Esercizi su aritmetica di macchina, analisi dell'errore e programmazione in ambiente Python

23 ottobre 2019

Nota: gli esercizi più impegnativi sono contrassegnati dal simbolo ($\star$).

Richiami/Notazioni:

- Insieme dei numeri di macchina *normali*:

$$\mathbb{F}(\beta, t, M_1, M_2) := \{x \in \mathbb{R} \mid x = \pm\, \beta^p \sum_{i=0}^{t} d_i \beta^{-i}\},$$

  dove:

  (i) $d_i \in \mathbb{N}$, $0 \leq d_i \leq \beta - 1$ per ogni $i = 0, \ldots, t$,

  (ii) $d_0 \neq 0$,

  (iii) $p \in \mathbb{Z}$, $M_1 + 1 \leq p \leq M_2 - 1$.

- Numeri di macchina *denormali* relativi a $\mathbb{F}(\beta, t, M_1, M_2)$:

$$\{x \in \mathbb{R} \mid x = \pm\, \beta^{M_1+1} \sum_{i=1}^{t} d_i \beta^{-i},\ \text{con almeno un } d_i \neq 0\}$$

- Precisione di macchina:
  $\texttt{eps} := \beta^{-t}$

- Arrotondamento:

$$\mathrm{fl} : x \in \mathbb{R} \mapsto \mathrm{fl}(x) \in \mathcal{M} = \mathbb{F} \cup \{0, \pm\texttt{Inf}\}$$

  oppure

$$\mathrm{fl} : x \in \mathbb{R} \mapsto \mathrm{fl}(x) \in \widetilde{\mathcal{M}} = \mathbb{F} \cup \{0, \pm\texttt{Inf}, \text{numeri denormali}\}$$

  a seconda che si assuma o meno l'utilizzo dei numeri denormali.

Nota: In assenza di precisazioni, per "numero di macchina" intenderemo un numero di macchina *normale*.

Esercizio 1. Esprimere i numeri 5, 21, 32, 35, 63, 255 in base 2 ed in base esadecimale.

Esercizio 2. Esprimere i numeri 0.1, 0.4, 0.5, 1.4, 2.5 in base 2.

Esercizio 3. Considerato l'insieme dei numeri di macchina $\mathbb{F}(2, 52, -1023, 1024)$, determinare:

1. il più grande numero di macchina,
2. il più piccolo numero di macchina strettamente positivo,
3. il più piccolo numero di macchina denormale strettamente positivo,
4. il più grande numero di macchina denormale,
5. la precisione di macchina (denotata con `eps`),
6. la distanza tra 1 ed il successivo numero di macchina,
7. l'insieme dei numeri reali x tali che $\text{fl}(x) = 0$ (si assuma l'utilizzo dei numeri denormali),
8. l'insieme dei numeri reali x tali che $\text{fl}(x) = 1$,
9. l'insieme dei numeri reali x tali che $\text{fl}(x) = 17$,
10. l'insieme dei numeri reali x tali che $\text{fl}(x) = 1 + \texttt{eps}$.

[Nota: si tratta dell'insieme dei numeri di macchina dello standard IEEE 754 a doppia precisione, utilizzato di default in Python.]

Esercizio 4. Ripetere l'esercizio 3 considerando l'insieme dei numeri di macchina $\mathbb{F}(16, 13, -63, 64)$. Confrontare i due insiemi di numeri di macchina appena considerati. [Nota: si tratta dell'insieme dei numeri di macchina dello standard *IBM floating-point double-precision 64-bit*.]

Esercizio 5. Lo standard IEEE 754 a *quadrupla precisione* adotta il formato in base 2 e memorizza i numeri in un registro a 128 bit, ripartiti nel modo seguente: 1 bit per il segno, 15 per l'esponente, 112 per la mantissa. Determinare il corrispondente valore dei parametri β, t, M_1, M_2 nella notazione $\mathbb{F}(\beta, t, M_1, M_2)$. Ripetere l'esercizio 3 per questo insieme di numeri di macchina.

Esercizio 6. Giustificare la seguente affermazione: considerato l'insieme dei numeri di macchina $\mathbb{F}(\beta, t, M_1, M_2)$, la distanza tra due numeri di macchina consecutivi appartenenti a $[\beta^p, \beta^{p+1}]$ é pari a β^{p-t}.

Esercizio 7. Giustificare la seguente affermazione: l'insieme dei numeri di macchina $\mathbb{F}(\beta, t, M_1, M_2)$ è costituito da $2(\beta - 1)\beta^t(M_2 - M_1 - 1)$ elementi.

Esercizio 8. Determinare il massimo errore relativo commesso nell'arrotondamento *per troncamento* (detto anche *round towards zero*), secondo il quale un numero $x \in \mathbb{R}$ viene arrotondato a $\text{fl}(x) \in \mathbb{F}(\beta, t, M_1, M_2)$ conservandone le $t+1$ cifre più significative, e scartando le cifre rimanenti.

Esercizio 9. Eseguire la seguente istruzione in Python e spiegare il risultato ottenuto:

```
1+(2**-51+2**-52+2**-53)==1+2**-50
```

Esercizio 10. Eseguire la seguente istruzione in Python e spiegare il risultato ottenuto:

```
3+(2**-52+2**-53)==(3+2**-52)+2**-53
```

Dedurne che l'addizione *floating-point* non gode della proprietà associativa.

Esercizio 11. Eseguire le seguenti istruzione in Python e spiegare il risultato ottenuto:

```
eps=2**-52
1+.5*eps-.5*eps==1
```

Esercizio 12. Eseguire le seguenti istruzioni Python e spiegare i risultati ottenuti:

```
v=np.array([40,1e-15,2e-15,3e-15])
w=v.copy()
w.sort()
w.sum()>v.sum()
```

Esercizio 13. Valutare mediante un esperimento Python il numero di FLOPS (*floating-point operations per second*) di cui è capace il proprio computer, distinguendo le prestazioni per ognuna delle 4 operazioni algebriche elementari. [Suggerimento: utilizzare il modulo `timeit` ed un ciclo di `for`.]

Esercizio 14. Determinare, mediante un esperimento Python, come avviene l'arrotondamento $x \mapsto \mathrm{fl}(x)$ secondo il metodo "Round to Nearest, Ties to Even" quando $|x|$ è maggiore del più grande numero di macchina.

Esercizio 15. Verificare mediante un esperimento Python che l'utilizzo di numeri denormali rallenta l'esecuzione delle operazioni floating-point.

Esercizio 16. Siano $x = \texttt{np.pi}$, $y = 3.1415$. Calcolare mediante Python l'errore relativo ε che si commette nell'approssimare x con y. Osservare che il numero di cifre significative corrette dell'approssimazione y è all'incirca pari a

$$-\log_{10}(\varepsilon) \ .$$

Spiegare e generalizzare questa osservazione. [Occorre importare il pacchetto NumPy mediante il comando `import numpy as np` prima di svolgere l'esercizio]

Esercizio 17. ($\star$) Mostrare che, se x e y sono numeri le cui rappresentazioni floating-point in base β hanno le prime q cifre coincidenti, ma non la $(q+1)$-esima, e lo stesso esponente, allora, considerato $\varepsilon = \frac{|x-y|}{|x|}$, si ha

$$\lfloor -\log_{10}(\varepsilon) \rfloor \leq q \leq \lceil -\log_{10}(\varepsilon) \rceil .$$

[Nota: $\lfloor\ \rfloor$ e $\lceil\ \rceil$ denotano rispettivamente le funzioni floor e ceiling.]

Esercizio 18. Eseguire le seguenti istruzioni Python:

```
x=0
delta=0.1
while x!=1:
    x=x+delta
```

Spiegarne il (non) funzionamento. [Suggerimento: si veda l'esercizio 2. Per interrompere l'esecuzione del programma, digitare la sequenza di tasti `CTRL + C`.]

Esercizio 19. Eseguire e spiegare il risultato delle seguenti istruzioni Python:

```
x=1e155
(x-x)*x
x**2-x**2
import numpy as np
x=np.array(x)
x**2-x**2
```

Esercizio 20. Siano $x_1 = 0.123456789123456789$, $x_2 = 0.123456789$. Supponiamo di lavorare su un calcolatore che utilizza l'insieme dei numeri di macchina $\mathbb{F}(10, 9, -10, 10)$. Confrontare il risultato dell'operazione $x_1 - x_2$ effettuata in aritmetica di macchina con il risultato esatto. Determinare la perdita di cifre significative dovuta all'operazione effettuata e verificare che questa sia in accordo con l'analisi del condizionamento delle operazioni elementari. Questo è un esempio di *fenomeno di cancellazione*: l'operazione in aritmetica di macchina ha comportato la "cancellazione" delle 8 cifre meno significative del risultato esatto.

Esercizio 21. Siano $a = 1/3$, $b = 0.3333333333$. Calcolare il valore esatto di $a - b$. Successivamente, calcolare l'espressione $a - b$ mediante Python. [Visualizzare le 16 cifre più significative del risultato.] Spiegare il risultato ottenuto.

Esercizio 22. Dimostrare che il numero di condizionamento del problema di calcolare la funzione $f(x) = \sqrt{x}$ è $\kappa = \frac{1}{2}$.

Esercizio 23. Il calcolo della funzione $f(x) = \sqrt{1+x} - 1$ è poco accurato per x "piccolo". Verificarlo con esperimenti al calcolatore. A cosa è dovuta questa perdita di cifre significative? Riscrivere l'espressione $\sqrt{1+x} - 1$ in modo equivalente, in modo tale da rendere il calcolo di $f(x)$ accurato anche per $x \approx 0$.

Esercizio 24. Ripetere l'esercizio 23 per $g(x) = \log(x+1) - \log(x)$. Questa volta il calcolo del valore della funzione è poco accurato per x "grande".

Esercizio 25. Trovare i numeri di condizionamento delle seguenti funzioni (dipenderanno da x). Determinare i valori di x per i quali il calcolo delle corrispondenti funzioni é mal condizionato:

1) $x - 1$,
2) e^x,
3) $\log(|x|)$,
4) $\sin(x)$,
5) $\sqrt{x^2 + 1} - |x|$.