

# Appunti di calcolo numerico

Matteo cocciniglia

2021-2022

## Contents

<b>1 Rappresentazione posizionale</b>	<b>1</b>
<b>2 Rappresentazione floating point</b>	<b>1</b>
<b>3 Errori di macchina e operazioni</b>	<b>2</b>
<b>4 matrici simmetriche</b>	<b>3</b>
4.1 Teorema di fattorizzazione di Gauss per matrici simmetriche . . .	3
4.2 dimosotrazione . . . . .	3

## 1 Rappresentazione posizionale

Per rappresentare un qualsiasi numero  $N \geq 1$  in notazione posizionale usando una base  $\beta$  (dove  $\beta = \text{numero di simboli } S - 1$ ) si può stabilire la relazione

$$N = (c_p c_{p-1} \cdots c_0)_\beta = (c_p \beta^p c_{p-1} \beta^{p-1} \cdots c_0).$$

Per un numero  $\alpha < 1$  si ha  $\alpha = (0.c_p c_{p-1} \cdots c_0)_\beta$  rappresentabile anche come la sommatoria  $\sum_{i=1}^{\infty} a_i \beta^{-i}$  quindi si ha che un qualsiasi numero  $\alpha \in \mathbb{R}$  è rappresentabile come

$$\alpha = \sum_{i=1}^{\infty} a_i \beta^{-i} \beta^p$$

dove  $p$  è il numero di simboli dopo il punto radice  $\neq 0$ .

Riassumendo: un qualsiasi numero  $\alpha \in \mathbb{R}$  è rappresentabile in notazione **Pozizionale** come  $\text{segno}(\alpha) m \beta^p$  dove  $m = \sum_{i=1}^{\infty} a_i \beta^{-i}$

## 2 Rappresentazione floating point

Per memorizzare un qualsiasi numero reale ma in uno spazio finito di memorie viene utilizzata la rappresentazione a **virgola mobile** che a differenza di altre rappresentazioni (es. virgola fissa) permette grande flessibilità. Il formato di rappresentazione è caratterizzato da 4 elementi:

$$\beta, t, L, U$$

- $\beta$  è la base usata, di norma base 2.
- $t$  è il numero massimo di numeri dedicato alla rappresentazione della mantissa, di norma **single precision** = 23 **double precision** = 52.
- $L, U$  sono rispettivamente il **Lower** e **Upper bound** per la rappresentazione di un esponente  $r$  quindi il numero minimo e massimo che si può rappresentare, di norma  $[-127, +128]$ .

altre caratteristiche del insieme dei numeri floating point

- è un insieme **discreto** e **finito**.
- è **simmetrico** al origine, quindi per ogni numero esiste l'opposto.
- la sua cardinalità è data da  $card(\beta, t, L, U) = 2(\beta - 1)\beta^{t-1}(U - L + 1)$  dove
  - $(U - L + 1)$  è il numero di tutti i possibili esponenti
  - $\beta^{t-1}$  è il numero di tutti i numero di valori che ogni posizione dedicata alla mantissa può prendere escluso un elemento.
  - $\beta - 1$  è il numero che il primo elemento può prendere (-1 perché non può essere zero).
  - 2 perché un valore può essere sia positivo che negativo.

### 3 Errori di macchina e operazioni

Dato che l'insieme dei **Floating point** è un insieme discreto e finito spesso si fa uso di **troncamento** e **arrotondamento** per rappresentare i numeri non appartenenti all'insieme di riferimento ottenendo una perdita di informazione chiamata **errore**.

L'**errore assoluto** nella rappresentazione di un numero  $\alpha$  è la differenza tra il numero e la sua rappresentazione, quindi si ottiene che:  $E_a = |\alpha - \alpha^*|$  dove  $\alpha^*$  è la rappresentazione macchiana.

L'errore relativo invece è dato da  $E_r = \frac{E_a}{|\alpha|}$ . Inoltre si ha che l'errore relativo ottenuto è sempre minore o uguale alla **precisione di macchina** rappresentata come  $k\beta^{t-1}$  dove

- $k = 1$  nel caso si utilizzi logiche di approssimazione tramite toroncamento.
- $k = \frac{1}{2}$  nel caso si usi arrotondamento.

questo risultato è chiamato **teorema dell'errore di rappresentazione dei numeri reali**.

Dato che l'insieme dei floating point  $F$  non è chiuso alle operazioni aritmetiche si possono verificare, oltre ai casi di **overflow** e **underflow**, dei casi in cui il

numero non è rappresentabile pur essendo compreso tra il numero massimo e il numero minimo rappresentabile del insieme, difatti si ha che un'operazione tra numeri floating point restituisce un numero reale che andrà trasformato seguendo la logica vista nel capitolo Rappresentazione floating point

Matteo Cocciniglia 02/03/2022

## 4 matrici simmetriche

look at pag 167

Ricordando che:

- Una matrice quadrata  $A$  è simmetrica se  $A = A^T$ .
- Per memorizzare una matrice in modo ottimizzato è richiesto  $\simeq \frac{n^2}{2}$  locazioni di memoria.

### 4.1 Teorema di fattorizzazione di Gauss per matrici simmetriche

Se  $A$  è simmetrica e tutti i suoi minori principali sono diversi da zero, allora esistono una matrice triangolare inferiore  $L$  una diagonale  $D$  con elementi  $\neq 0$  tali che  $A = LDL^T$

### 4.2 dimostrazione

Sfruttando l'ipotesi che tutti i minori di  $A$  sono  $\neq 0$  dai risultati del **teorema di fattorizzazione di Gauss** si ottiene che  $A = LU$  dove  $L$  è una matrice triangolare inferiore con diagonale unitaria e  $U$  matrice diagonale superiore.

Si definisce  $D = \begin{pmatrix} u_{11} & 0 & \cdots \\ 0 & u_{22} & & \\ \vdots & & \ddots & \\ & & & u_{nn} \end{pmatrix}$ , quindi si ha che  $A = LDD^{-1}U$

dato che  $DD^{-1}$  è la matrice identità il risultato non è alterato.

Provare che  $D^{-1}U \stackrel{?}{=} L^T$ . Sfruttando la simmetria della matrice  $A = A^T$  si ottiene  $A = (LDD^{-1}U)^T$ .