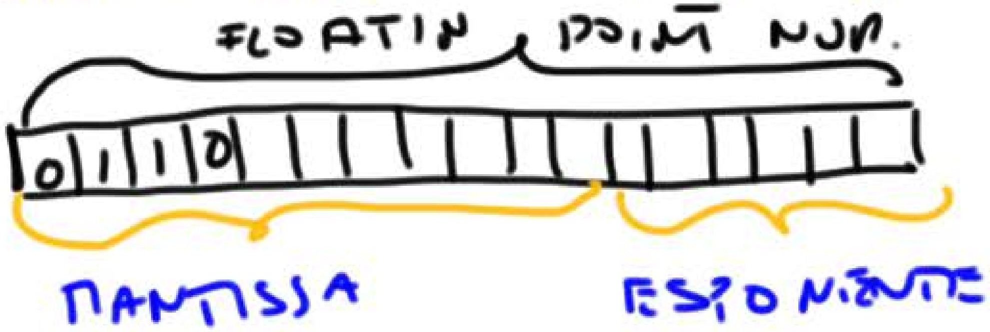
METODO DI EULERO ESPLICITO E FLOATING POINT

QUANDO SI FANNO OPERAZIONI IN ARITMETICA NON ESATTA (FLOATING POINT) SI INTRODUCONO ERRORI



$a\, 2^{b}$ ← ESPONENTE

↑ MANTISSA

Esempio 32 bit 8 esponente 24 mantissa

$$a_0 + a_1 2 + a_2 2^2 + \dots + a_7 2^7 = 0 \dots 255$$

se 1 bit è usato per segno possiamo rapp numeri

-127 a $+128$ ⇒ RANGE $10^{\pm 38}$

mantissa 1 bit segno 23 bit per numero in $[0,1)$

$$\frac{b_1}{2} + \frac{b_2}{2^2} + \dots + \frac{b_{23}}{2^{23}} \Rightarrow \text{num + piccolo } 10^{-7} \text{ circa}$$

CONSEGUENZE RAPPRESENTAZIONE F.P.

a NUM F.P.

b NUM F.P.

$a+b$ IN F.P. SI SCRIVE $a \oplus b \neq a+b$

NUM F.P.   ARITMETICA ESATTA

POCHI NUMERI REALI SONO RAPP. IN F.P.

ESEMPI 0,1 NON HA RAPPRESENTAZIONE ESATTA IN F.P.

0,1	0
0,2	0
0,4	0
0,8	0
1,6	1
0,2	0

HA SERIE INFINITA PERIODICA IN BASE 2

MODELLO PER F.P.

$$a \oplus b = (a+b)(1+\delta) \qquad |\delta| \le \varepsilon$$

$$a \ominus b = (a-b)(1+\delta) \qquad |\delta| \le \varepsilon$$

$$a \otimes b = ab(1+\delta)$$

$$a \oslash b = \frac{a}{b}(1+\delta)$$

$\varepsilon \simeq 10^{-7}$ per singolo precisione (32 bit)

$\varepsilon \simeq 10^{-15}$ per doppia precisione (64 bit)

IN SINTESI

$$\begin{cases} y' = f(x,y) \\ y(0) = y_0 \end{cases} \qquad y_{n+1} = y_n + h f(x_n, y_n) + r_n$$

ERRORE NELLA COMPUTAZIONE DOVUTO AL F.P.
NOTA: r_n È DIFFICILE DA STIMARE. ANZI MOLTO DIFFICILE.

$y(x)$ soluzione esatta

$\varepsilon_n = y(x_n) - y_n$

$$y(x_{n+1}) - y(x_n) - h f(x_n, y(x_n)) = \tau_n = O(h^2)$$

$$y_{n+1} - y_n - h f(x_n, y_n) = r_n$$

f è Lip. in y

$$\varepsilon_{n+1} - \varepsilon_n - h\left[f(x_n, y(x_n)) - f(x_n, y_n)\right] = \tau_n - r_n$$

$$|\varepsilon_{n+1}| \le |\varepsilon_n| + hL|\varepsilon_n| + |\tau_n| + |r_n|$$

$$\varepsilon_{k+1} - \varepsilon_k - h\left[f(x_k, y(x_k)) - f(x_k, y_k)\right] = \tau_k - r_k$$

$$|\varepsilon_{k+1}| \le |\varepsilon_k| + hL|\varepsilon_k| + |\tau_k| + |r_k|$$

$$\le |\varepsilon_k| \underbrace{(1+hL)}_{A} + \underbrace{(|\tau_k| + |r_k|)}_{B \ge}$$

$$B \ge \max|\tau_k| + \max|r_k|$$

STESSA DIMOSTRAZIONE PER EULERO ESPLICITO

SENZA r_k

$$|\varepsilon_k| \le B\,\frac{A^k - 1}{A-1} \le B\,\frac{A^M - 1}{A-1} \le B\,\frac{A^M}{hL} \qquad A^M \le C_2$$

$$B = \max|\tau_k| + \max|r_k| \le C_1 h^2 + \eta$$

Numero piccolo

$$|\varepsilon_k| \le (C_1 h^2 + \eta)\,\frac{C_2}{hL} = \left(\frac{C_1 C_2}{L}\right) h + \left(\frac{C_2 \eta}{L}\right)\frac{1}{h}$$

piccola

CONCLUSIONE

INTRODUCENDO ERRORE DOVUTO A F.P. LA STIMA DELL'ERRORE DIVENTA

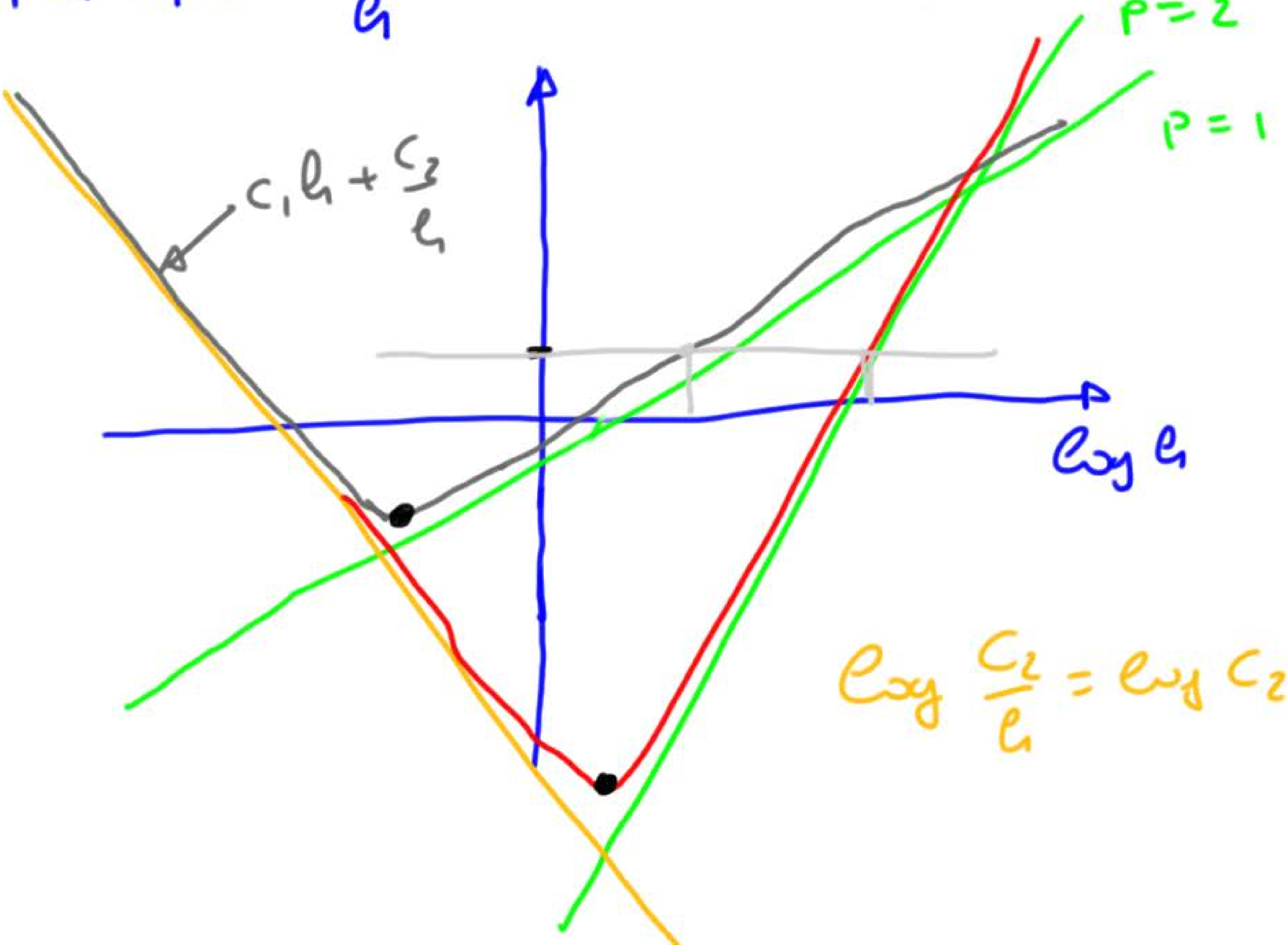
$$|\varepsilon_h| \le C_1 h + \frac{C_2}{h}$$

Numero piccolo per h moderatamente grande. Esplode per h piccolo

PERCHE ORDINE ELEVATO (IN GENERE) E' MEGLIO?

$$\log c_1 h^p = \log c_1 + p(\log h)$$

$$|\varepsilon_n| \le c_1 h^p + \frac{c_2}{h}$$



$$\log \frac{c_2}{h} = \log c_2 - \log h$$

R.K. VA BENE, SI OTTENGONO METODI DI ORDINE ELEVATO I.E. $|E_n| \le c_1 h^p + \frac{c_2}{h}$ p grande

R.K. PERO' COSTANO TANTO

SI POSSONO COSTRUIRE METODI PIU' ECONOMICI CON GLI STESSI ORDINI?

R.K. E' METODO A 1 PASSO CIOE'

$$y_{k+1} = \phi(y_k, h)$$

POSSIAMO USARE INFORMAZIONI DEI PASSI PRECEDENTI PER MIGLIORARE ORDINE

$$y_{k+1} = \phi(y_k, y_{k-1}, \dots, y_{k-e}, h)$$

ESEMPIO (NON FUNZIONANTE)

$$\begin{cases} y' = f(x,y) \\ y(0) = y_a \end{cases} \qquad y'(x_n) = f(x_n, y(x_n))$$

$$y'(x_n) = \frac{y(x_{n+1}) - y(x_n)}{h} + O(h)$$

$$= \frac{y(x_n) - y(x_{n-1})}{h} + O(h)$$

$$= \frac{y(x_{n+1}) - y(x_{n-1})}{2h} + O(h^2)$$

$$y(x_{n+1}) - y(x_{n-1}) = 2h\, f(x_n, y(x_n)) + O(h^3)$$

$$y_{n+1} - y_{n-1} = 2h\, f(x_n, y_n)$$

$$y_{n+1} = \phi(y_n, y_{n-1}, h) = y_{n-1} + 2h\, f(x_n, y_n)$$

$$y_{k+1} = y_{k-1} + 2h f(x_k, y_k)$$

$$y_0 = y_a$$

manca y_1 per partire, di solito si usa un R.K. per calcolare i primi punti e poi si va di multistep

REGIONE DI STABILITA' DEL METODO

$$f(x,y) = \lambda y$$

$$y_{k+1} = y_{k-1} + 2\underbrace{h\lambda}_{s} y_k = 2s\, y_k + y_{k-1}$$

PROBLEMA: RISOLVERE RICORRENZA LINEARE.

SOLUZIONE DI :

$$y_{k+1} = 2s\,y_k + y_{k-1}$$

CERCO SOLUZIONE DELLA FORMA $y_k = z^k$

$$z^{k+1} = 2s\,z^k + z^{k-1} \qquad z^{k-1}\big(\underbrace{z^2 - 2sz - 1}_{=0}\big) = 0$$

$$z_{1,2} = \frac{2s \pm \sqrt{4s^2+4}}{2} = s \pm \sqrt{s^2+1}$$

LA SOLUZIONE GENERALE È DEL TIPO

$$\boxed{y_k = c_1 z_1^k + c_2 z_2^k}$$

$$c_1 z_1^{k+1} + c_2 z_2^{k+1} = 2s\left(c_1 z_1^k + c_2 z_2^k\right) + c_1 z_1^{k-1} + c_2 z_2^{k-1}$$

$$c_1 z_1^{k-1}\big(\underbrace{z_1^2 - 2s z_1 - 1}_{=0}\big) + c_2 z_2^{k-1}\big(\underbrace{z_2^2 - 2s z_2 - 1}_{=0}\big) = 0$$

SOLUZIONE DI :

$$y_{k+1} = 2s\, y_k + y_{k-1}$$

$$y_k = c_1 z_1^k + c_2 z_2^k \qquad z_1 = s + \sqrt{1+s^2} \qquad z_2 = s - \sqrt{1+s^2}$$

LE COSTANTI c_1 e c_2 SI CALCOLANO RISOLVENDO

$$\left.\begin{aligned} y_0 &= c_1 z_1^0 + c_2 z_2^0 = c_1 + c_2 \\ y_1 &= c_1 z_1 + c_2 z_2 \end{aligned}\right\} \Rightarrow c_1 \text{ e } c_2$$

$$|y_k| \to \infty \quad \text{se} \quad |z_1| \text{ o } |z_2| > 1$$

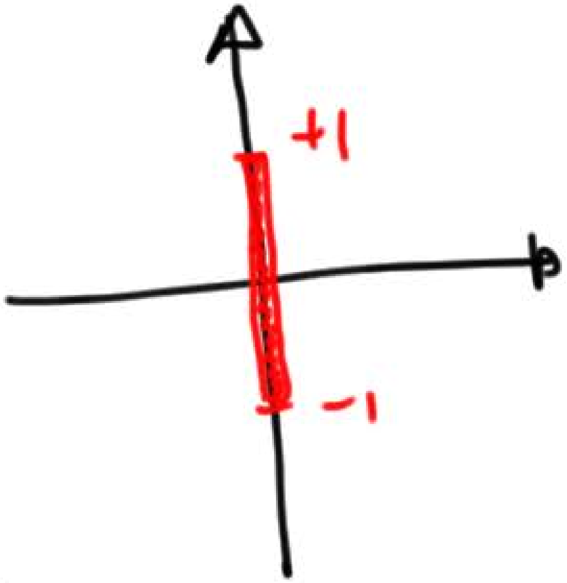
REGIONE DI STABILITA:

punti $s \in \mathbb{C}$ tali che $|z_1|$ e $|z_2| < 1$

cioè

$$|z_1| = |s + \sqrt{1+s^2}| < 1$$

$$|z_2| = |s - \sqrt{1+s^2}| < 1$$



ESERCIZIO: CALCOLARE REGIONE DI STABILITA'