eCAMPUS UNIVERSITÀ

DiSTA

Corso: Analisi Numerica

Docente: Roberto Piersanti

Calcolo degli autovalori e fondamenti della matematica numerica

Lezione 4.8a

Operazioni floating point

Fondamenti della matematica numerica

- **La rappresentazione dei numeri macchina**

$$\mathrm{fl}^t(x) = \pm(0.\alpha_1\alpha_2\dots\alpha_{t-1}\alpha_t)_\beta \cdot \beta^e$$

- Operazioni in aritmetica (finita) in virgola mobile

  ✓ Numeri macchina (regioni di **overflow** and **underflow**)

  ✓ Errori di arrotondamento (**round-off**)

  ✓ **Precisione di macchina**

  ✓ Le **operazioni di macchina**

Fondamenti della matematica numerica (numeri macchina)

- Definizione di <u>numero macchina</u>, con una **nomenclatura rigorosa**

$$\mathrm{fl}^t(x) = \pm(0.\alpha_1\alpha_2\ldots\alpha_{t-1}\alpha_t)_\beta \cdot \beta^e$$

$\mathrm{fl}^t(x)$
- ✓ Segno: $\pm$
- ✓ Mantissa: $\alpha_1\alpha_2\ldots\alpha_{t-1}\alpha_t$ (t cifre significative)
- ✓ Base: β
- ✓ Esponente (caratteristica): e

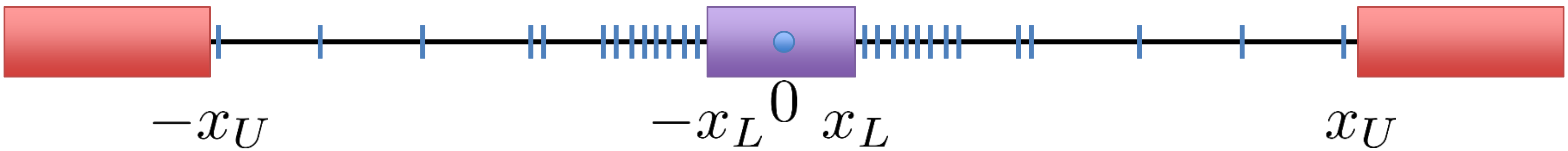
1	t					1
$\pm$	α_1	α_2			α_t	e

$$0 \leq \alpha_k \leq \beta - 1 \qquad 0 > L \leq e \leq U > 0 \quad L, U \in \mathbb{N}$$

$$\alpha_1 \neq 0$$

Limite inferiore (L) Limite superiore (U)

Fondamenti della matematica numerica (numeri macchina)



- <u>Idealizzazione della retta reale</u> come **rappresentazione in macchina**
- **Numeri rappresentabili:** • 0 , | $(-x_U, x_L) \cup (x_L, x_U)$
- **Numeri non rappresentabili:**

  $|x| > x_U$ regione di <u>overflow</u> $|x| < x_U$ regione di <u>underflow</u>

$$\mathbb{F} = \{\mathrm{fl}^t(x) = \pm(0.\alpha_1\alpha_2\ldots\alpha_k\ldots\alpha_{t-1}\alpha_t)_\beta \cdot \beta^e, 0 \leq \alpha_k \leq \beta-1, \alpha_1 \neq 0, L \leq e \leq U\}$$

- N **numero totale di numeri macchina** a disposizione è

$$N = N(t,\beta,U,L) = \underset{\pm}{2}(\beta - \underset{\alpha_1}{1})\underset{\alpha_k}{\beta^{t-1}}(U - \underset{e}{L} + 1) + \underset{0}{1}$$

Fondamenti della matematica numerica (numeri macchina)

- **Esempio** (accademico, non reale)

$$\beta = 2,\ \ t = 3,\ \ L = -2,\ \ U = 3$$

$$N = 2(\beta - 1)\beta^{t-1}(U - L + 1) + 1 = 2(2-1)2^{2-1}(3 + 2 + 1) + 1 = 49$$

- **Errore di rappresentazione** introdotto dal calcolatore

Errore di arrotondamento di un numero reale
Errori di round-off

Fondamenti della matematica numerica (round off)

- Consideriamo un numero reale x in base β e il suo floating $\mathrm{fl}^t(x)$

$$x = \pm(0.\alpha_1\alpha_2\dots\alpha_{t-1}\alpha_t\alpha_{t+1}\dots)_\beta \cdot \beta^e$$

$$\mathrm{fl}^t(x) = \pm(0.\alpha_1\alpha_2\dots\alpha_{t-1}\tilde{\alpha}_t)_\beta \cdot \beta^e$$

$$\tilde{\alpha}_t = \begin{cases} \alpha_t & 0 \le \alpha_{t+1} < \beta/2 \\ 1+\alpha_t & \beta/2 \le \alpha_{t+1} \le \beta \end{cases}$$

$$\Rightarrow \quad x - \mathrm{fl}^t(x) = \pm(0.00\dots0\alpha_{t+1}\dots)_\beta \cdot \beta^e \iff \alpha_{t+1} < \beta/2$$

$$\Rightarrow \quad x - \mathrm{fl}^t(x) = \pm(0.00\dots\alpha_t\alpha_{t+1}\dots)_\beta \cdot \beta^e \iff \alpha_{t+1} \ge \beta/2$$

- Si può **stimare l'errore assoluto e relativo**

$$|x - \mathrm{fl}^t(x)| \le= (0.00\dots0\frac{\beta}{2})_\beta \cdot \beta^e = \frac{1}{2}\beta^{e-t} \qquad \frac{|x - \mathrm{fl}^t(x)|}{|x|} \le \frac{1}{2}\beta^{1-t}$$