OPERAZIONI SUI NUMERI FINITI

Dati $x, y \in F(\beta, t, L, U)$, non è detto che il risultato di una operazione tra x e y sia un elemento di F.

$$x \circ y = fl(x \bullet y) \quad x, y \in F$$

$$\bullet = \nearrow \begin{matrix} + \\ - \end{matrix} \quad \searrow \begin{matrix} * \\ / \end{matrix}$$

1. **eseguire l'operazione tra x e y**
2. **rappresentare il risultato entro F.**

TEOREMA

Siano $x, y \in F(\beta, t, L, U)$. Allora

$$\frac{|fl(x \bullet y) - x \bullet y|}{|x \bullet y|} \leq k\beta^{1-t}$$

ove $k = 1$ o $1/2$ a seconda che la rappresentazione sia per troncamento o per arrotondamento oppure

$$fl(x \bullet y) = (x \bullet y)(1 + \epsilon) \qquad |\epsilon| \leq k\beta^{1-t}.$$

Somma algebrica

x e $y \in F(\beta, t, L, U)$:

$$\begin{aligned} x &= xm\ \beta^{xe} \\ y &= ym\ \beta^{ye} \end{aligned}$$

$z = zm\ \beta^{ze} = fl(x \pm y)$.

1. Si scala il numero con esponente più basso in modo che gli addendi abbiano lo stesso esponente (quello più grande);
2. si esegue la somma delle mantisse;
3. si considerano le prime t cifre più significative (con troncamento o arrotondamento), ponendole in zm.
4. Si normalizza il risultato aggiustando l'esponente di z in modo che la mantissa zm sia < 1;.

ESEMPI

$\beta = 10, t = 5$, arrotondamento.

- $x = .64937\ 10^7; y = .53726\ 10^4$
  1. Scalatura di y:
     $y = .00053726\ 10^7$;
  2. Somma delle mantisse:
     $.64932 + .00053726 = .64985726$
  3. Arrotondemanto del risultato alla 5 cifra:
     $zm = .64986$
  4. Non necessaria la normalizzazione $ze = 7$

$$z = .64986\ 10^7$$

.

- $x = .64937\ 10^7; y = .53726\ 10^7$
  1. Hanno già lo stesso esponente: non c'è bisogno di scalare gli addendi;
  2. Somma delle mantisse:
     $.64932 + .53726 = 1.18658$
  3. Non è necessario arrotondare
     $zm = 1.18658$
  4. Normalizzazione del risultato: divido la mantissa per 10 e sommo 1 all'esponente
     $zm = .11866$, $ze = 7 + 1$.

$$z = .11866\ 10^8$$

- $x = .75869\ 10^2; y = .75868\ 10^2$
  1. Hanno già lo stesso esponente: non c'è bisogno di scalare gli addendi;
  2. Differenza delle mantisse:
     $.75869 - .75868 = .00001$;
  3. Non è necessario arrotondare
     $zm = .00001$
  4. Normalizzazione del risultato: moltiplico la mantissa per 10^4 e sottraggo 4 all'esponente
     $zm = .1$
  5. $ze = 2 - 4$

$$z = .1\ 10^{-2}$$

.

In questo caso si ha una cancellazione di cifre; succede quando si esegue una differenza tra due quantità circa uguali.Vediamo cosa succede se i dati (x e y) sono affetti da errore.

$$x = fl(.75868531\ 10^2) = .75869\ 10^2$$
$$E_{ax} = 4.69\ 10^{-4} \quad E_{rx} = .6181\ 10^{-5} \leq \tfrac{1}{2}\ 10^{-4}$$

$$y = fl(.75868100\ 10^2) = .75868\ 10^2$$
$$E_{ay} = 1.\ 10^{-4} \quad E_{ry} = .1318\ 10^{-5} \leq \tfrac{1}{2}\ 10^{-4}$$

Il risultato esatto vale $.431\ 10^{-3}$, ma

$$E_a = |.10\ 10^{-2} - .431\ 10^{-3}| = .0569\ 10^{-2}$$

$$E_r = \frac{.569\ 10^{-3}}{.431\ 10^{-3}} \simeq 1.320186$$

La cancellazione determina una amplificazione dell'errore sui dati.

- $x = .62379\ 10^7; y = .32881\ 10^1$
  1. Scalatura di y:
     $y = .00000032881\ 10^7$;
  2. Somma delle mantisse
     $.62379 + .00000032881 = .62379032881$;
  3. Arrotondamento della mantissa del risultato
     $zm = .62379$
  4. Non è necessario normalizzare il risultato $ze = 7$

  $z = .62379\ 10^7$ anche se $y \neq 0$.

Quando $x + y = x$ con $y \neq 0$, si verifica un errore di incolonnamento. Questo capita ogni volta che $|y| \leq \frac{u}{\beta}|x|$. Non esiste un solo elemento neutro per la somma.

PRODOTTO

x e $y \in F(\beta, t, L, U)$:

$$x = xm\ \beta^{xe}$$
$$y = ym\ \beta^{ye}$$

$z = zm\ \beta^{ze} = fl(x \cdot y)$.

1. Si esegue il prodotto $xm \cdot ym$
2. Si esegue troncamento o arrotondamento del risultato a t cifre;
3. Si sommano gli esponenti tenendo conto di una eventuale normalizzazione del risultato.

ESEMPI. $\beta = 10, t = 5$, arrotondamento.

- $x = .11111\ 10^{7}; y = .10202\ 10^{-2}$
  1. Prodotto $.111111 * .10202 = .0113354422$
  2. Arrotondamento $zm = .11335$;
  3. Calcolo dell'esponente con normalizzazione $ze = 7 - 2 - 1 = 4$
     $z = .11335\ 10^{4}$.

QUOZIENTE

x e $y \in F(\beta, t, L, U)$:

$$\begin{aligned} x &= xm\ \beta^{xe} \\ y &= ym\ \beta^{ye} \end{aligned}$$

$z = zm\ \beta^{ze} = fl(x/y)$.

1. Si scala x in modo che $xm < ym$
2. Si esegue xm/ym;
3. Si esegue troncamento o arrotondamento alle t cifre più significative che vengono poste in zm;
4. Calcolo dell'esponente.

ESEMPI. $\beta = 10, t = 5$, arrotondamento.

- $x = .62500\ 10^{0}; y = .12500\ 10^{-2}$
  1. Scalatura di x: $x = .062500\ 10^{1}$
  2. Divisione delle mantisse.$06250/.12500 = .5$;
  3. $ze = 1 + 2 = 3$
     $z = .5\ 10^{3}$.

Osservazioni

Le operazioni tra numeri finiti si riconducono a:

1. **operazioni tra numeri del tipo $(.w_1 w_2 ... w_\tau)$ con $\tau \geq t$;**
2. **moltiplicazioni o divisioni per β^k (k intero);**
3. **somme e sottrazioni di esponenti.**

Le operazioni di tipo 3 sono operazioni tra numeri fixed point.

Le operazioni di tipo 2 comportano scorrimenti verso sinistra o destra di k posizioni.

Le operazioni di tipo 1 sono riconducibili a operazioni tra numeri fixed point. Infatti $.w_1 ... w_\tau = w_1 ... w_\tau \ \beta^{-\tau}$. Pertanto si eseguono operazioni tra numeri fixed point e poi si moltiplica per opportuni fattori di scala, con operazioni di tipo 2.

ESEMPIO.
$.312 * .13 = 312 \ 10^{-3} * 13 \ 10^{-2}$.
Si esegue $312 * 13 = 4056$ **e poi** $4056 \ 10^{-5} = .04056$.

Osservazione importante!

La ridefinizione delle operazioni di macchina comporta la non validità delle proprietà formali.

Dati $x, y \in F$, non è detto che $x \circ y \in F$; infatti può essere che si verifichi OVERFLOW. F non è chiuso rispetto alle operazioni.

1. **Vale la proprietà commutativa per $+$ e per $*$;**
2. **$\exists 0$ tale che $fl(\alpha + 0) = fl(\alpha)$;**
3. **$\exists 1$ tale che $fl(\alpha \cdot 1) = fl(\alpha)$;**
4. **$\forall \alpha, \exists -\alpha$ tale che $fl(\alpha - \alpha) = 0$.**

Ma gli elementi neutri rispetto a somma e prodotto e l'opposto di un numero rispetto alla somma non sono unici.

NON VALGONO:

1. **associativa per il prodotto e la somma;**
2. **distributiva;**
3. **legge di annullamento del prodotto.**

ESEMPI

$\beta = 10, t = 7$, arrotondamento.

$$x = .1234567\ 10^0;\ y = .6666325\ 10^4;\ z = -.6666325\ 10^4$$

. Facciamo vedere che non vale l'associativa della somma: $fl((x+y)+z) \neq fl(fl(x+y)$

1. $fl(fl(x+y)+z) = .123\ 10^0$.

$$\begin{aligned} fl(x+y) &= fl((.6666325 + .00001234567)\ 10^4) = .6666448\ 10^4 \\ fl(fl(x+y)+z) &= fl((.6666448 - .6666325)\ 10^4) = .123\ 10^0 \end{aligned}$$

SI HA CANCELLAZIONE SU DATI PERTURBATI.

2. $fl(x+(y+z)) = .1234567\ 10^0$

$$\begin{aligned} fl(y+z) &= 0 \\ fl(x+fl(y+z)) &= .1234567\ 10^0 \end{aligned}$$

IN QUESTO CASO LA CANCELLAZIONE NON DA PROBLEMI.

Osservazione importante:

L'errore commesso nel calcolo di un'espressione dipende da come questa viene calcolata, o meglio dall'algoritmo usato per calcolarla.

ALGORITMO 1	ALGORITMO 2
$s \leftarrow x + y$	$s \leftarrow y + z$
$s \leftarrow s + z$	$s \leftarrow s + x$

ESEMPIO

$\beta = 10, t = 2$, **troncamento.**
$x = .91\ 10^1; y = .92\ 10^1; z = .10\ 10^0$.
Facciamo vedere la non validità della distributiva della moltiplicazione rispetto alla somma, cioè

$$fl(x \cdot fl(y + z)) \neq fl(fl(xy) + fl(xz))$$

Infatti:

1. $fl(y + z) = fl((.92 + .010)\ 10^1) = .93\ 10^1$
   $fl(x\ fl(y + z)) = fl(.91\ 10^1 * .93\ 10^1) = .84\ 10^2$

2. $fl(xy) = fl(0.8372\ 10^2) = .83\ 10^2$
   $fl(xz) = .91\ 10^0$

$$\begin{aligned} fl(fl(xy) + fl(xz)) &= fl(.83\ 10^3 + .91\ 10^0) \\ &= fl((.83 + .0091)\ 10^2) \\ &= .83\ 10^2 \end{aligned}$$

ESEMPIO

$\beta = 10, t = 7, L = -50, U = 49.$
$x = .2\ 10^{-27}; y = .1\ 10^{-26}; z = .2\ 10^{-9}$

$$fl\left(\frac{z}{xy}\right) \neq fl\left(\frac{z}{x}\frac{1}{y}\right)$$

1. $fl(xy) = fl(.2\ 10^{-52}) = 0$ **UNDERFLOW** $\Longrightarrow$ **non vale la legge di annullamento del prodotto**
   $fl(z/fl(xy))$ **non calcolabile.**

2. $fl(z/x) = 1.0\ 10^{18} = .1\ 10^{19}$
   $fl(1/y) = .1\ 10^{28}$
   $fl(fl(z/x) * fl(1/y)) = .1\ 10^{46}$

Poichè gli ERRORI DI ARROTONDAMENTO nelle operazioni capitano potenzialmente in ogni operazione, ogni risultato intermedio ne può essere influenzato. L'accumulo degli errori è detto PROPAGAZIONE DEGLI ERRORI DI ARROTONDAMENTO.

ESEMPIO

$\beta = 10,\ t = 5$, arrotondamento

Si vuole calcolare $(x - y)/z$ dove

$$\begin{array}{ccccccccc} x & = & .554617 & & y & = & .554601 & z & = & .1\ 10^{-n} \\ & \downarrow & & & & \downarrow & & & & \\ fl(x) & = & .55462 & & fl(y) & = & .55460 & & & \end{array}$$

Il risultato esatto dell'espressione è $.16\ 10^{-4+n}$.

$fl(x - y) = .00002 = .2\ 10^{-4}$,

$E_a = |.16\ 10{-}4 - .2\ 10^{-4}| = .04\ 10^{-4}$.

$fl(fl(x - y)/z) = .2\ 10^{-4+n}$, $E_a = .04\ 10^{-4+n}$

L'errore assoluto è amplificato di 10^n.

CAUSE DI ERRORE

Calcolo di un'espressione

$$y = \varphi(x)$$

φ ↗ **RAZIONALE (operazioni algebriche, radici, potenze)**

φ ↘ **NON RAZIONALE (logaritmi, esponenziali, funzioni trigonometriche**

↓

approssimazione razionale, p.es. serie di Taylor

Restringiamoci al caso in cui $\varphi(x)$ è una funzione razionale. In questo caso ci sono due cause di errore.

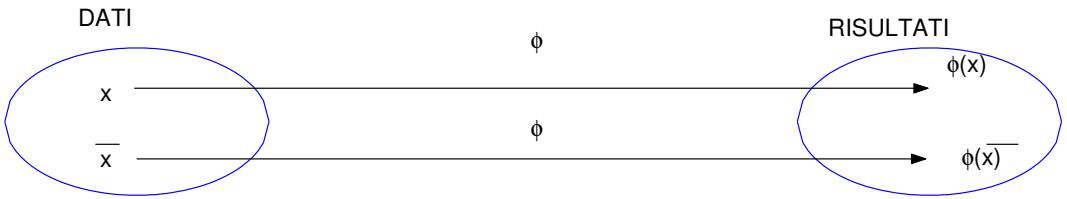
1. **Errore nella rappresentazione dei dati iniziali**

2. **Errore nella rappresentazione del risultato delle operazioni**

1. ERRORE INERENTE O SUI DATI INIZIALI

.

Dati perturbati
Operazioni esatte



Si studia $|\varphi(\bar{x}) - \varphi(x)|$ in relazione a $|\bar{x} - x|$.

$$E_{dati} = |\varphi(\bar{x}) - \varphi(x)|$$

Se $\epsilon_x = \frac{|\bar{x}-x|}{|x|}$ errore relativo sui dati iniziali

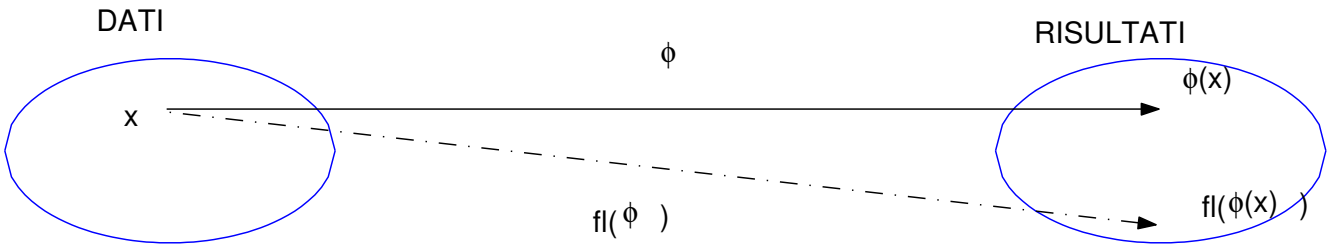
$$\epsilon_{dati} = \frac{|\varphi(\bar{x}) - \varphi(x)|}{|\varphi(x)|}$$

errore relativo sui risultati.
Se ϵ_{dati} è molto grande rispetto a ϵ_x si ha MAL CONDIZIONAMENTO: piccole perturbazioni sui dati iniziali provocano grosse perturbazione sui risultati finali (PROBLEMA MAL CONDIZIONATO)
Il mal condizionamento dipende dal problema, ossia da φ e non dal modo in cui φ è calcolato.

2. ERRORE NELLE OPERAZIONI DI MACCHINA O ERRORE DI ARROTONDAMENTO

.

Dati esatti
Operazioni con errori



Si studia $|fl(\varphi(x)) - \varphi(x)|$ in relazione alla precisione di macchina.

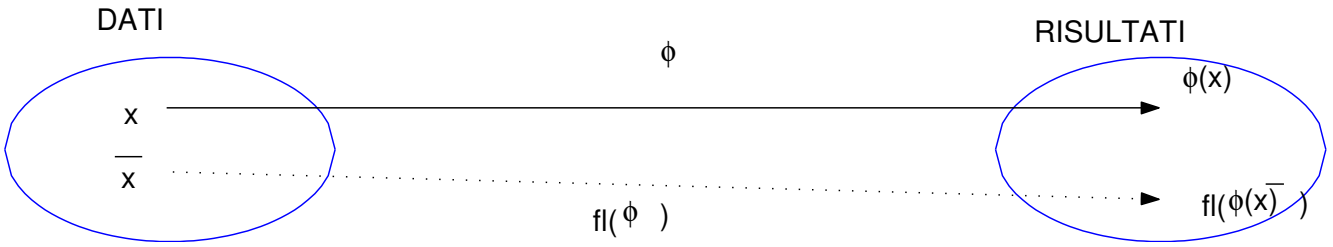
$$E_{alg} = |fl(\varphi(x)) - \varphi(x)|.$$

Analisi di stabilità: un algoritmo è stabile se non è troppo sensibile agli effetti degli errori di arrotondamento. La stabilità dipende dall'ordine con cui sono eseguite le operazioni di macchina, quindi dipende dall'algoritmo.

Infine occorre studiare l'effetto combinato dell'errore sui dati iniziali e dell'aritmetica inesatta.Si studia

$$E_{tot} = |fl(\varphi(\bar{x})) - \varphi(x)|$$

in relazione a $|x - \bar{x}|$ e all'aritmetica finita.



Parametri introdotti per l'analisi dell'errore

$$E_{dati} = \varphi(\bar{x}) - \varphi(x)$$

Errore assoluto sui dati iniziali

$$E_{alg} = fl(\varphi(x)) - \varphi(x)$$

Errore assoluto algoritmico

$$E_{tot} = fl(\varphi(\bar{x})) - \varphi(x) = E_{alg} + E_{dati}$$

Errore assoluto totale

$$\epsilon_{dati} = \frac{\varphi(\tilde{x}) - \varphi(x)}{\varphi(x)}$$

Errore relativo sui dati iniziali

$$\epsilon_{alg} = \frac{fl(\varphi(x)) - \varphi(x)}{\varphi(x)}$$

Errore relativo algoritmico

$$\epsilon_{tot} = \frac{fl(\varphi(\bar{x})) - \varphi(x)}{\varphi(x)} = \epsilon_{dati} + \epsilon_{alg}$$

Errore relativo totale

$$
\begin{aligned}
\epsilon_{tot} &= \frac{fl(\varphi(\tilde{x}))-\varphi(\tilde{x})+\varphi(\tilde{x})-\varphi(x)}{\varphi(x)} \\
&= \frac{\varphi(\tilde{x})-\varphi(x)}{\varphi(x)} + \frac{fl(\varphi(\tilde{x}))-\varphi(\tilde{x})}{\varphi(\tilde{x})}\left(\frac{\varphi(\tilde{x})-\varphi(x)+\varphi(x)}{\varphi(x)}\right) \\
&= \epsilon_{dati} + \epsilon_{alg}\left(1+\epsilon_{dati}\right) = \\
&\simeq \epsilon_{dati} + \epsilon_{alg}
\end{aligned}
$$

In una analisi del I ordine, $\epsilon_{dati}\epsilon_{alg}$ è trascurato.