CALCOLO NUMERICO

PROF: LUCA GEMIGNANI
DIPARTIMENTO DI INFORMATICA (FIBONACCI)
ESERCITATORI: POLONI, DEL CORSO

INTRODUZIONE

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

↓ Limite Somme di Riemann

$$x_i = a + \frac{(b-a)}{N} \cdot i \qquad i = 0, \dots, N$$

$$\int_a^b f(x)\,dx = \lim_{N\to\infty} \sum_{j=1}^{N} f(x_{j-1})\,(x_j - x_{j-1})$$

$f(x_{j-1})$ ↑ altezza rettangolo; $(x_j - x_{j-1})$ ↑ base rettangolo

Problemi del corso:
- Soluzione di sistemi lineari $Ax = b$
- Calcolo autovettori e autovalori
- Soluzioni di equazioni non lineari $f(x) = 0$ $\quad x - \sin x = 0$
- Integrazione numerica $\int_a^b f(x)\,dx$

FINE INTRODUZIONE

MASTER COPY
Tel. 050 8312126
Cell. 388 9837745

Aritmetica di un Calcolatore

I calcolatori operano maggiormente in aritmetica binaria.
L'insieme dei numeri rappresentabili con un calcolatore è finito.

Sia $B \in \mathbb{N} > 1$ la base con cui contiamo

Se scrivo $(6,56)_7 = 6 \cdot 7^0 + 5 \cdot 7^{-1} + 6 \cdot 7^{-2}$

Questo indica la base

Teorema di Rappresentazione in Base:

Sia $x \in \mathbb{R}$, $x \neq 0$, allora esistono e sono univocamente determinati:

① Una sequenza di numeri

$$\{d_i\}_{i \in \mathbb{N}}, \quad d_1 \neq 0, \quad 0 \leq d_i \leq B-1, \ d_i \text{ non è definitivamente uguale a } B-1.$$

② $p \in \mathbb{Z}$

tale che $x = \text{sgn}(x)\, B^p \underbrace{\sum_{i=1}^{+\infty} d_i B^{-i}}_{\text{Mantissa}}$

p = esponente della rappresentazione

Lo 0 ha una sua rappresentazione particolare in macchine

Ex. 1,125 in base 10 posso rappresentarlo in diversi modi:

- $10^1 (0,1125)$ → Questa va bene viene chiamata rappresentazione normalizzata
- $10^0 (1,125)$ → Questa non va bene perché la cifra davanti la virgola non è zero
- $10^2 (0,01125)$ → Questa non va bene perché la prima cifra dopo la virgola (d_1) deve essere diversa da 0.

Se ho un numero periodico $0,\bar{9}$

$0,\bar{9} = \sum_{i=1}^{+\infty} 9 \cdot 10^{-1}$ Serie geometrica

$$\sum_{i=1}^{+\infty} 9 \cdot 10^{-1} = 9 \sum_{i=1}^{+\infty} 10^{-1} = \frac{9}{10} \sum_{i=0}^{+\infty} 10^{-i}$$

$$\left.\begin{aligned} \sum_{k=0}^{n} a^k &= S_n \\ \sum_{k=0}^{n+1} a^k &= S_{n+1} \end{aligned}\right\} \quad \begin{aligned} S_{n+1} &= S_n + a^{n+1} \quad \text{2 eq. in 2 incognite} \\ S_{n+1} &= 1 + a S_n \end{aligned}$$

$$S_n = \frac{1 - a^{n+1}}{1 - a}$$

Se $a = 1$ va a $+\infty$

Se $|a| < 1$ $a^{n+1} = 0$ quindi il limite fa $\frac{1}{1-a}$

quindi

$$0,\bar{9} = \frac{9}{10} \sum_{i=0}^{+\infty} 10^{-i} = \frac{9}{10} \cdot \frac{1}{1 - \frac{1}{10}} = \frac{9}{10} \cdot \frac{10}{9} = 1$$

$$0,\bar{9} = 1$$

Posso scriverlo

$10^1 (0,1)$ OK!

$10^0 (0,9999\ldots 9)$ Non va bene perché di è definitivamente uguale a $B-1$

La macchina ha $B = 2$

Se devo rappresentarlo su 32 bit

±	± Esponente	Mantissa
1 bit	8 bit	23 bit

$$p = \pm 1 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 + 1 \cdot 2^4 + 1 \cdot 2^5 + 1 \cdot 2^6 = 2^7 - 1 = 127$$

Esponente $-127 < p < 127$

Posso rappresentare questi numeri

$$\left\{ X = \pm\, 2^{p} \sum_{i=1}^{23} d_i\, 2^{-i} \quad -127 \le p \le 128 \,,\ d_1 \neq 0 \quad 0 < d_i < 1 \right\} \cup \{0\}$$

Insieme dei Numeri di Macchina

$$F(B, t, m, M) = \left\{ x \in \mathbb{R} \,:\, x = \pm B^{p} \sum_{i=1}^{t} d_i B^{-i},\ d_1 \neq 0,\ 0 \le d_i \le B-1,\ -m \le p \le M \right\} \cup \{0\}$$

Proprietà:

① Simmetrico rispetto all'origine

② È un insieme finito

③ Il più piccolo numero di macchina positivo si chiama ω

$$\omega = B^{-m}\left(1 \cdot B^{-1}\right) = B^{-m-1}$$

④ Il più grande numero di macchina si chiama Ω omega

$$\Omega = B^{M}(B-1)\sum_{i=1}^{t} B^{-i}$$

Ex.

$F(2, 3, 2, 2)$



I numeri in questo caso sono più addensati in ω

A noi interessa l'errore relativo cioè $\left|\frac{\tilde{a} - a}{a}\right|$

non quello assoluto $|\tilde{a} - a|$

Cioè ai numeri più grandi possiamo accettare più errori.

Se voglio inserire un numero $x \in \mathbb{R}$ nella macchina, $x > 0$

ho 2 possibilità

- $x \in [\omega, \Omega]$
- $x \notin [\omega, \Omega]$ → overflow (numero troppo grande)
  → underflow (numero troppo piccolo)

master COPY COPISTERIA 0521/983743 328/1212388

A noi interessa $x \in [\omega, \Omega]$

$$x = \pm B^p \sum_{i=1}^{+\infty} d_i B^{-i} \qquad d_1 \neq 0 \qquad 0 \leq d_i \leq B-1$$

↓ La mantissa può anche essere infinita, la macchina opera un troncamento (trn)

$$\tilde{x} = \pm B^p \sum_{i=1}^{t} d_i B^{-i}$$

La macchina memorizza $\tilde{x}$ e si crea quindi un errore

$$\varepsilon_x = \frac{\tilde{x} - x}{x} \quad \text{con } x \neq 0$$

Errore di rappresentazione

TEOREMA:

$$|\varepsilon_x| \leq B^{1-t} = u = \text{precisione di macchina}$$

Il parametro t lo decide la macchina

Dim:

$$x = B^p \sum_{i=1}^{+\infty} d_i B^{-i} \qquad \tilde{x} = B^p \sum_{i=1}^{t} d_i B^{-i}$$

$$|\varepsilon_x| = \left|\frac{\tilde{x} - x}{x}\right| = \frac{x - \tilde{x}}{x} = \frac{B^p \sum_{i=t+1}^{+\infty} d_i B^{-i}}{x}$$

x è sicuramente $\geq B^p B^{-1}$ con $d_i = 1$

$$\frac{B^p \sum_{i=t+1}^{+\infty} d_i B^{-i}}{x} \leq \frac{B^p \sum_{i=t+1}^{+\infty} d_i B^{-i}}{B^p B^{-1}}$$

Il d_i massimo è $B-1$, mentre $\sum_{i=t+1}^{+\infty} B^{-i} = B^{-(t+1)} + B^{-(t+2)} + \dots$

$$= B^{-(t+1)} \left(1 + B^{-1} + B^{-2} + \dots\right), \text{ quindi}$$

$$\frac{B^p \sum_{i=t+1}^{+\infty} d_i B^{-i}}{B^p B^{-1}} \leq \frac{(B-1) B^{-(t+1)} \sum_{i=0}^{+\infty} B^{-i}}{B^{-1}} = \frac{(B-1) B^{-(t+1)} \cdot \frac{1}{1-\frac{1}{B}}}{B^{-1}} =$$