Avviso

I contenuti di queste annotazioni non sono esaustivi ai fini del buon esito dell'esame. Fare riferimento alle lezioni tenute in aula ed ai testi consigliati:

- G. Monegato, Fondamenti di Calcolo Numerico. Ed. CLUT
- A. Quarteroni, F. Saleri, 'Calcolo Scientifico, esercizi e problemi risolti con Matlab e Octave' Springer.

Obiettivi dell'Analisi Numerica

Trovare **algoritmi** che risolvono un problema matematico:

- **nel minor tempo possibile**
- **con la massima accuratezza possibile**

"L'analisi numerica è l'arte di dare risposta numerica ad un problema matematico *ben-posto* mediante un calcolatore automatico digitale."

Un problema è **ben posto** se **esiste un'unica soluzione** e questa dipende un modo **continuo** dai dati del problema.

Risoluzione numerica di un modello

- Modello Matematico: rappresentazione matematica di un problema reale.
- Metodo Numerico: tecnica per calcolare quantità matematiche complesse.
- Algoritmo Numerico: Sequenza di istruzioni che realizzi il metodo numerico.
- Implementazione dell'algoritmo numerico: traduzione in un linguaggio per una determinata architettura hardware/software.
- Interpretazione **corretta** dei risultati.

Analisi Numerica e Calcolo Scientifico

- Analisi Numerica: Studio matematico di **metodi numerici** (tecnicamente non è necessario il calcolatore).
- Calcolo Scientifico: **Ricerca di Algoritmi** che siano affidabili, flessibili, portabili.

Questa distinzione non è rigorosa: L'analista numerico si occupa anche della scrittura degli algoritmi.

Calcolo Numerico e Simbolico

- Calcolo Simbolico: $(a+b)^2 = a^2 + 2ab + b^2$.
- Calcolo Numerico: $2 + 3.34 = 5.34$.
  Nel calcolo numerico si effettuano approssimazioni:

  $$1/3 = 0,3333\dots$$

  non si possono scrivere tutte le cifre!

Metodo numerico ed algoritmi

- **metodo numerico**: descrizione **matematica** dei calcoli da effettuare per risolvere il problema matematico.
- **algoritmo**: sequenza precisa di azioni (tipicamente istruzioni per il calcolatore) per ottenere il risultato.

Ad ogni metodo numerico, possono corrispondere più algoritmi che lo implementano.

Gli Errori

Una scienza si dice esatta quando...
si conoscono gli errori!

Sorgenti di errore:

- Semplificazioni del modello
- Errori nei dati
- Errori di troncamento del metodo numerico
- Errori di approssimazione nei dati e nei calcoli

Nomenclatura degli Errori

Sia $\tilde{x}$ il valore vero di una certa quantità e x il suo valore affetto da errore:

- Errore assoluto: $\varepsilon_a = |x - \tilde{x}|$.
- Errore relativo: $\varepsilon_r = \frac{|x-\tilde{x}|}{|\tilde{x}|}$.
- Errore relativo percentuale: $\varepsilon_{\%} = 100\,\frac{|x-\tilde{x}|}{|\tilde{x}|}$.

il risultato di un calcolo si scrive:

$$x \pm \varepsilon_a \Leftrightarrow x - \varepsilon_a \leq \tilde{x} \leq x + \varepsilon_a$$

Propagazione degli errori

In seguito ad operazioni di calcolo, gli errori tendono ad accumularsi.

- Somma e sottrazione: l'errore assoluto del risultato è la somma degli errori assoluti degli addendi:

$$C = A \pm B \Longrightarrow \varepsilon_{a,C} = \varepsilon_{a,A} + \varepsilon_{a,B}$$

- Moltiplicazione (e divisione): l'errore relativo del risultato è la somma degli errori relativi dei fattori:

$$C = A\,B \Longrightarrow \varepsilon_{r,C} = \varepsilon_{r,A} + \varepsilon_{r,B}$$

Sistemi di numerazione

Sistema posizionale: ad ogni cifra si assegna un peso a seconda della posizione. Il peso è specificato dalla **base** del sistema che si utilizza, e dall'esponente, che corrisponde alla posizione della cifra. Il valore del numero è dato dalla somma delle cifre pesate.
Nella base 10 scrivere il numero 23.75, significa:

$$2 \cdot 10^1 + 3 \cdot 10^0 + 7 \cdot 10^{-1} + 5 \cdot 10^{-2}$$

Lo stesso numero nella base 2 (binaria) si scrive 10111.11, ossia:

$$1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2}$$

Altre basi usate in ambito informatico sono la 8 (ottale) e 16 (esadecimale).

I numeri nel calcolatore

Comunemente i numeri sono memorizzati nel calcolatore secondo la base binaria. Nella pratica si utilizzano due schemi di memorizzazione in base al “tipo” di numero da memorizzare:

- I numeri interi sono memorizzati come multipli di byte.
- I numeri reali (floating-point), secondo un opportuno schema.

I numeri nel calcolatore 2

Un numero reale $a \in R$, può essere trascritto come:

$$a = m \times N^q,$$

dove

- N è la base scelta per la rappresentazione
- m la mantissa (o frazione)
- q l'esponente.

Nel caso dei numeri reali, N^q è un fattore moltiplicativo ed m “contiene l'informazione principale” del numero reale a.
Se $N^{-1} \leq |m| < 1$ la rappresentazione si dice normalizzata.

Rappresentazione in virgola mobile

$$\pm 0.mmmmmmmmm... \times N^{qqqq..}$$

- Numeri di cifre mantissa ed esponente fissati
- IEEE 754: floating point in base 2 (1985, 7 anni per decidere!)
- IEEE 854: floating point in base arbitraria (1987) http://standards.ieee.org/findstds/standard/854-1987.html
- IEEE 754-2008: ha incluso i due precedenti standard http://standards.ieee.org/findstds/standard/754-2008.html

Approssimazione dei numeri

Esistono due tecniche di approssimazione dei numeri:

1. Troncamento: si escludono i bits a destra dell'ultimo bit di m.
2. Arrotondamento: si aggiunge $\frac{1}{2}N^{-bit_m}$ alla mantissa e poi si esegue il troncamento.

bit_m è il numero di bit che sono messi a disposizione per memorizzare la mantissa dei numeri floating-point.
I numeri di mantissa rappresentabili sono separati della quantità N^{-bit_m} (granularità della mantissa).

Approssimazione dei numeri - Esempio

Supponiamo di avere un 'computer esempio' con 2 byte per la mantissa. Vediamo come viene memorizzato il numero

$$a = 0.1$$

La rappresentazione di a in cifre binarie è la seguente:

$$a = 0.00011001100110011001100\ldots$$

si tratta di un numero periodico. Il primo passo è la normalizzazione:

$$a = 0.11001100110011001100\ldots \times 2^{-3}$$

Approssimazione dei numeri - Esempio

poi considerare i primi 16 bit della parte decimale, ed effettuare un troncamento:

$$\bar{a} = 0.1100110011001100 \times 2^{-3}$$

Questo è il numero arrotondato che è memorizzato nel calcolatore.
Se ora lo rappresentiamo nuovamente nella base decimale si ha:

$$\bar{a} = 0.09999847412109\ldots$$

Approssimazione dei numeri - Esempio

Il numero differisce, seppur di poco, da quello originale.
Vediamo l'effetto sull'errore assoluto ε_a:

$$|a - \bar{a}| = 1.5258789\ldots 10^{-6} \leq 2^{-3-16} = 1.9073486\ldots 10^{-6}$$

Spaziatura dei numeri macchina

$$\bar{a} = 0.1100110011001100 \times 2^{-3}$$

Il numero macchina successivo a questo si ottiene aggiungendo un bit all'ultimo della mantissa:

$$\bar{a}_1 = 0.1100110011001101 \times 2^{-3}$$

che in rappresentazione decimale corrisponde a:

$$\bar{a}_1 = 0.1000003814697\ldots$$

Come si vede il numero 0.1 non è rappresentabile sul calcolatore e si ha:

$$\bar{a} < 0.1 < \bar{a}_1$$

Spaziatura dei numeri macchina 2

Nella rappresentazione floating point la spaziatura dei numeri macchina non è uniforme, perchè dipende anche dall'esponente.
Se $a = \pm 0.mmmmmmmmmmmm \times N^{qqqqq}$, la spaziatura nell'intorno di a è $N^{-bit_m} \times N^{qqqqq}$.
Ad esempio, in binario quando la mantissa passa da 0.111111111 a 0.100000000 la spaziatura raddoppia.

Operazioni macchina

- In generale la proprietà associativa non vale:

$$(a \oplus b) \oplus c \neq a \oplus (b \oplus c)$$

- la proprietà commutativa continua a valere:

$$a \oplus b = b \oplus a$$

dove il simbolo $\oplus$ indica la somma effettuata al calcolatore.
Si sottolinea che questo non dipende da “errori di calcolo”

nella somma bit a bit, ma sempre dal fatto che si ha a disposizione un numero limitato di cifre binarie che costringe all'arrotondamento.

Operazioni macchina - Esempio

Si vuole eseguire la somma $0.1 + 0.00001$. Il primo addendo lo abbiamo già scritto, il secondo ($b = 1 \times 10^{-5}$) è:

$$b = 0.1010011111000101101011 \ldots \times 2^{-16}$$

che sul 'computer esempio' è approssimato da (troncamento):

$$\bar{b} = 0.1010011111000101 \times 2^{-16}$$

che corrisponde a $\bar{b} = 0.99984 \ldots 10^{-5}$. Per effettuare la somma dobbiamo riportare (operazione di *allineamento*) questo numero allo stesso esponente di $\bar{a}$, per cui:

$$\bar{b} = 0.00000000000001010011111000101 \times 2^{-3}$$

Operazioni macchina - Esempio

purtroppo abbiamo a disposizione solo 16 bit per la mantissa e si deve fare una seconda approssimazione.

$$\bar{\bar{b}} = 0.0000000000000101 \times 2^{-3}$$

che in decimale corrisponde a $\bar{\bar{b}} = 0.95367 \ldots 10^{-5}$. Si capisce che c'è un ulteriore peggioramento nella rappresentazione di b.
Il risultato finale della somma sarà

$$\bar{a} + \bar{\bar{b}} = 0.1000080108 \ldots$$

L'effetto del secondo arrotondamento è tanto più marcato quanto più è grande il rapporto tra gli addendi della somma.

Precisione di macchina

Questi effetti degli arrotondamenti si possono formalizzare con il concetto di *precisione macchina*.
La **precisione macchina** è il più grande numero $\mathbf{u}$ in virgola mobile tale che:

$$1 \oplus x = 1 \quad \text{se} \quad x \leq \mathbf{u}$$

Il simbolo $\oplus$ specifica che l'operazione di addizione è stata effettuata sul calcolatore.

Precisione di macchina 2

Si ha

$$1 \oplus \mathbf{u} = 1$$

Conseguenze sulla proprietà associativa:

$$(1 \oplus \mathbf{u}) \oplus \mathbf{u} = 1$$

mentre

$$1 \oplus (\mathbf{u} \oplus \mathbf{u}) = 1 \oplus 2\mathbf{u} \neq 1$$

Precisione di macchina 3

La precisione macchina viene anche chiamata **precisione relativa**: Fissa un **ordine di grandezza** oltre il quale i numeri più piccoli non sono sentiti nelle operazioni di somma algebrica.

$$|\bar{b}| \leq |\bar{a}|\ \mathbf{u} \Longrightarrow \bar{a} \oplus \bar{b} = \bar{a}$$

Stima a priori della precisione macchina:

$$\mathbf{u} = \frac{1}{2} N^{-bit_m + 1}.$$

Nota: le moderne FPU sono provviste dei *guard digits* oltre la mantissa, per eseguire più correttamente le operazioni.

Cancellazione sottrattiva

Perdita di cifre significative quando si sottraggono due numeri quasi uguali.
Esempio in base decimale con 7 digit di mantissa:

$$a = 0.123456789 \qquad b = 0.123456666$$

$$a - b = 0.123 \times 10^{-6}$$

$$\bar{a} = 0.1234567 \qquad \bar{b} = 0.1234566$$

$$\bar{a} - \bar{b} = 0.1 \times 10^{-6}$$

che significa 1 sola cifra di accuratezza.
L'accuratezza delle 7 cifre iniziali è andata perduta.

Cancellazione sottrattiva 2

Vediamo con un altro esempio come si può superare l'inconveniente in un caso specifico. Calcolare

$$x = a - \sqrt{a^2 - b}$$

se $|b| \ll |a|$, x è la differenza di due numeri quasi uguali: c'è cancellazione sottrattiva (se $a > 0$). Per evitare il problema:

$$x = a - \sqrt{a^2 - b}\ \frac{a + \sqrt{a^2 - b}}{a + \sqrt{a^2 - b}} = \frac{b}{a + \sqrt{a^2 - b}}$$

Cancellazione

La cancellazione si verifica particolarmente in:

- moltissime somme di quantità ”piccole” (rispetto al totale).
- somme i cui addendi sono molto differenti in ordine di grandezza.
- molte somme i cui addendi hanno segno variabile.