehat{\varepsilon}_1,$$

essendo $\widehat{\varepsilon}_1$ e $\widehat{\varepsilon}_2$ dello stesso ordine di grandezza. Per cui,

$$\varepsilon_{\text{in}} \doteq \widehat{\varepsilon}_1 + \frac{x_2}{x_2 + 1} \widehat{\varepsilon}_2 \approx \frac{x_2}{x_2 + 1} \widehat{\varepsilon}_2$$

ha ordine di grandezza superiore a quello di $\widehat{\varepsilon}_1$ e $\widehat{\varepsilon}_2$.

Esercizio. Si studi il condizionamento di

$$f(x) = \sqrt{x_1 + x_2} - \sqrt{x_1}, \quad x \in \mathbb{R}^2 \text{ con } x_1, x_2 > 0,$$

e la stabilità dei due algoritmi basati sulle due diverse espressioni

$$f(x) = \sqrt{x_1 + x_2} - \sqrt{x_1} = \frac{x_2}{\sqrt{x_1 + x_2} + \sqrt{x_1}}.$$

Gli indici di condizionamento di f sono

$$\begin{aligned} K_1(f, x) &= \frac{x_1}{f(x)} \cdot \frac{\partial f}{\partial x_1}(x) = \frac{x_1}{\sqrt{x_1 + x_2} - \sqrt{x_1}} \cdot \left(\frac{1}{2\sqrt{x_1 + x_2}} - \frac{1}{2\sqrt{x_1}} \right) \\ &= \frac{x_1}{\sqrt{x_1 + x_2} - \sqrt{x_1}} \cdot \frac{\sqrt{x_1} - \sqrt{x_1 + x_2}}{2\sqrt{x_1 + x_2}\sqrt{x_1}} \\ &= -\frac{1}{2} \sqrt{\frac{x_1}{x_1 + x_2}} \end{aligned}$$

e

$$\begin{aligned} K_2(f, x) &= \frac{x_2}{f(x)} \cdot \frac{\partial f}{\partial x_2}(x) = \frac{x_2}{\sqrt{x_1+x_2} - \sqrt{x_1}} \cdot \frac{1}{2\sqrt{x_1+x_2}} \\ &= \frac{x_2}{x_2} \cdot \frac{\sqrt{x_1+x_2} + \sqrt{x_1}}{2\sqrt{x_1+x_2}} = \frac{\sqrt{x_1+x_2} + \sqrt{x_1}}{2\sqrt{x_1+x_2}}. \end{aligned}$$

Ora

$$|K_1(f, x)| = \left| -\frac{1}{2} \sqrt{\frac{x_1}{x_1+x_2}} \right| \leq \frac{1}{2}$$

e

$$K_2(f, x) = \frac{\sqrt{x_1+x_2} + \sqrt{x_1}}{2\sqrt{x_1+x_2}} \leq 1.$$

Per cui, f è ben condizionata su ogni dato.

Analizziamo per la stabilità il primo algoritmo, quello basato su

$$f(x) = \sqrt{x_1+x_2} - \sqrt{x_1}.$$

Si ha

$$\begin{aligned} 1 \quad a &= x_1 + x_2 & \beta_a &\doteq \gamma_1 \\ 2 \quad b &= \sqrt{a} & \beta_b &\doteq \frac{1}{2}\beta_a + \gamma_2 \doteq \frac{1}{2}\gamma_1 + \gamma_2 \\ 3 \quad c &= \sqrt{x_1} & \beta_c &\doteq \gamma_3 \\ 4 \quad y &= b - c & \varepsilon_{\text{alg}} &\doteq \frac{b}{b-c}\beta_b - \frac{c}{b-c}\beta_c + \gamma_4 \doteq \frac{1}{2} \frac{b}{b-c}\gamma_1 + \frac{b}{b-c}\gamma_2 - \frac{c}{b-c}\gamma_3 + \gamma_4. \end{aligned}$$

Così

$$\begin{aligned} M_1(x) &= \frac{1}{2} \cdot \frac{b}{b-c} = \frac{1}{2} \cdot \frac{\sqrt{x_1+x_2}}{\sqrt{x_1+x_2} - \sqrt{x_1}} = \frac{1}{2} \cdot \frac{\sqrt{x_1+x_2}(\sqrt{x_1+x_2} + \sqrt{x_1})}{x_2} \\ &= \frac{1}{2} \cdot \frac{\sqrt{x_2}\sqrt{x_2}}{x_2} \cdot \sqrt{\frac{x_1}{x_2} + 1} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \\ &= \frac{1}{2} \sqrt{\frac{x_1}{x_2} + 1} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \\ M_2(x) &= 2M_1(x) = \sqrt{\frac{x_1}{x_2} + 1} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \\ M_3(x) &= -\frac{c}{b-c} = -\frac{\sqrt{x_1}}{\sqrt{x_1+x_2} - \sqrt{x_1}} = -\frac{\sqrt{x_1}(\sqrt{x_1+x_2} + \sqrt{x_1})}{x_2} \\ &= -\sqrt{\frac{x_1}{x_2}} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \end{aligned}$$

Avendosi

$$\begin{aligned}\frac{x_1}{x_2} &\leq M_1(x) = \frac{1}{2}\sqrt{\frac{x_1}{x_2} + 1} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \leq \frac{x_1}{x_2} + 1 \\ 2\frac{x_1}{x_2} &\leq M_2(x) = 2M_1(x) \leq 2\left(\frac{x_1}{x_2} + 1\right) \\ 2\frac{x_1}{x_2} &\leq |M_3(x)| = \sqrt{\frac{x_1}{x_2}} \left(\sqrt{\frac{x_1}{x_2} + 1} + \sqrt{\frac{x_1}{x_2}} \right) \leq 2\left(\frac{x_1}{x_2} + 1\right).\end{aligned}$$

Per cui, l'algoritmo è instabile se e solo se $\frac{x_1}{x_2}$ è grande.

Analizziamo per la stabilità il secondo algoritmo, quello basato su

$$f(x) = \frac{x_2}{\sqrt{x_1 + x_2} + \sqrt{x_1}}.$$

Si ha

$$\begin{aligned}1 \quad a &= x_1 + x_2 \quad \beta_a \doteq \gamma_1 \\ 2 \quad b &= \sqrt{a} \quad \beta_b \doteq \frac{1}{2}\beta_a + \gamma_2 \doteq \frac{1}{2}\gamma_1 + \gamma_2 \\ 3 \quad c &= \sqrt{x_1} \quad \beta_c \doteq \gamma_3 \\ 4 \quad d &= b + c \quad \beta_d \doteq \frac{b}{b+c}\beta_b + \frac{c}{b+c}\beta_c + \gamma_4 \doteq \frac{1}{2}\frac{b}{b+c}\gamma_1 + \frac{b}{b+c}\gamma_2 + \frac{c}{b+c}\gamma_3 + \gamma_4 \\ 5 \quad y &= \frac{x_2}{d} \quad \varepsilon_{\text{alg}} \doteq -\beta_d + \gamma_5 \doteq -\frac{1}{2}\frac{b}{b+c}\gamma_1 - \frac{b}{b+c}\gamma_2 - \frac{c}{b+c}\gamma_3 - \gamma_4 + \gamma_5.\end{aligned}$$

Così

$$\begin{aligned}M_1(x) &= -\frac{1}{2} \cdot \frac{b}{b+c} = -\frac{1}{2} \cdot \frac{\sqrt{x_1 + x_2}}{\sqrt{x_1 + x_2} + \sqrt{x_1}} \\ M_2(x) &= 2M_1(x) = -\frac{\sqrt{x_1 + x_2}}{\sqrt{x_1 + x_2} + \sqrt{x_1}} \\ M_3(x) &= -\frac{c}{b-c} = -\frac{\sqrt{x_1}}{\sqrt{x_1 + x_2} + \sqrt{x_1}} \\ M_4(x) &= -1.\end{aligned}$$

Avendosi $|M_j(x)| \leq 1$, $j \in \{1, 2, 3, 4\}$, l'algoritmo è stabile per ogni dato.

Esercizio. Si studi il condizionamento di

$$f(x) = x_1x_2 + x_1x_3, \quad x \in \mathbb{R}^3,$$

e la stabilità dei due algoritmi basati sulle due diverse espressioni

$$f(x) = x_1x_2 + x_1x_3 = x_1(x_2 + x_3).$$

Gli indici di condizionamento di f sono

$$\begin{aligned} K_1(f, x) &= \frac{x_1}{f(x)} \cdot \frac{\partial f}{\partial x_1}(x) = \frac{x_1}{x_1x_2 + x_1x_3} \cdot (x_2 + x_3) = 1 \\ K_2(f, x) &= \frac{x_2}{f(x)} \cdot \frac{\partial f}{\partial x_2}(x) = \frac{x_2}{x_1x_2 + x_1x_3} \cdot x_1 = \frac{x_2}{x_2 + x_3} = \frac{1}{1 + \frac{x_3}{x_2}} \\ K_3(f, x) &= \frac{x_3}{f(x)} \cdot \frac{\partial f}{\partial x_3}(x) = \frac{x_3}{x_1x_2 + x_1x_3} \cdot x_1 = \frac{x_3}{x_2 + x_3} = \frac{1}{1 + \frac{x_2}{x_3}}. \end{aligned}$$

Per cui, f è mal condizionata se e solo se $\frac{x_2}{x_3}$ è vicino a -1 .

Analizziamo per la stabilità il primo algoritmo, quello basato su

$$f(x) = x_1x_2 + x_1x_3.$$

Si ha

$$\begin{aligned} 1 \quad a &= x_1x_2 \quad \beta_a \doteq \gamma_1 \\ 2 \quad b &= x_1x_3 \quad \beta_b \doteq \gamma_2 \\ 3 \quad y &= a + b \quad \varepsilon_{\text{alg}} \doteq \frac{a}{a+b}\beta_a + \frac{b}{a+b}\beta_b + \gamma_3 \doteq \frac{a}{a+b}\gamma_1 + \frac{b}{a+b}\gamma_2 + \gamma_3. \end{aligned}$$

Così

$$\begin{aligned} M_1(x) &= \frac{a}{a+b} = \frac{x_1x_2}{x_1x_2 + x_1x_3} = \frac{x_2}{x_2 + x_3} = \frac{1}{1 + \frac{x_3}{x_2}} \\ M_2(x) &= \frac{b}{a+b} = \frac{x_1x_3}{x_1x_2 + x_1x_3} = \frac{x_3}{x_2 + x_3} = \frac{1}{1 + \frac{x_2}{x_3}} \end{aligned}$$

Per cui, l'algoritmo è instabile se e solo se $\frac{x_2}{x_3}$ è vicino a -1 .

Analizziamo per la stabilità il secondo algoritmo, quello basato su

$$f(x) = x_1(x_2 + x_3)$$

Si ha

$$\begin{aligned} 1 \quad a &= x_2 + x_3 \quad \beta_a \doteq \gamma_1 \\ 2 \quad y &= x_1 \cdot a \quad \varepsilon_{\text{alg}} \doteq \beta_a + \gamma_2. \end{aligned}$$

Così

$$M_1(x) = 1$$

e l'algoritmo è stabile per ogni dato.

Esercizio. Si studi il condizionamento di

$$f(x) = \log \frac{x+1}{x}, \quad x > 0,$$

e la stabilità dei tre algoritmi basati sulle tre diverse espressioni

$$f(x) = \log \frac{x+1}{x} = \log \left(1 + \frac{1}{x} \right) = \log(x+1) - \log x.$$

L'indice di condizionamento di f è

$$\begin{aligned} K(f, x) &= \frac{x}{f(x)} \cdot f'(x) = \frac{x}{\log \left(1 + \frac{1}{x} \right)} \cdot \frac{1}{1 + \frac{1}{x}} \cdot \left(-\frac{1}{x^2} \right) \\ &= -\frac{\frac{1}{x}}{\left(1 + \frac{1}{x} \right) \log \left(1 + \frac{1}{x} \right)} \end{aligned}$$

Avendosi

$$\begin{aligned} \lim_{x \rightarrow 0} K(f, x) &= 0 \\ \lim_{x \rightarrow +\infty} K(f, x) &= -1 \end{aligned}$$

si conclude che f è ben condizionata su ogni dato.

Analizziamo per la stabilità il primo algoritmo, quello basato su

$$f(x) = \log \frac{x+1}{x}.$$

Si ha

$$\begin{aligned} 1 \quad a &= x+1 \quad \beta_a \doteq \gamma_1 \\ 2 \quad b &= \frac{a}{x} \quad \beta_b \doteq \beta_a + \gamma_2 \doteq \gamma_1 + \gamma_2 \\ 3 \quad y &= \log b \quad \varepsilon_{\text{alg}} \doteq \frac{1}{\log b} \beta_b + \gamma_3 \doteq \frac{1}{\log b} \gamma_1 + \frac{1}{\log b} \gamma_2 + \gamma_3. \end{aligned}$$

Così

$$M_1(x) = M_2(x) = \frac{1}{\log b} = \frac{1}{\log \left(1 + \frac{1}{x} \right)}$$

Per cui, l'algoritmo è instabile se e solo se x è grande.

Analizziamo per la stabilità il secondo algoritmo, quello basato su

$$f(x) = \log \left(1 + \frac{1}{x} \right).$$

Si ha

$$\begin{aligned} 1 \quad a &= \frac{1}{x} \quad \beta_a \doteq \gamma_1 \\ 2 \quad b &= 1+a \quad \beta_b \doteq \frac{1}{1+a} \beta_a + \gamma_2 \doteq \frac{1}{1+a} \gamma_1 + \gamma_2 \\ 3 \quad y &= \log b \quad \varepsilon_{\text{alg}} \doteq \frac{1}{\log b} \beta_b + \gamma_3 \doteq \frac{1}{\log b \cdot (1+a)} \gamma_1 + \frac{1}{\log b} \gamma_2 + \gamma_3. \end{aligned}$$

Così

$$\begin{aligned} M_1(x) &= \frac{1}{\log b \cdot (1+a)} = \frac{1}{\log(1+\frac{1}{x}) \cdot (1+\frac{1}{x})} \\ M_2(x) &= \frac{1}{\log b} = \frac{1}{\log(1+\frac{1}{x})} \end{aligned}$$

Per cui, l'algoritmo è instabile se e solo se x è grande.

Analizziamo la stabilità del terzo algoritmo, quello basato su

$$f(x) = \log(x+1) - \log x.$$

Si ha

$$\begin{aligned} 1 \quad a &= x+1 \quad \beta_a \doteq \gamma_1 \\ 2 \quad b &= \log a \quad \beta_b \doteq \frac{1}{\log a} \beta_a + \gamma_2 \doteq \frac{1}{\log a} \gamma_1 + \gamma_2 \\ 3 \quad c &= \log x \quad \beta_c \doteq \gamma_3 \\ 4 \quad y &= b-c \quad \varepsilon_{\text{alg}} \doteq \frac{b}{b-c} \beta_b - \frac{c}{b-c} \beta_c + \gamma_4 \doteq \frac{b}{b-c} \cdot \frac{1}{\log a} \gamma_1 + \frac{b}{b-c} \gamma_2 - \frac{c}{b-c} \gamma_3 + \gamma_4 \end{aligned}$$

Così

$$\begin{aligned} M_1(x) &= \frac{b}{b-c} \cdot \frac{1}{\log a} = \frac{1}{b-c} = \frac{1}{\log(1+\frac{1}{x})} \\ M_2(x) &= \frac{b}{b-c} = \frac{\log(x+1)}{\log(1+\frac{1}{x})} = \frac{\log x + \log(1+\frac{1}{x})}{\log(1+\frac{1}{x})} \\ &= \frac{\log x}{\log(1+\frac{1}{x})} + 1 = -\frac{\log \frac{1}{x}}{\log(1+\frac{1}{x})} + 1 \\ M_3(x) &= -\frac{c}{b-c} = -\frac{\log x}{\log(1+\frac{1}{x})} = \frac{\log \frac{1}{x}}{\log(1+\frac{1}{x})} \end{aligned}$$

Per cui, l'algoritmo è instabile se e solo se x è grande.

Esercizio. Si supponga di essere nella Situazione 4 con il dato x tale che $x = \hat{x} = \text{fl}(\hat{x})$, cioè x è un dato che non si ottiene da una misurazione e le componenti $x_1, \dots, x_n$ di x sono numeri di macchina. E' utile in questo caso sostituire l'algoritmo instabile per calcolare i valori di f con uno stabile?

In questo caso si ha

$$\varepsilon_y = \varepsilon_{\text{alg}}$$

(risulta

$$\varepsilon_{\text{in}} = 0 \text{ e } \varepsilon_{\text{mac}} = 0).$$

Per cui, è utile passare da un algoritmo instabile ad uno stabile, in quanto ε_y passerà dall'avere "in generale" ordine di grandezza maggiore di eps ad avere ordine di grandezza non superiore a eps.