Relazione primo Laboratorio ALAN

Derqui Andrea 5229435
Eric Fasce Canessa 5299197
Luca Ninivaggi 5312342

Esercizio 1:

Ci viene proposto di dimostrare la proprietà associativa, andando a usare un'espressione parametrica al variare di i: $a = (d_0 + 1) \cdot 10^i$, con $i = 0, 1, \ldots, 6$, $b = (d_1 + 1) \cdot 10^{20}$, $c = -b$,

Dobbiamo verificare$(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}$ e $\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$,ma vedendo i risultati in output questo non accade.

```
        ESERCIZIO 1

i = 0
        (a+b)+c: 0
        a+(b+c): 6
i = 1
        (a+b)+c: 0
        a+(b+c): 60
i = 2
        (a+b)+c: 0
        a+(b+c): 600
i = 3
        (a+b)+c: 0
        a+(b+c): 6000
i = 4
        (a+b)+c: 65536
        a+(b+c): 60000
i = 5
        (a+b)+c: 589824
        a+(b+c): 600000
i = 6
        (a+b)+c: 6.02931e+06
        a+(b+c): 6e+06
```

Possiamo notare come il primo risultato sia di fatto quello errato, mentre il secondo è algebricamente corretto. Questo perché, poiché $\boldsymbol{c} = -\boldsymbol{b}$ ($\boldsymbol{b}$ e $\boldsymbol{c}$ possono causare una cancellazione, che genera la propagazione di un errore che può essere anche molto elevato), possiamo riscrivere l'espressione come

$(\boldsymbol{a} + \boldsymbol{b}) - \boldsymbol{b}$ e $\boldsymbol{a} + (\boldsymbol{b} - \boldsymbol{b})$.

Nel caso di $\boldsymbol{a} + (\boldsymbol{b} - \boldsymbol{b})$ abbiamo la cancellazione effettuata come primo calcolo: questo impedisce all' errore di essere propagato, e perciò l'operazione $(\boldsymbol{b} - \boldsymbol{b})$ si risolve con il valore 0; ci sommiamo poi $\boldsymbol{a}$ e otteniamo quindi il valore di $\boldsymbol{a}$.

Invece nell' espressione $(\boldsymbol{a} + \boldsymbol{b}) - \boldsymbol{b}$ noi andiamo prima a effettuare una somma $(\boldsymbol{a} + \boldsymbol{b})$ e poi andiamo a effettuare una potenziale cancellazione: questa si verifica nel caso in cui $\boldsymbol{a}$ sia diversi ordini di grandezza inferiore a $\boldsymbol{b}$, e perciò la sottrazione può essere paragonata alla cancellazione $\boldsymbol{b} - \boldsymbol{b}$: essendo $\boldsymbol{b}$ un valore molto grande, il suo errore relativo fa "esplodere" il calcolo, e per questo notiamo errori.

All' aumentare di $\boldsymbol{a}$, la cancellazione diventa più simile a una normale sottrazione (cioè $\underline{\boldsymbol{a}}$ e $\boldsymbol{b}$ sono dei un "simile" ordine di grandezza) e quindi l'operazione non amplifica eccessivamente l'errore (ipoteticamente, aumentando ancora $\boldsymbol{a}$, si dimostrerebbe la proprietà associativa).

-Esercizio 2:

L'esercizio 2 ci propone di approssimare la funzione $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{e}\hat{}\boldsymbol{x}$ per due punti del dominio: $\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{5}$ e $\boldsymbol{x2} = \boldsymbol{30}$;

Per approssimare una funzione usiamo il polinomio di Taylor, nella forma:

$$f_N(x) = \sum_{n=0}^{N} \frac{x^n}{n!}$$

Calcoliamo il polinomio utilizzando diversi valori di $\boldsymbol{N}$, per poter osservare come questi fanno variare il valore approssimato, mentre usiamo $\boldsymbol{exp}(\boldsymbol{x})$ per controllare quale sarebbe il valore "corretto".

Sottolineiamo innanzitutto come per potere utilizzare un algoritmo capace di poter eseguire tale sommatoria, abbiamo avuto la necessità di andare a "scomporre" ogni elemento della sommatoria, poiché andare a calcolare $\boldsymbol{x}\hat{}\boldsymbol{n}$ e dividerlo per $\boldsymbol{n}!$ non è possibile con un valore "grande" di $\boldsymbol{x}$ o di $\boldsymbol{n}$ (abbiamo un overflow sia su numeratore che denominatore, e quindi dobbiamo modificare l'espressione in modo da mantenere il suo risultato algebrico ma modificandone i passaggi intermedi).

Perciò $\frac{x^n}{n!}$ lo consideriamo come $\left(\frac{x}{n}\right) * \left(\frac{x}{n-1}\right) * \left(\frac{x}{n-2}\right) * \ldots * \left(\frac{x}{1}\right)$; in questo modo abbiamo ridotto l'operazione in parti più piccole, evitando che vada in overflow.

```
x = 0.5; exp(0.5) = 1.64872

        3 -> 1.64583
        10 -> 1.64872
        50 -> 1.64872
        100 -> 1.64872
        150 -> 1.64872

x = 30; exp(30) = 1.06865e+13

        3 -> 4981
        10 -> 2.3883e+08
        50 -> 1.06833e+13
        100 -> 1.06865e+13
        150 -> 1.06865e+13
```

Come possiamo notare, per il punto $x = 0,5$ possiamo calcolare $e\text{^}x$ andando a usare anche solo $N = 3$, per avere un'approssimazione corretta ("corretta" paragonata al valore dato da $exp(x)$, nonostante anch' esso sia approssimato e perturbato da errori) fino a 3 cifre decimali; andando a utilizzare $N = 10$, possiamo avere fino a 5 cifre decimali, e così via.

Nel caso di $x = 30$, abbiamo invece che il valore calcolato è estremamente errato per $N = 3$ e $N = 10$, mentre inizia ad avvicinarsi con $N = 50$.

Questo perché la funzione data restituisce un buona approssimazione dell'esponenziale per punti sul dominio vicino allo $\mathbf{0}$ (indipendentemente dal grado N, $f(x) = 1$), mentre più ci allontaniamo dall' origine, maggiore sarà il valore dell'immagine (tende a infinito, con una velocità sempre maggiore); perciò se volessimo approssimare dei valori "grandi" avremmo bisogno di un polinomio di grado sempre maggiore.

```
x = -0.5; exp(-0.5) = 0.606531

        3 -> 0.604167
        10 -> 0.606531
        50 -> 0.606531
        100 -> 0.606531
        150 -> 0.606531

x = -30; exp(-30) = 9.35762e-14

        3 -> -4079
        10 -> 1.21255e+08
        50 -> 8.78229e+08
        100 -> -1.24766e-08
        150 -> -1.24893e-08
```

Andiamo poi a vedere come l'approssimazione si comporta nel caso in cui ci i valori presi sul dominio siano negativi.

Come nel caso dei positivi, $x = -0.5$ viene approssimato quasi correttamente: questo si spiega analogamente a prima.

Nel caso di $x = -30$, invece i valori risultano errati, e anche in questo caso per poterli approssimare dobbiamo utilizzare un grado elevato del polinomio: abbiamo però inoltre il problema che anche che il risultato si avvicina a quello atteso, il segno potrebbe essere opposto, e perciò nella pratica questo può portare diversi problemi (es. la funzione indica una probabilità, e se questa fosse negativa?).

Il valore estremamente negativo del polinomio con $N = 3$ lo possiamo inoltre spiegare osservando che è un polinomio di grado dispari, e perciò a $-\infty$ tende a $-\infty$.

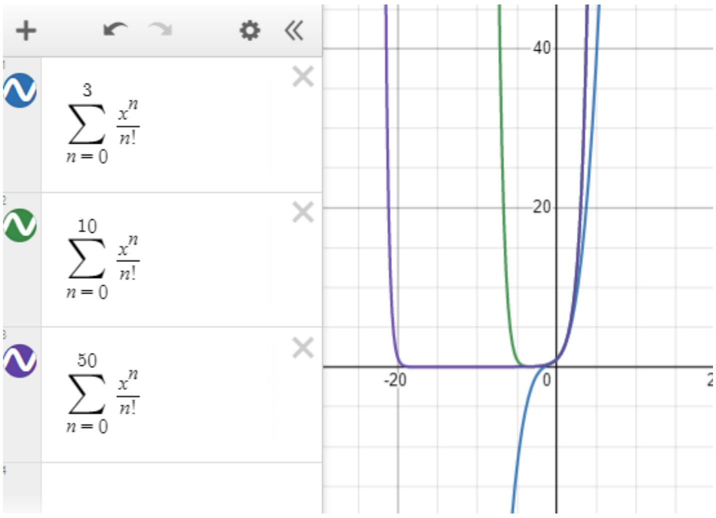
Questo poiché la nostra funzione è descritta da una sommatoria, i cui elementi sono (nel caso in cui **x** sia negativa) alternati positivi e negativi: se il grado del polinomio è pari, avremo come elemento più significativo un valore positivo, altrimenti negativo, e perciò tenderà a $-\infty$.

Notando che $f(x) = e\verb|^|x$ ha la proprietà per cui $f(-x) = 1/e\verb|^|x$, modifichiamo il modo in cui calcoliamo $f(x)$ nel caso di x negativi: calcoliamo invece $1/f(-x)$.

```
x = -0.5; exp(-0.5) = 0.606531

        3 -> 0.607595
        10 -> 0.606531
        50 -> 0.606531
        100 -> 0.606531
        150 -> 0.606531

x = -30; exp(-30) = 9.35762e-14

        3 -> 0.000200763
        10 -> 4.18709e-09
        50 -> 9.36041e-14
        100 -> 9.35762e-14
        150 -> 9.35762e-14
```

Possiamo ora osservare come i valori si avvicino al risultato atteso, con la stessa velocità del primo esempio: se $x = -0.5$ il suo valore dipenderà da quanto calcolato da $f(0.5)$, e perciò risulta correttamente approssimato (3 cifre decimali) anche solo con $N = 3$.

Invece per $x = -30$, andiamo a calcolare $f(30)$, e questo dipende perciò dal valore di N: infatti, $N = 3$ e $N = 10$ risultano molto errati, mentre con $N = 50$ iniziamo ad avvicinarci al valore corretto.



Per poter confutare le nostre osservazioni, possiamo vedere come i grafici confermino i risultati ottenuti, e di come più ci allontaniamo dall' origine, maggiore sarà il grado necessario per poter approssimare correttamente la funzione $e\verb|^|x$.

Notiamo inoltre come N pari/dispari descriva un diverso comportamento a $-\infty$, e capiamo perché l'immagine dei valori negativi possa essere anch' essa negativa.

-Esercizio 3

L' esercizio 3 consiste nel calcolare la precisione di macchina ***eps*** per la precisione singola e doppia, rispettivamente tipi float e double.
eps è un valore costante che rappresenta il più piccolo valore diverso da 0 e vale $eps = 2^{-d}$, tale per cui $1 + 2^{-d} > 1$.

```
        ESERCIZIO 3. Precisione di macchina

Singola precisione: 5.96046e-08. Il valore d corrisponde a 24.
Doppia precisione: 1.11022e-16. Il valore d corrisponde a 53.
```

Tramite cicli di iterazione, eleviamo 2 alla -i (iteratore di un ciclo). Verificando la condizione $1 + 2^{-d} > 1$, otterremmo sia il valore di ***eps*** sia l'intero i al quale l'iteratore si è fermato, che corrisponde alla precisone di bit che il tipo usa nella rappresentazione.

I valori ottenuti corrispondono a quanto stimato, poiché gli esperimenti sono stati fatti in c, dove i float hanno una mantissa di 24 bit e i double di 53.

OUTPUT DELL’ ESEGUIBILE:

ESERCIZIO 1

i = 0

(a+b)+c: 0

a+(b+c): 6

i = 1

(a+b)+c: 64

a+(b+c): 60

i = 2

(a+b)+c: 608

a+(b+c): 600

i = 3

(a+b)+c: 6016

a+(b+c): 6000

i = 4

(a+b)+c: 60000

a+(b+c): 60000

i = 5

(a+b)+c: 600000

a+(b+c): 600000

i = 6

(a+b)+c: 6e+06

a+(b+c): 6e+06

ESERCIZIO 2 algoritmo 1. f(x)=e^x

x = 0.5; exp(0.5) = 1.64872

3 -> 1.64583

10 -> 1.64872

50 -> 1.64872

100 -> 1.64872

150 -> 1.64872

x = 30; exp(30) = 1.06865e+13

3 -> 4981

10 -> 2.3883e+08

50 -> 1.06833e+13

100 -> 1.06865e+13

150 -> 1.06865e+13

x = -0.5; exp(-0.5) = 0.606531

3 -> 0.604167

10 -> 0.606531

50 -> 0.606531

100 -> 0.606531

150 -> 0.606531

x = -30; exp(-30) = 9.35762e-14

3 -> -4079

10 -> 1.21255e+08

50 -> 8.78229e+08

100 -> -1.24766e-08

150 -> -1.24893e-08

ESERCIZIO 2 algoritmo 2. Nota che f(-x) = 1/f(x)

x = -0.5; exp(-0.5) = 0.606531

3 -> 0.607595

10 -> 0.606531

50 -> 0.606531

100 -> 0.606531

150 -> 0.606531

x = -30; exp(-30) = 9.35762e-14

3 -> 0.000200763

10 -> 4.18709e-09

50 -> 9.36041e-14

100 -> 9.35762e-14

150 -> 9.35762e-14

ESERCIZIO 3. Precisione di macchina

Singola precisione: 5.42101e-20. Il valore d corrisponde a 64.

Doppia precisione: 5.42101e-20. Il valore d corrisponde a 64.