LABORATORIO DI CALCOLO

L'analisi numerica è la disciplina che si occupa di risolvere problemi scientifici mediante il calcolo numerico. Questa disciplina non ha status inferiore ai calcoli analitici e, in generale, consente di avere informazioni su un sistema che, altrimenti, sarebbero completamente inaccessibili.

Come spesso avviene quando si affronta la soluzione di problemi complessi, anche nell'analisi numerica il miglior metodo per risolvere un problema si trova spesso dopo aver trovato la soluzione stessa del problema o, almeno, dopo aver ottenuto informazioni sulle proprietà rilevanti per la sua soluzione.

Così può accadere che vi siano diversi approcci per risolvere numericamente un problema ma non sono tutti equivalenti. Alcuni possono risultare particolarmente efficienti ed altri inaffidabili. Similmente, un algoritmo può risultare il migliore per affrontare un certo tipo di problema e non esserlo per risolvere problemi di tipo diverso. In sostanza, proprio come accade nel calcolo analitico, l'esperienza e la conoscenza di un certo tipo di problema aiuta ad individuare e trovare le tecniche numeriche migliori. Dunque è bene usare la forza bruta della potenza di calcolo di un computer solamente dopo aver ben riflettuto sul metodo migliore per attaccare un problema con metodi numerici.

Analogamente ai risultati di misure sperimentali, il risultato di un calcolo numerico è determinato a meno di una certa incertezza o errore. Escludendo errori di programmazione o di funzionamento di un computer, i tipi di errore risultanti da un calcolo numerico sono sostanzialmente due:

- errori di arrotondamento
- errori di troncamento.

Errori di arrotondamento

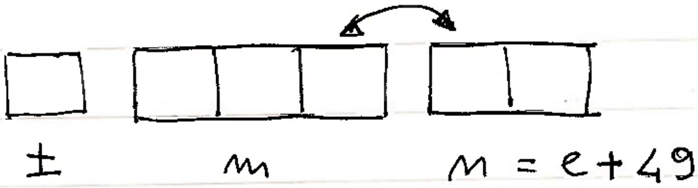
Consideriamo il caso tipico di un calcolo numerico con numeri reali che, in generale, hanno rappresentazione decimale infinita. Un calcolatore tiene in memoria i numeri reali in forma decimale e, quindi, deve evidentemente troncare la serie decimale. La precisione (semplice, doppia, quadrupla) stabilisce quante cifre significative il calcolatore tiene in memoria.

Vediamo più esplicitamente come un numero reale è tenuto nella memoria di un calcolatore. Usiamo inizialmente la base 10:

$$x = \pm\, m \cdot 10^{e}$$

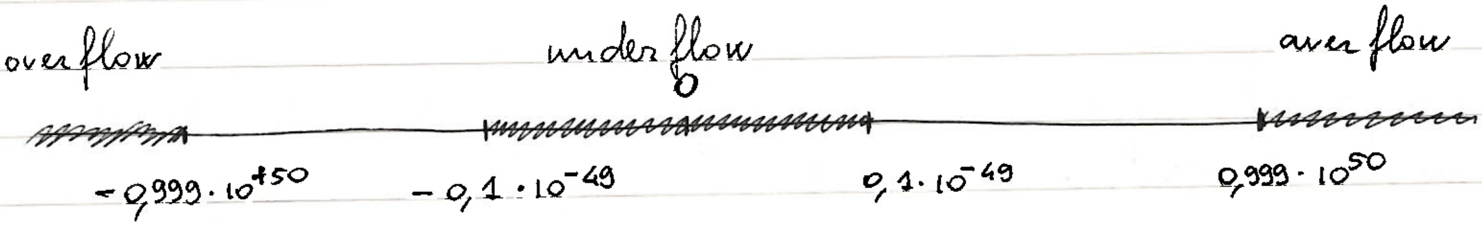
$\pm$ = segno
m = mantissa $\qquad 10^{-1} = 0,1 \leq m < 1 = 10^{0}$
e = esponente

Supponiamo di avere a disposizione 6 cifre per caratterizzare i nostri numeri. Allora potremo distribuire le cifre nel seguente modo:

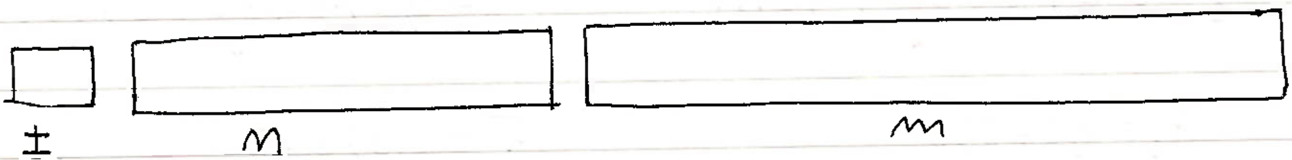


$m : 100 \longrightarrow 999$

$n : 00 \longrightarrow 99 \Rightarrow e : -49 \longrightarrow 50$



I calcolatori funzionano in base 2: 0 e 1 (attivato/non-attivato).



$$x = \pm\, m_2\, 2^{e_2}$$

$\pm = 0, 1$

$n = e_2 + \text{offset} : 00\ldots0 \longrightarrow 111\ldots1$

$1 = 2^0 \leq m_2 < 2^1 = 2 \rightsquigarrow 1,\boxed{\ldots} \qquad m_2 : 00\ldots0 \longrightarrow 111\ldots1$

ESEMPIO. $101{,}11_2 = 1{,}0111 \cdot 2^2 = (-1)^0 (1 + 0{,}0111) \cdot 2^{(129-127)}$ offset

$\boxed{0}\,\boxed{10000001}\,\boxed{011100\ldots0}$

$129 = 128 + 1 = 2^7 + 2^0$

segno	esp	mantissa	
semplice: 1	$8 \rightarrow (10^{-38} - 10^{38})$	23 (7 cifre)	32 bit = 4 byte
doppia: 1	$11 \rightarrow (10^{-308} - 10^{308})$	52 (16 cifre)	64 bit = 8 byte
quadrupla: 1	$15 \rightarrow (10^{-4932} - 10^{4932})$	112 (34 cifre)	128 bit = 16 byte

Quanto detto implica che si deve prestare una certa attenzione quando si effettuano calcoli al computer. Ad esempio:

- sottrazione di due numeri vicini

$x_1 = 0,27832|5743\ldots$ $\qquad$ $x_2 = 0,27845|1327\ldots$

$\bar{x}_1 = 0,27833$ $\qquad$ $\bar{x}_2 = 0,27845$

$$\bar{x}_2 - \bar{x}_1 = 0,00012 \rightsquigarrow 0,12\underbrace{000}_{\text{non significative}} \cdot 10^{-3}$$

$$x_2 - x_1 = 0,12\boxed{5584}\ldots \cdot 10^{-3}$$

Gli effetti degli arrotondamenti si ripercuotono nei passi successivi nel caso di un calcolo un po' più complesso. Ad esempio calcoliamo la seguente funzione

$$f(x) \begin{cases} \to x\left(\sqrt{x+1} - \sqrt{x}\right) \\ \to x \dfrac{1}{\sqrt{x+1}+\sqrt{x}} \end{cases} \quad x = 500 \Rightarrow \quad \begin{matrix} 500\left(22.383 - 22.361\right) = 11,000 \\ \dfrac{500}{22.383+22.361} = 11,175 \end{matrix} \; !$$

$$f(500) = 11,174755\ldots$$

- addizione di numeri di diverso ordine di grandezza

4 cifre $\quad x_1 = 0,1000 \qquad x_2 = 0,3000 \cdot 10^{-5}$

$$x_1 + x_2 = 0,1000 + 0,00003$$

$$= 0,1000|3 \longrightarrow 0,1000$$

$\boxed{\text{ESERCIZIO 1}}$

Discutiamo ora la seconda sorgente di errore nei calcoli numerici.

Errori di troncamento

In generale, a meno di casi particolarmente semplici, un calcolo numerico implica l'approssimazione di una successione infinita di operazioni con una serie ~~in~~finita. Esempi:

- Calcolo di una funzione $\sin(x) = \sum \ldots \to$ somma finita
- Calcolo di un integrale $\int_a^b f(x)dx \to \sum \to$ somma finita

- risoluzione di un'equazione differenziale $\longrightarrow$ rapporti incrementali
- Calcolo $f(x)=0$ $\longrightarrow$ metodi iterativi

In questi casi il troncamento introduce un errore che è importante saper stimare. Un algoritmo più o meno buono si distingue proprio in questo: a parità di sforzo computazionale ϵ è più piccolo

$$x_{vero} = x_{\text{calcolato}} \pm \epsilon$$

[redacted]

Condizionamento e stabilità

Consideriamo il sistema di equazioni differenziali

$$\begin{cases} y_1'(x) = y_2(x) \\ y_2'(x) = y_1(x) \end{cases} \Rightarrow \begin{aligned} y_1(x) &= Ae^{x} + Be^{-x} \\ y_2(x) &= Ae^{x} - Be^{-x} \end{aligned}$$

Dalla condizione iniziale $y_1(0) = 1 = -y_2(0) \Rightarrow A=0$ e $B=1$.

Supponiamo ora di risolvere numericamente il problema, cioè di calcolare $y_1(x)$ e $y_2(x)$ ad una certa serie di punti $x_1, x_2, \dots$ date le condizioni iniziali. Per effetto di piccoli arrotondamenti è come se calcolassimo la soluzione con valori leggermente perturbati rispetto a quelli esatti. L'effetto è che il contributo della parte esponenzialmente crescente può diventare non nullo anche se con coefficiente piccolo. Ora, però, al crescere di x il peso relativo di e^x rispetto a e^{-x} diventa rapidamente dominante ⇒ ERRORE! Dunque anche una piccola deviazione dalla soluzione esatta porta inevitabilmente ad un errore rilevante ad x abbastanza grandi.

Questo è un esempio di un problema intrinsecamente "mal-condizionato": l'unica via d'uscita è quella di usare elevata precisione e non considerare x troppo grandi.

Una quantità che ci può dare indicazioni sul fatto che un problema è potenzialmente "mal-condizionato" è la seguente. Consideriamo un certo algoritmo per il calcolo di una funzione f. Sia poi $\tilde{x}$ vicino a x. Allora

$$\varepsilon(x) = \left| \frac{\tilde{x} - x}{x} \right| \quad \text{errore relativo su } x$$

$$\varepsilon[f(x)] = \left| \frac{f(\tilde{x}) - f(x)}{f(\tilde{x})} \right| \quad \text{errore relativo su } f$$

numero di condizionamento $\quad K = \dfrac{\varepsilon[f(x)]}{\varepsilon(x)} = \left| \dfrac{x f'(x)}{f(x)} \right|$ → grande : mal-condizionato / → piccolo : ben-condizionato

"Risposta quasi giusta per una domanda quasi giusta".

Per l'esempio precedente

$$\left|\frac{x\, y_1'(x)}{y_1(x)}\right| = \left|\frac{x\, y_2(x)}{y_2'(x)}\right| = x^2 \left|\frac{y_2(x)}{x\, y_2'(x)}\right|$$

In sostanza il condizionamento caratterizza la sensibilità della soluzione di un problema a piccoli cambiamenti nei dati. Se tale dipendenza è grande (K grande) il problema è numericamente delicato e "mal-condizionato". Caso contrario se K è piccolo. (Vedi esercizio 2)

Una questione differente è invece la stabilità che è legata al metodo numerico per risolvere un determinato problema. Ad esempio

$$y'(x) = -y(x) \qquad y(0) = 1$$

$$\Downarrow$$

$$y(x) = e^{-x}$$

Una piccola fluttuazione $y(0) = 1+\epsilon \Rightarrow y(x) = (1+\epsilon)e^{-x}$

$$K = \left| x \frac{-e^{-x}}{e^{-x}} \right| = |x| \qquad \text{ben posto}$$

Ci sono però dei metodi numerici di risoluzione di equazioni differenziali che, per il loro funzionamento, introdurrebbero nella soluzione di quel problema anche degli esponenziali crescenti. Questo pone un problema dato che essi potrebbero diventare dominanti rispetto alla.

soluzione corretta ⇒ problema.

Si vede dunque qui che, pur essendo il problema ben-condizionato, il metodo numerico utilizzato può portare a risultati errati. In questo caso si dice che il metodo numerico è instabile (rivedere il calcolo della serie $\frac{1}{n^2}$ dell'esercizio 1).

È importante comprendere bene la differenza tra il condizionamento di un problema e la stabilità di un metodo numerico considerato per risolverlo.
Un problema mal-condizionato può essere risolto accuratamente - ammesso che sia possibile! - solo tramite calcoli di elevata precisione e, comunque, sempre con il rischio di avere risultati errati.
Un problema ben-posto può essere risolto accuratamente con qualunque metodo che sia stabile per quel determinato problema.
Un metodo che è instabile per un certo calcolo potrebbe dare risultati accurati per un certo range di valori ma, inevitabilmente, finirà per dare risultati sbagliati se usato su un range troppo ampio. Lo stesso metodo potrebbe essere invece perfettamente stabile per altri tipi di problemi.

Consigli di programmazione

Un buon codice non è solo ~~un~~ un programma che fa quello che deve in modo efficiente ma anche che, leggendolo, si capisca come lo faccia.
Un punto importante nello sviluppare un buon ~~codice~~ metodo di programmazione è quello di pensare di leggere il codice dopo un mese o, ancora meglio, che qualcun altro legga il vostro codice. Questo è un aspetto ~~ma~~ di rilievo. Infatti nel futuro, nel caso si lavori con calcolatori, vi capiterà di dover mettere mano modificando~~lo~~ e/o ampliando un codice scritto da altri. Se la stesura non è chiara e comprensibile, ~~la~~ il compito può diventare davvero arduo. La chiarezza aiuta anche nella fase di debugging cioè nella fase di controllo che il codice sia corretto e nella ricerca di eventuali errori di programmazione.

- Evitare espressioni criptiche

$j=0,1,2$ $\qquad$ $K=(2-j)(1+3j)/2$ $\qquad$??

$$\left.\begin{array}{l} K=j+1 \\ \text{if } (K=3) \text{ then } K=0 \end{array}\right] \begin{array}{c} \text{permutazione} \\ (0,1,2)\to(1,2,0) \end{array}$$

- Battezzare variabili e strutture con nomi informativi

sum, prod, tot, tmp, ...

- Raggruppare istruzioni simili o correlate

Inizializzazioni, Definizioni,

```
tmp = a
a = b        scambia a <-> b
b = tmp
```

Spezzettare parti di codice omogenei (cicli do)

<u>Commentare</u>!!

- Lavorare in forma modulare (subroutine)

Calcolo della norma di un vettore: è utile definire una subroutine di calcolo

```
norma(vec_in, norm_out)

....

call norma(....)
```

Comandi importanti

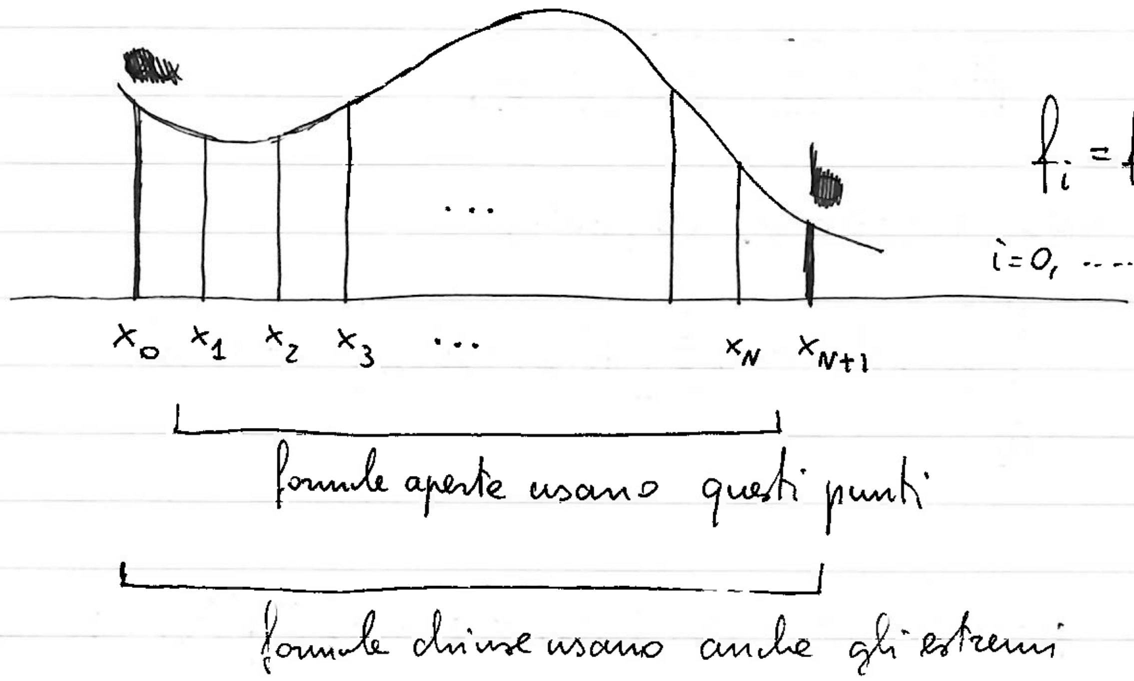
- Cicli : $n = 1, \dots, max$ do-loop; for-loop
- If : progressione condizionata del codice
- Input/output.

INTEGRAZIONE DI FUNZIONI

Ci occupiamo ora di effettuare il calcolo di integrali del tipo

$$\int_a^b f(x)\,dx$$

Supponiamo che a e b siano valori finiti. Più tardi si considererà il caso di limiti di integrazione infiniti.
I metodi di quadratura (cioè deterministici) sono basati in un modo o nell'altro, nell'approssimare l'integrale nella somma opportuna del valore di f calcolato in vari punti all'interno del range $[a,b]$. I metodi di integrazione di Newton-Cotes prendono i punti in cui calcolare f equidistanziati nell'intervallo di integrazione.



$$f_i = f(x_i) \qquad i = 0, \dots, N+1$$

Le formule di Newton-Cotes si distinguono in chiuse o aperte a seconda che i punti estremi di integrazione siano inclusi o no.
Le formule aperte possono essere utili nel caso la funzione

$f(x)$ abbia singolarità integrabili in a oppure b. Similmente $f(a)$ (oppure b) potrebbe assumere un valore finito ma risultante da una forma di indecisione $\frac{0}{0}$ oppure $\frac{\infty}{\infty}$.

I mattoni base delle formule di integrazione sono delle regole per integrare su un piccolo numero di intervalli e poi sommare i vari contributi. All'aumentare di tale numero si possono trovare delle formule che danno il valore esatto dell'integrale per polinomi di ordine progressivamente più elevato.

Nel seguito indichiamo con h la distanza tra due ascisse successive (sono equispaziate!).

$\boxed{\text{FORMULE CHIUSE}}$

REGOLA DEL TRAPEZIO (2 punti)

$$\int_{x_1}^{x_2} f(x)\,dx = h\,\frac{1}{2}\left[f(x_1)+f(x_2)\right] + O(h^3)$$

DIM

Scriviamo $f(x) = f(x_1) + (x-x_1)\,f'(x_1) + \frac{1}{2}(x-x_1)^2 f''(x_1) + \frac{1}{6}(x-x_1)^3 f'''(x_1)$

Integrando $\int f = h\,f(x_1) + \frac{1}{2}h^2 f'(x_1) + O(h^3)$

Poiché $f'(x_1) = f'(x_2) - h\,f''(x_2) + O(h^2)$ riscriviamo

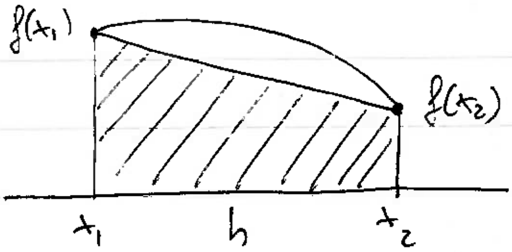
$$\int f = h\, f(x_1) + \frac{1}{2} h^2 f'(x_2) + O(h^3) \qquad (*)$$

Poiché vale anche $f(x) = f(x_2) + (x - x_2) f'(x_2) + \frac{1}{2}(x - x_2)^2 f''(x_2) + \dots$

segue $\int f = h\, f(x_2) - \frac{1}{2} h^2 f'(x_2) + O(h^3) \qquad (**)$

$\blacksquare$

Sommando (*) e (**) si ottiene la formula del trapezio. Graficamente essa si interpreta nel seguente modo:



REGOLA DI SIMPSON (3 punti)

$$\int_{x_1}^{x_3} f(x)\,dx = \frac{h}{3}\left[f(x_1) + 4 f(x_2) + f(x_3) \right] + O(h^5)$$

DIM

Scriviamo

$$f(x) \begin{cases} \to f(x_1) + (x - x_1) f'(x_1) + \frac{1}{2}(x - x_1)^2 f''(x_1) + \frac{1}{6}(x - x_1)^3 f'''(x_1) + \frac{1}{24}(x - x_1)^4 f^{IV}(x_1) + \dots \\ \to f(x_3) + (x - x_3) f'(x_3) + \frac{1}{2}(x - x_3)^2 f''(x_3) + \frac{1}{6}(x - x_3)^3 f'''(x_3) + \frac{1}{24}(x - x_3)^4 f^{IV}(x_3) + \dots \end{cases}$$

Integrando

$$\int f \begin{cases} \to 2h\, f(x_1) + \frac{1}{2}(2h)^2 f'(x_1) + \frac{1}{6}(2h)^3 f''(x_1) + \frac{1}{24}\cdot(2h)^4 f'''(x_1) + O(h^5) \\ \to 2h\, f(x_3) - \frac{1}{2}(2h)^2 f'(x_3) + \frac{1}{6}(2h)^3 f''(x_3) - \frac{1}{24}(2h)^4 f'''(x_3) + O(h^5) \end{cases}$$

Sommando abbiamo

$$2\int f = 2h\left[f(x_1)+f(x_3)\right] + 2h^2\left[f'(x_1)-f'(x_3)\right] + \frac{4}{3}h^3\left[f''(x_1)+f''(x_3)\right] +$$
$$+ \frac{2}{3}h^4\left[f'''(x_1)-f'''(x_3)\right] + \mathcal{O}(h^5)$$

Scriviamo ora

$$f'(x_1) = f'(x_2) - h f''(x_2) + \frac{1}{2}h^2 f'''(x_2) + \mathcal{O}(h^3)$$

$$f'(x_3) = f'(x_2) + h f''(x_2) + \frac{1}{2}h^2 f'''(x_2) + \mathcal{O}(h^3)$$

$$f''(x_1) = f''(x_2) - h f'''(x_2) + \mathcal{O}(h^2)$$

$$f''(x_3) = f''(x_2) + h f'''(x_2) + \mathcal{O}(h^2)$$

$$f'''(x_1) = f'''(x_2) + \mathcal{O}(h)$$

$$f'''(x_3) = f'''(x_2) + \mathcal{O}(h)$$

Sostituendo

$$2\int f = 2h\left[f(x_1)+f(x_3)\right] - 2h^2\left[2h f''(x_2)\right] + \frac{8}{3}h^3 f''(x_2) + \mathcal{O}(h^5)$$
$$= 2h\left[f(x_1)+f(x_3)\right] - \frac{4}{3}h^3 f''(x_2) + \mathcal{O}(h^5) \qquad (*)$$

Vale poi che

$$f(x) = f(x_2) + (x-x_2)f'(x_2) + \frac{1}{2}(x-x_2)^2 f''(x_2) + \frac{1}{6}(x-x_2)^3 f'''(x_2) + \frac{1}{24}(x-x_2)^4 f^{IV}(x_2) + \dots$$

Integrando $\int f = 2h\, f(x_2) + \frac{1}{3} h^3 f''(x_2) + O(h^5)$ $\qquad (**)$

Da $(*)$ e $(**)$ si ha allora

$$6\int f = 2h\left[f(x_1)+f(x_3)\right] \cancel{- \frac{4}{3} h^3 f''(x_2)} + 8h\, f(x_2) \cancel{+ \frac{4}{3} h^3 f''(x_2)} + O(h^5)$$

da cui la formula iniziale.
Usando 4 punti si ottiene la formula di Simpson $\frac{3}{8}$

$$\int_{x_1}^{x_4} f(x)\,dx = \frac{3h}{8}\left[1\, f(x_1) + 3\, f(x_2) + 3\, f(x_3) + 1\, f(x_4)\right] + O(h^5)$$

che, però, è dello stesso ordine della formula di Simpson usuale.

<u>REGOLA DI BODE</u> (5 punti) (BOOLE)

$$\int_{x_1}^{x_5} f(x)\,dx = \frac{h}{45}\left[14 f(x_1) + 64 f(x_2) + 24 f(x_3) + 64 f(x_4) + 14 f(x_5)\right] + O(h^7)$$

Si possono costruire formule analoghe considerando un numero sempre maggiore di punti. Da notare che, però, a partire da 8 punti ci sono pesi negativi e gli errori di arrotondamento possono giocare un ruolo rilevante nell'accuratezza.

FORMULE ESTESE CHIUSE

Vogliamo ora scrivere le formule estese all'intervallo di integrazione $[a = x_1, \, b = x_N]$ a partire da quelle valutate poco sopra. È utile riesprimere la accuratezza in termini di numero N di punti. Infatti

$$h = \frac{b-a}{N-1} \qquad \left(f_i = f(x_i) \right)$$

REGOLA DEL TRAPEZIO ESTESA

$$\int_{x_1}^{x_N} f(x)\,dx = h\left[\frac{1}{2} f_1 + f_2 + f_3 + \ldots + f_{N-1} + \frac{1}{2} f_N\right] + O\left(\frac{1}{N^2}\right)$$

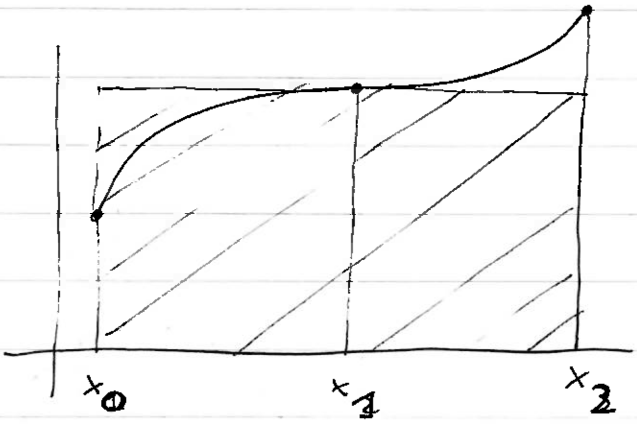
REGOLA DI SIMPSON ESTESA

$$\cancel{\int_{x_1}^{x_N} f(x)\,dx = h\left[\frac{3}{8} f_1 + \frac{7}{6} f_2 + \frac{23}{24} f_3 + f_4 + f_5 + \ldots + f_{N-4} + f_{N-3} + \frac{23}{24} f_{N-2} + \frac{7}{6} f_{N-1} + \frac{3}{8} f_N\right] + O\left(\frac{1}{N^4}\right)}$$

$$\int_{x_1}^{x_N} f(x)\,dx = \frac{h}{3}\left[f_1 + 4 f_2 + 2 f_3 + 4 f_4 + 2 f_5 + \ldots + 2 f_{N-2} + 4 f_{N-1} + f_N\right] + O\left(\frac{1}{N^4}\right)$$

FORMULE APERTE

Le formule aperte non sono molto convenienti per i seguenti motivi. ① Non sono banalmente "impilabili" per costruire delle formule estese ② La loro accuratezza è inferiore alle formule di integrazione gaussiana. Giusto un paio di esempi.



$$f(x) = f(x_1) + (x-x_1) f'(x_1) + \frac{1}{2}(x-x_1)^2 f''(x_1) + \dots$$

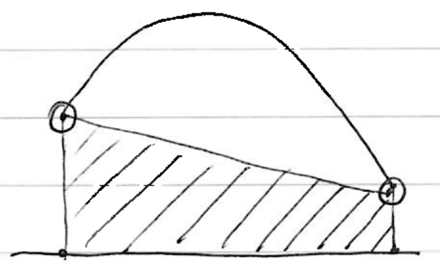
Integrando
$$\int_{x_0}^{x_2} f(x)\,dx = 2h f(x_1) + O(h^3)$$

Oppure
$$\int_{x_0}^{x_5} f(x)\,dx = \frac{h}{24}\left[55 f(x_1) + 5 f(x_2) + 5 f(x_3) + 55 f(x_4)\right] + O(h^5)$$

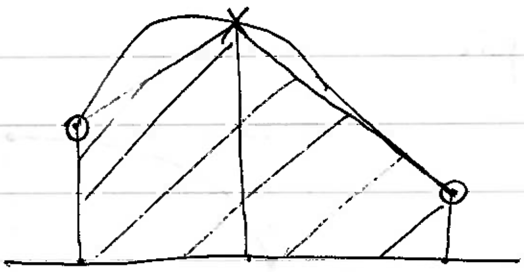
ALGORITMI

Le formule presentate vengono utilizzate in modo ricorsivo riducendo progressivamente il valore di h e verificandone la stabilità. In particolare, la regola del trapezio si presta a tale tipo di operazione e consente di aumentare rapidamente il numero di punti senza dover ricominciare il calcolo ogni volta.

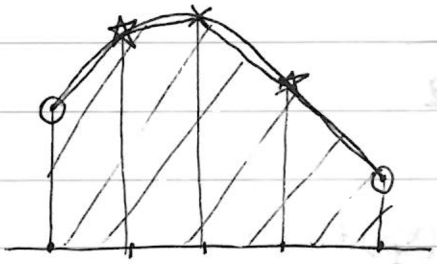
Consideriamo infatti il seguente esempio:



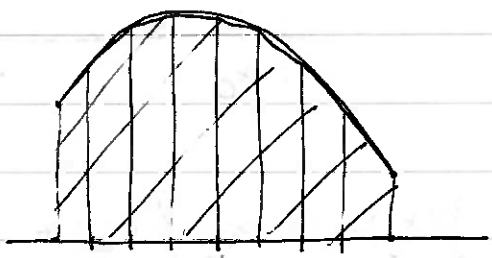
$1 = 2^0$



$2 = 2^1$



$4 = 2^2$



$8 = 2^3$

Sia $J \geq 1$ e dividiamo l'intervallo di integrazione $(b-a)$ in 2^J intervalli. Evidentemente

$$h = \frac{b-a}{2^J} \qquad N = 2^J + 1 \quad \# \text{ punti}$$

Dalla regola del trapezio estesa abbiamo

$$T_J = (2h)\left[\frac{1}{2} f_0 + f_2 + f_4 + \ldots + f_{2(N-1)} + \frac{1}{2} f_{2N}\right]$$

$$T_{J+1} = h\left[\frac{1}{2} f_0 + f_1 + f_2 + \ldots\ldots + f_{2(N-1)} + f_{2N-1} + \frac{1}{2} f_{2N}\right]$$

Segue dunque la regola ricorsiva

$$\boxed{T_{J+1} = \frac{1}{2} T_J + \sum_{K=1}^{N} f(x_{2K-1})}$$

Inoltre si vede immediatamente che la formula di Simpson estesa all'ordine J è data da

$$\boxed{S_{J+1} = \frac{1}{3}\left[4\,T_{J+1} - T_J\right]}$$

Metodo di Romberg

Il metodo di Romberg consiste nell'utilizzare i risultati della formula del trapezio per migliorare l'accuratezza del calcolo. Il metodo è una generalizzazione della formula di ricorrenza : infatti dai valori del calcolo con la regola del trapezio otteniamo il valore con la più accurata formula di Simpson.

L'assunzione della formula di Romberg è che nella formula del trapezio estesa gli errori ~~[illegible]~~ compaiono solo con potenze pari:

$$\int_a^b f(x)\,dx = T(f,h) + a_1 h^2 + a_2 h^4 + a_3 h^6 + \dots$$

o, equivalentemente

$$\int_a^b f(x)\,dx = T(f,N) + \frac{a_1'}{N^2} + \frac{a_2'}{N^4} + \frac{a_3'}{N^6} + \dots$$

Considerando un numero dimezzato di punti ($h \to 2h$):

$$\int_a^b f(x)\,dx = T(f,2h) + 4a_1 h^2 + 16 a_2 h^4 + \dots$$

da cui segue subito che

$$\int_a^b f(x)\,dx = \frac{4\,T(f,2h) - T(f,h)}{3} + b_2 h^4 + b_3 h^6 + \dots$$

che è esattamete la formula di Simpson ~~generali~~ estesa.

L'affermazione può essere generalizzata: Sia

$$Q = R(h, K-1) + c_1 h^{2K} + c_2 h^{2K+2} + \dots$$

$$Q = R(2h, K-1) + c_1 4^K h^{2K} + c_2 4^{K+1} h^{2K+2} + \dots$$

e così

$$Q = \frac{4^K R(h,K-1) - R(2h,K-1)}{4^K - 1} + c_2' h^{2K+2} + \dots$$

e dunque $$R(h,K) = \frac{4^K R(h,K-1) - R(2h,K-1)}{4^K - 1}$$

Ricordando che h è legato a J e, quindi, al numero di punti, definiamo

$$R(J,0) = T(J) \longrightarrow$$ calcolo dell'integrale con la regola del trapezio esteso.

definiamo poi

$$R(J,1) = \frac{4\,R(J,0) - R(J-1,0)}{4-1}$$

$$R(J,2) = \frac{4^2\,R(J,1) - R(J-1,1)}{4^2-1}$$

e, in generale

$$R(J,K) = \frac{4^K\,R(J,K-1) - R(J-1,K-1)}{4^k-1}$$

In modo schematico

J	N punti	$R(J,0)$ (regola trapezio)	$R(J,1)$ (Simpson)	formula $1/N^6$	formula $1/N^8$
0	2	$R(0,0)$			
1	$2+1$	$R(1,0)$	$R(1,1)$		
2	$4+1$	$R(2,0)$	$R(2,1)$	$R(2,2)$	
3	$8+1$	$R(3,0)$	$R(3,1)$	$R(3,2)$	$R(3,3)$

<u>NOTA</u>: usando i metodi discussi si possono fare integrali con limiti infiniti mediante cambi di variabile che ~~[illegible]~~ mappano ad un range finito di integrazione.

Metodo di Gauss

Descriviamo ora il metodo di integrazione di Gauss. In questo metodo i punti in cui si calcola la funzione integranda non sono più equidistanziati. Il metodo di Gauss consente di ottenere una accuratezza elevata nel caso di funzioni abbastanza lisce. Il problema è quello di calcolare integrali del tipo

$$I = \int_a^b W(x) f(x)\, dx$$

dove $W(x)$ è una funzione di peso. Il metodo di Gauss consiste nell'avere

$$I = \sum_{j=1}^{N} w_j f(x_j)$$

dove w_j sono dei coefficienti fissati, indipendenti da f e x_j sono dei punti all'interno di (a,b), anch'essi fissati in modo indipendente da f.

Il valore di w_j e x_j viene fissato richiedendo che, se $f(x)$ è un polinomio di grado uguale o inferiore a $2N-1$, allora il valore dato da I per l'integrale è esatto.

Per l'insieme dei polinomi, definiamo il prodotto scalare tra due polinomi f e g:

$$\langle f | g \rangle = \int_a^b W(x) f(x) g(x)\, dx$$

Due polinomi sono ortogonali se $\langle f|g\rangle = 0$; un polinomio è normalizzato se $\langle f|f\rangle = 1$.
Una base di polinomi ortonormali può essere ottenuta mediante la formula di ricorrenza:

$$p_{-1}(x) = 0 \qquad p_0(x) = 1$$

$$p_{j+1}(x) = (x - a_j)\, p_j(x) - b_j\, p_{j-1}(x) \qquad j = 0, 1, \dots$$

Imponendo che $\langle p_{j+1} | p_j\rangle = 0$ si ha:

$$0 = \langle p_{j+1} | p_j\rangle = \langle (x - a_j) p_j | p_j\rangle - b_j \langle p_j | p_{j-1}\rangle = \langle x p_j | p_j\rangle - a_j \langle p_j | p_j\rangle$$

$$\Downarrow$$

$$\boxed{a_j = \frac{\langle x p_j | p_j\rangle}{\langle p_j | p_j\rangle}}$$

Similmente, da $\langle p_{j+1} | p_{j-1}\rangle = 0$, segue

$$0 = \langle p_{j+1} | p_{j-1}\rangle = \langle (x - a_j) p_j | p_{j-1}\rangle - b_j \langle p_{j-1} | p_{j-1}\rangle =$$

$$= \langle x p_j | p_{j-1}\rangle - b_j \langle p_{j-1} | p_{j-1}\rangle =$$

$$= \langle p_j | x p_{j-1}\rangle - b_j \langle p_{j-1} | p_{j-1}\rangle = \langle p_j | p_j\rangle - b_j \langle p_{j-1} | p_{j-1}\rangle$$

Poiché $x p_{j-1} = p_j + a_{j-1} p_{j-1} + b_{j-1} p_{j-2}$ vale allora

$$\boxed{b_j = \frac{\langle p_j | p_j\rangle}{\langle p_{j-1} | p_{j-1}\rangle}}$$

<u>TEO</u>: gli zeri dei polinomi sono reali, con molteplicità 1 e all'interno dell'intervallo (a,b).

<u>DIM</u>.
Siano $x_1, \dots, x_m$ gli zeri del polinomio $p_n(x)$ (grado n), all'interno di (a,b) in cui esso cambia segno. m può essere zero (nessuno zero) oppure, al massimo, pari a n (numero totale di zeri).

Il prodotto $(x-x_1)\dots(x-x_m)\,p_n(x)$ non cambia mai segno in (a,b). Data l'ortogonalità di $p_n(x)$ con ogni polinomio di grado $<n$, vale che

$$\int_a^b (x-x_1)\dots(x-x_m)\,p_n(x)\,W(x) = 0 \qquad \text{se } m<n$$

Poiché $(x-x_1)\dots(x-x_m)\,p_n(x)$ non cambia mai segno e $W(x) \geq 0$, si ha che $m=n$. Dunque

$$p_n(x) = A(x-x_1)\cdot\ldots\cdot(x-x_n) \qquad \text{con } x_1,\dots,x_n \in (a,b)$$

Si prova poi che ogni zero di p_{n+1} si trova tra due zeri di p_n:

● = zeri di p_n ; × = zeri di p_{n+1}.

Siano ora gli $x_1, \dots, x_N$ gli zeri del polinomio $p_N(x)$, allora i pesi sono dati da

$$\boxed{w_j = \frac{\langle p_{N-1} | p_{N-1} \rangle}{p_{N-1}(x_j)\, p_N'(x_j)}}$$

Infatti, sia $p_N(x) = (x - a_{N-1})\, p_{N-1}(x) - b_{N-1}\, p_{N-2}(x)$.

Vale che $p_N'(x) = p_{N-1}(x) + p_{N-1}'(x)(x - a_{N-1}) - b_{N-1}\, p_{N-2}'(x)$

Poiché p_{N-1}' è di grado $N-2$ e p_{N-2}' di grado $N-3$, si ha

$$\langle p_N'(x) | p_{N-1}(x) \rangle = \langle p_{N-1} | p_{N-1} \rangle N$$

Dalla formula di Gauss:

$$\langle p_N'(x) | p_{N-1}(x) \rangle = \sum_{j=1}^{N} w_j\, p_N'(x_j)\, p_{N-1}(x_j)$$

Segue che

$$w_j = \frac{\langle p_{N-1} | p_{N-1} \rangle}{p_N'(x_j)\, p_{N-1}(x_j)}$$

Un'ultima propietà importante è che la formula di Gauss è, in realtà, esatta per polinomi fino all'ordine $2N-1$. Sia, infatti, f un polinomio di grado $\leq 2N-1$. Allora

$$f = p_N\, q + r$$

dove q e r sono polinomi di grado $\leq N-1$. Dato che p_N è ortogonale a tutti i polinomi di grado $\leq N-1$, si ha:

$$\int_a^b f(x) W(x)\, dx = \int_a^b p_N(x)\, q(x)\, W(x) + \int_a^b r(x) W(x)\, dx = \int_a^b r(x) W(x)\, dx$$

Poiché $r(x)$ è di ordine $\leq N-1$ vale

$$\int_a^b r(x) W(x)\, dx = \sum_{j=1}^{N} r(x_j)\, w_j$$

ed essendo x_j gli zeri di p_N, vale anche

$$f(x_j) = p_N(x_j)\, q(x_j) + r(x_j) = r(x_j)$$ e dunque

$$\int_a^b f(x) W(x)\, dx = \sum_{j=1}^{N} f(x_j)\, w_j$$

è esatta.

Riportiamo ora le funzioni peso $W(x)$, gli intervalli e le formule di ricorrenza per i set più usati di polinomi ortogonali.

LEGENDRE

$$W(x) = 1 \qquad [-1, 1]$$

$$(j+1)\, P_{j+1}(x) = (2j+1)\, x\, P_j(x) - j\, P_{j-1}$$

CHEBYSHEV

$$W(x) = \frac{1}{\sqrt{1-x^2}} \qquad (-1, 1)$$

$$T_{j+1} = 2x\, T_j - T_{j-1}$$

JACOBI

$$W(x) = (1-x)^{\alpha} (1+x)^{\beta} \qquad (-1, 1)$$

$$c_j\, P^{(\alpha,\beta)}_{j+1}(x) = (d_j + e_j x)\, P^{(\alpha,\beta)}_j(x) - f_j\, P^{(\alpha,\beta)}_{j-1}(x)$$

$$c_j = 2(j+1)(j+\alpha+\beta+1)(2j+\alpha+\beta)$$

$$d_j = (2j+\alpha+\beta+1)(\alpha^2-\beta^2)$$

$$e_j = (2j+\alpha+\beta)(2j+\alpha+\beta+1)\cdot(2j+\alpha+\beta+2)$$

$$f_j = 2(j+\alpha)(j+\beta)(2j+\alpha+\beta+2)$$

LAGUERRE

$$W(x) = x^{\alpha} e^{-x} \qquad (0, +\infty)$$

$$(j+1)\, L^{\alpha}_{j+1}(x) = (-x+2j+\alpha+1)\, L^{\alpha}_j(x) - (j+\alpha)\, L^{\alpha}_{j-1}(x)$$

HERMITE

$$W(x) = e^{-x^2} \qquad (-\infty, +\infty)$$

$$H_{j+1}(x) = 2x\, H_j(x) - 2j\, H_{j-1}(x)$$

NOTA: stima dell'errore.
Sappiamo che se f è un polinomio di grado $\leq 2N-1$, allora la formula di Gauss è esatta. Considerando lo sviluppo di Taylor:

$$f(x) = \sum_{m=1}^{2N-1} \frac{x^m}{m!} f^{(m)}(\xi) + \frac{x^{2N}}{(2N)!} f^{(2N)}(\xi) + \dots \qquad \xi \in (a,b)$$

Dunque possiamo scrive

$$\cancel{\int_a^b W f(x)\,dx - \sum_{j=1}^{N} W_j f(x_j)}$$

$$f(x) = \tilde{p}_{2N-1}(x) + \frac{f^{(2N)}(\xi)}{(2N)!} \tilde{p}_{2N}(x) + \dots$$

$$\int W(x) f(x) - \sum w_j f(x_j) = \frac{f^{(2N)}(\xi)}{(2N)!} \langle p_N | p_N \rangle$$

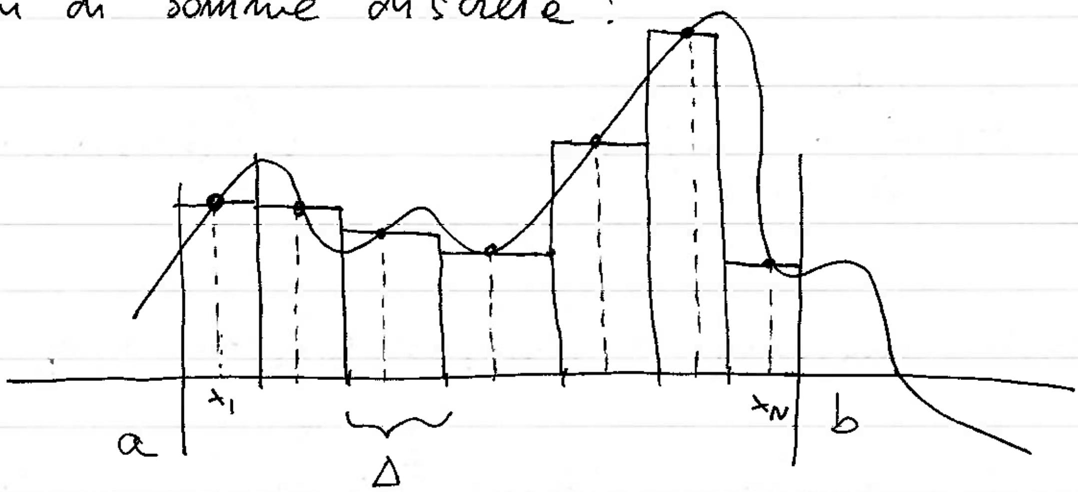
poiché $p_{2N} = \tilde{p}_N \left(\tilde{p}_N + \tilde{p}_{N-1} \right)$ con p_N, p_{N-1}, p_{2N-1} polinomi opportuni.

$\boxed{\text{ESERCIZI } 3 \to 7}$

Metodo Monte Carlo

Il nome evocativo di questo metodo fa già comprendere che il caso gioca un ruolo fondamentale in questo tipo di approccio. E' a prima vista controintuitivo che in un calcolatore, cioè una macchina deterministica, si vogliano utilizzare numeri casuali. Vedremo però la grande utilità di questo metodo.

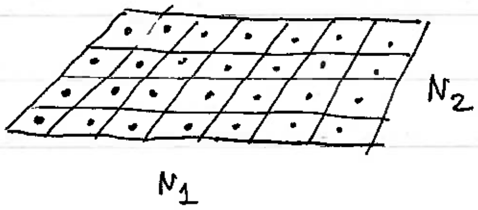
Ricordiamo brevemente l'idea del calcolo di un integrale in termini di somme discrete:



$$A = \sum_{i=1}^{N} \Delta f(x_i) = \frac{b-a}{N} \sum_{i=1}^{N} f(x_i)$$

Generalizzando al caso di M dimensioni, avremo

$$A = \frac{(b_1-a_1)}{N_1} \cdot \ldots \cdot \frac{(b_M-a_M)}{N_2} \sum_{m_1=1}^{N_1} \ldots \sum_{m_m=1}^{N_m} f(x_{m_1}, \ldots x_{m_m})$$



L'idea del metodo Monte Carlo è quella di generare i punti x_i a caso nell'intervallo (a,b) con distribuzione uniforme e la stima per l'integrale è data ancora da

$$A = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i)$$

x_i = random uniformi in (a,b).

Analogamente per l'integrale multidimensionale

$$A = \frac{V^{(M)}}{N} \sum_{i=1}^{N} f(\vec{x}_i) \qquad V^{(M)} = (b_1-a_1) \cdot \ldots \cdot (b_M - a_M)$$

$\vec{x}_i$ = random uniformi in $(a_1,b_1) \times (a_2,b_2) \times \ldots \times (a_M, b_M)$

Definendo la media $\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$, allora

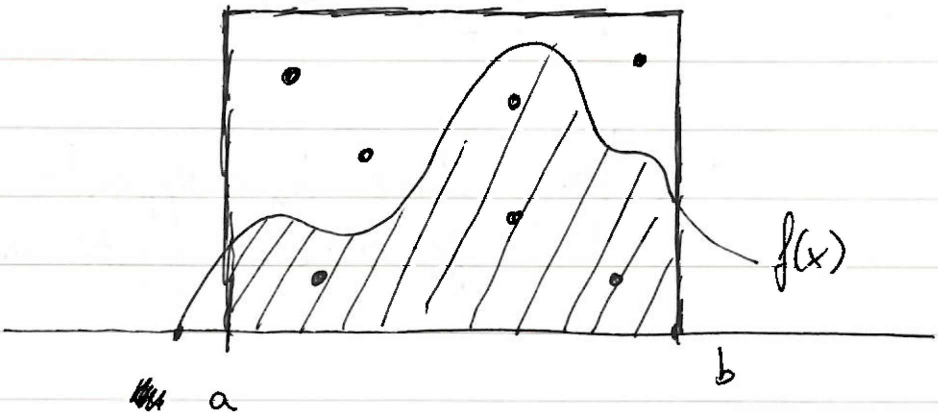
$$A = V^{(M)} \langle f \rangle$$

È importante notare come nel primo caso abbiamo M sommatorie mentre nel caso del metodo Monte Carlo ne abbiamo una sola.

Il metodo Monte Carlo diviene molto più conveniente in termini di tempo di calcolo rispetto ai metodi deterministici che abbiamo considerato quando il numero di variabili di integrazione diviene elevato. Inoltre, come vedremo tra poco, Il metodo Monte Carlo è anche molto semplice ed utile da usare quando il dominio di integrazione non è semplice e regolare. In tale situazione l'integrazione deterministica può diventare estremamente complicata.

~~ESERCIZIO 8~~ $\boxed{\text{ESERCIZIO 8}}$

Un approccio Monte Carlo alternativo è quello detto di hit and miss. In sostanza



si definisce una regione al cui interno è racchiusa l'area da calcolare e di cui è calcolabile l'area. Nel caso sopra il rettangolo R. Si prendono poi a caso con di probabilità uniforme N punti in R e si valuta la frazione N_h di punti che sono appartenenti alla regione di cui si vuole calcolare l'area. Avremo allora che

$$A = R\,\frac{N_h}{N}$$

Notiamo che il metodo di hit & miss è equivalente al metodo di integrazione Monte Carlo descritto poco sopra.

la seguente

Si può definire infatti ~~una~~ funzione di due variabili: ~~costa~~ ~~d~~ nella regione R:

$$g(x,y) = \begin{cases} 0 & \text{se } (x,y) \notin A \\ 1 & \text{se } (x,y) \in A \end{cases} \qquad (x,y) \in R \quad !!$$

Allora

$$\int_R dx\,dy\; g(x,y) = R\,\frac{1}{N}\sum_{i=1}^{N} g(x_i, y_i)$$

$$= R\,\frac{N_h}{N}$$ dove N_h è appunto la frazione di punti che sono in A.

ERRORE

È importante avere una stima dell'errore del metodo Monte Carlo. Infatti questo consente di avere informazioni sulla accuratezza della stima numerica.
Dato che il metodo Monte Carlo si basa sul calcolo di una media

$$A = V^{(M)} \langle f \rangle$$

è naturale calcolare l'errore in termini dell'errore della media. Dati $x_1, \dots, x_N$ valori con media $\langle x \rangle$, l'errore è dato da

$$err(\langle x \rangle) = \frac{\sigma}{\sqrt{N}}$$

essendo σ la deviazione standard

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$$

$$\langle x^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i^2 \quad ; \quad \langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$$

dunque

$$err(A) = \frac{V^{(M)}}{\sqrt{N}} \sqrt{\langle f^2 \rangle - \langle f \rangle^2} = \frac{V^{(M)}}{\sqrt{N}} \sqrt{\langle (f - \langle f \rangle)^2 \rangle}$$

È importante notare che l'accuratezza della stima Monte Carlo aumenta con $\sqrt{N}$, essendo N la statistica raccolta.

Va poi notato che la formula precedente è esatta nel caso che le misure $f(x_i)$ siano distribuite gaussianamente. Nel caso di distribuzioni di tipo differente, $err(A)$ fornisce solo un ordine di grandezza per le fluttuazioni.

Osserviamo, però, che in una tale situazione possiamo sfruttare il seguente importante teorema: il teorema del limite centrale.

<u>TEO</u>: Sia $x_1, x_2, \dots, x_N$ una sequenza di N variabili random indipendenti e con la medesima distribuzione di pobabilità con media μ e ~~varianza~~ deviazione standard $\sigma > 0$. Allora, al crescere di N, la distribuzione della media del campione delle variabili, si approssima ad una distribuzione

gaussiana con media μ e deviazione standard $\sigma/\sqrt{N}$, indipendentemente dalla forma della ~~p~~ distribuzione di probabilità iniziale.

In altre parole, se definiamo la nuova variabile random:

$$S^{(N)} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

allora la distribuzione di pobabilità di $S^{(N)}$ si approssima ad una gaussiana per $N \to \infty$.

Per la dimostrazione, vedi foglio allegato.

Così se anche $f(x_i)$ non sono distribuiti secondo una gaussiana,

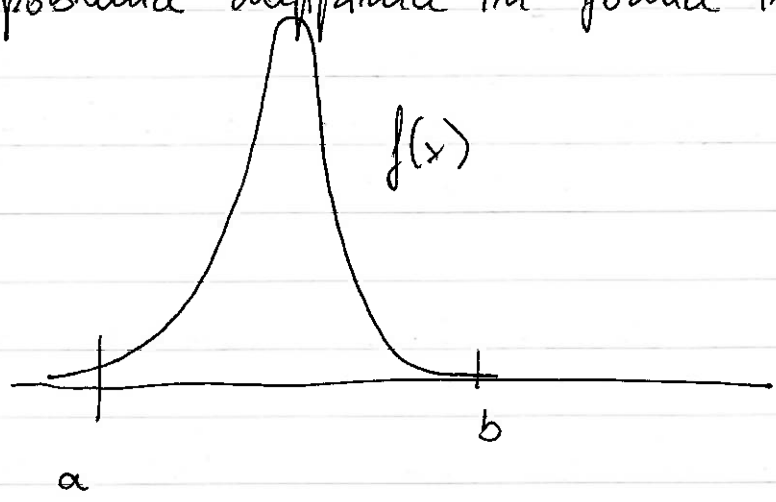
$$F^{(m)} = \sum_{i=1}^{M} f(x_i)$$

ha una distribuzione che, per $m \to \infty$, diviene gaussiana e, dunque, fornisce una corretta stima dell'errore sull'integrale.

$\boxed{\text{ESERCIZIO 9}}$, $\boxed{\text{ESERCIZIO 10}}$

Importance sampling

L'errore nell'integrazione Monte Carlo è $\frac{\sigma}{\sqrt{N}}$: è possibile ridurlo aumentando la statistica N oppure riducendo σ.
L'importance sampling ha proprio questo secondo obiettivo.
Vediamo il problema dapprima in forma intuitiva:



Generare in modo uniforme i punti in (a,b) non è il modo più efficiente di fare il calcolo dato che il maggior contributo proviene dai punti intorno al picco. L'integrazione Monte Carlo ha il suo massimo di efficienza per una distribuzione che ~~[illegible]~~ sia il più possibile uniforme.

Così se troviamo una funzione $g(x) > 0$, normalizzata, che nell'intervallo (a,b) è tale che

$$\frac{f(x)}{g(x)}$$

è più piatta, abbiamo raggiunto l'obiettivo di migliorare l'efficienza del nostro Monte Carlo.

Cioè:

$$I = \int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{g(x)}\,g(x)\,dx = \int_a^b \frac{f(x)}{g(x)}\,dG(x)$$

essendo $G(x) = \int_a^x g(y)\,dy$. Con cambio di variabile $z = G(x)$

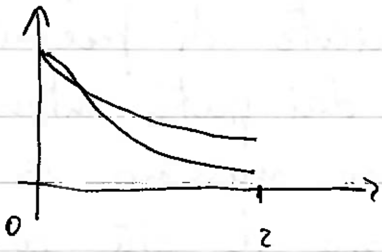
$$I = \int_{G(a)}^{G(b)} \frac{f(G^{-1}(z))}{g(G^{-1}(z))}\,dz = \int_0^1 \frac{f(G^{-1}(z))}{g(G^{-1}(z))}\,dz \simeq \frac{1}{N}\sum_{i=1}^{N} \frac{f(G^{-1}(z_i))}{g(G^{-1}(z_i))}$$

I valori di z sono distribuiti in modo uniforme in $(0,1)$.

Il problema è quindi ricondotto al calcolo di G^{-1}. L'alternativa è quella di generare punti secondo la distribuzione $g(x)$.

ESEMPIO 1

$$I = \int_0^2 e^{-x^2}\,dx$$



Poniamo $f(x) = e^{-x^2}$ e $g(x) = a e^{-x}$. Allora

$$G(x) = a\int_0^x e^{-y}\,dy = -a e^{-y}\Big|_0^x = a(1 - e^{-x})$$

Normalizzando $\Rightarrow 1 = a(1 - e^{-2}) \Rightarrow a = \frac{1}{1 - e^{-2}}$

Poniamo $z = \frac{1 - e^{-x}}{1 - e^{-2}} \Rightarrow e^{-x} = 1 - (1 - e^{-2})z$

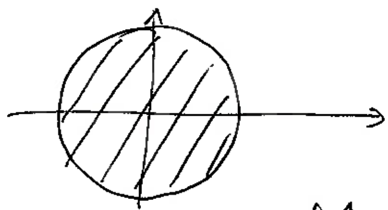
$$x = -\ln\left[1 - (1 - e^{-2})z\right]$$

E quindi

$$I = (1-e^{-2}) \int_0^1 \frac{f(x)}{e^{-x}}\, dz \qquad \text{dove} \quad x = -\ln\left[1-(1-e^{-2})z\right]$$

ESEMPIO 2

$$f(x) = \frac{1}{\sqrt{\pi}}\, e^{-x^2} \qquad (-\infty, +\infty)$$



$$g(x,y) = \frac{1}{\pi}\, e^{-(x^2+y^2)}$$

$$G(r) = \int \frac{1}{\pi}\, e^{-(x^2+y^2)}\, \Theta\left(-\frac{x^2+y^2}{r^2}+1\right) dx\, dy \qquad \Theta(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$$

$$= \int_0^r 2R\, e^{-R^2}\, dR = 1 - e^{-r^2}$$

Poniamo allora $z = G(r) = 1 - e^{-r^2}$ con z random in $(0,1)$.

$z' = 1-z \Rightarrow$ random in $(0,1)$ $\quad 1-z' = 1-e^{-r^2}$

$$r = \sqrt{-\log(z')} \Rightarrow x = r\cos(2\pi\alpha)$$

$\alpha \in (0,1)$ random

$$\boxed{\text{ESERCIZI } 11-13}$$

EQUAZIONI DIFFERENZIALI

La soluzione di problemi di equazioni differenziali ordinarie può essere sempre ridotto allo studio di un sistema di equazioni differenziali di primo ordine. Infatti, se abbiamo ad esempio

$$\frac{d^2y}{dx^2} + q(x)\frac{dy}{dx} = r(x)$$

possiamo riscrivere

$$\begin{cases} \dfrac{dy}{dx} = z(x) \\[2ex] \dfrac{dz}{dx} + q(x)z(x) = r(x) \end{cases}$$

Più in generale avremo

$$\frac{dy_i(x)}{dx} = f_i\left(x\,;\, y_1, \dots, y_N\right) \qquad i = 1, \dots, N$$

Noi ci occuperemo poi del caso di equazioni differenziali con valori iniziali $y_i(0) = a_i$. L'idea della risoluzione ~~dell~~ di equazioni differenziali ordinarie si basa sostanzialmente sul metodo di Eulero. Tale metodo corrisponde a discretizzare la derivata e poi calcolare la modifica delle funzioni y_i passo passo al variare di x.

Indichiamo con h lo step elementare per x e vogliamo risolvere

$$\frac{dy}{dx} = f(x,y)$$

Discretizzando e posto $y(x_0)=y_0$ e $x_n = x_0 + nh$

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

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

Questa è la formula del metodo di Eulero. Questa è una formula al primo ordine in h e, in generale, costituisce una discretizzazione troppo grossolana.

Un'approssimazione migliore si ha nel seguente modo:

$$y\left(x+\frac{h}{2}\right) = y(x) + \frac{h}{2}\frac{dy}{dx}(x) + \frac{1}{2}\left(\frac{h}{2}\right)^2\frac{d^2y}{dx^2}(x) + \frac{1}{6}\left(\frac{h}{2}\right)^3\frac{d^3y}{dx^3} + \dots$$

$$y\left(x-\frac{h}{2}\right) = y(x) - \frac{h}{2}\frac{dy}{dx}(x) + \frac{1}{2}\left(\frac{h}{2}\right)^2\frac{d^2y}{dx^2}(x) - \frac{1}{6}\left(\frac{h}{2}\right)^3\frac{d^3y}{dx^3} + \dots$$

da cui

$$y\left(x+\frac{h}{2}\right) - y\left(x-\frac{h}{2}\right) = h\frac{dy}{dx}(x) + O(h^3)$$

$x-\frac{h}{2} \to x_n$

$x+\frac{h}{2} \to x_{n+1}$ $\Rightarrow$ $y_{n+1} = y_n + h f\left(x_n + \frac{h}{2}, y_n + k_1\right) + O(h^3)$

$k_1 = \frac{h}{2} f(x_n, y_n)$

Posto
$$K_1 = h f(x_n, y_n)$$
$$K_2 = h f\left(x_n + \frac{h}{2}, \ y_n + \frac{K_1}{2}\right)$$
$$y_{n+1} = y_n + K_2 + O(h^3)$$

Questa è la formula di Runge-Kutta al secondo ordine.

Posto infatti

$x_n \to x$
$$y(x+h) = y(x) + h\frac{dy}{dx}(x) + \frac{1}{2}h^2 \frac{d^2y}{dx^2}(x) + O(h^3)$$

$y_n \to y(x)$
$$= y(x) + h f(x, y(x)) + \frac{1}{2}h^2 \frac{d f(x,y(x))}{dx^2} + O(h^3)$$

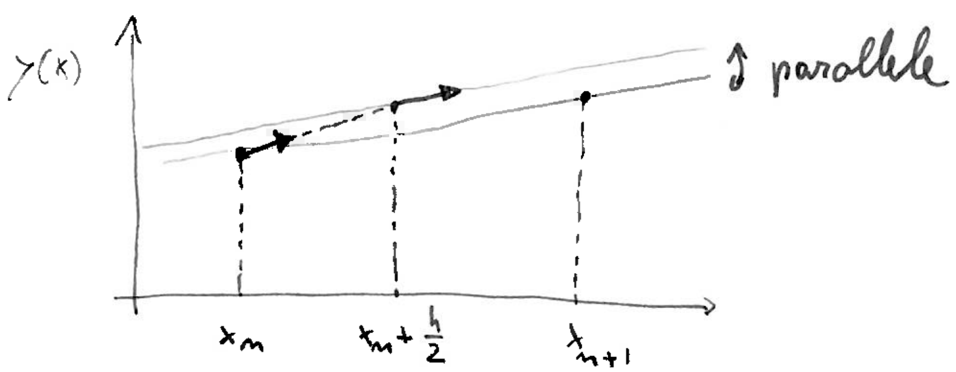
$y_{n+1} \to y(x+h)$

D'altra parte

$$f\left(x + \frac{h}{2}, y(x) + \frac{1}{2}h f(x,y(x))\right) = f(x,y(x)) + \frac{h}{2}\frac{\partial f}{\partial x}(x,y) + \frac{1}{2}h f(x,y(x))\frac{\partial f}{\partial y}$$

Poiché $f(x,y(x)) = \frac{dy}{dx}$ si ha $= f(x, y(x)) + \frac{h}{2}\frac{d}{dx}\left(f(x,y(x))\right)$.

Confrontando con la precedente si verifica l'identità a meno di termini di ordine $O(h^3)$. In termini geometrici la formula al secondo ordine si può così interpretare



La formula di gran lunga più utilizzata è, però, quella di Runge-Kutta al quarto ordine.

$$K_1 = h f(x_n, y_n)$$

$$K_2 = h f\left(x_n + \frac{h}{2}, y_n + \frac{K_1}{2}\right)$$

$$K_3 = h f\left(x_n + \frac{h}{2}, y_n + \frac{K_2}{2}\right)$$

$$K_4 = h f\left(x_n + h, y_n + K_3\right)$$

$$y_{n+1} = y_n + \frac{1}{6}\left(K_1 + 2K_2 + 2K_3 + K_4\right) + \mathcal{O}(h^5)$$

Questa formula è, in generale, migliore di quella al secondo ordine. "Migliore" nel senso che un passo h doppio è possibile rispetto a quella di ordine due e l'accuratezza è più elevata. Non è però sempre automatico che "ordine superiore" corrisponde a "accuratezza superiore".

Metodo del leap-frog

Il metodo del leapfrog è un metodo di integrazione del secondo ordine abbastanza utilizzato nello studio di sistemi dinamici. L'integrazione leap-frog corrisponde a calcolare posizioni e velocità ad intervalli di tempo alternati. Per esempio, se la posizione è nota a tempi interi, la velocità è nota a tempi interi più mezzo step.

Così ad esempio

$$x_{n+1} = x_n + v_{n+\frac{1}{2}}\, dt$$

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + a_n\, dt$$

dove a_n sono le accelerazioni.

~~Passo~~ Controllo dello step

In generale, il metodo di Runge-Kutta è equipaggiato anche ~~di~~ di un metodo di controllo della taglia del passo di integrazione. In regioni con variazioni molto limitate si può scegliere un passo abbastanza grande senza perdere accuratezza e, allo stesso tempo, non sprecare tempo macchina con un passo ~~di~~ inutilmente piccolo. In altre regioni, si possono avere delle variazioni tali da richiedere una riduzione del passo per non avere una perdita sensibile nell'accuratezza.

Dato che nel nostro uso non avremo necessità di accuratezze molto elevate non discuteremo l'implementazione di questa modifica. È bene comunque ~~sapere~~ esserne al corrente nel caso, in futuro, si debba presentare la necessità.

Più semplicemente noi faremo simulazioni a passo progressivamente più piccolo fino a che la differenza tra un passo di 2h e due passi di h sia sempre contenuta al di sotto di un certo valore di soglia:

$$y(x+2h) = y_1 + (2h)^5 \phi + O(h^6)$$

$$y(x+2h) = y_2 + 2(h^5) \phi + O(h^6)$$

allora $\Delta = |y_2 - y_1|$

e richiediamo che $\Delta \leqslant$ soglia.

ZERI DI UNA FUNZIONE

Il calcolo degli zeri di una funzione corrisponde a trovare le soluzioni di

$$f(x) = 0$$

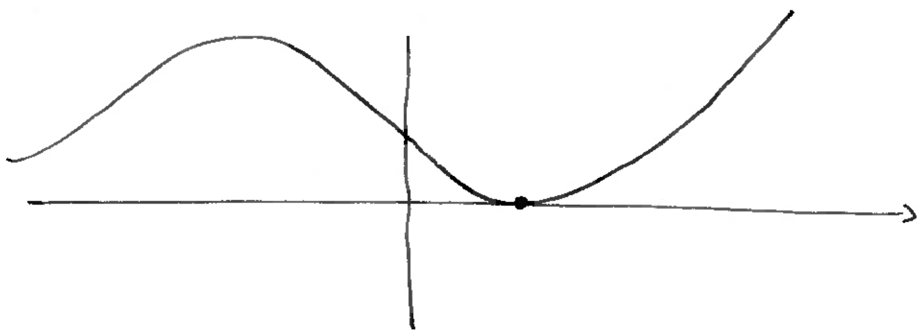
Noi tratteremo solo il caso unidimensionale; il problema in più dimensioni è molto più complesso. In generale, i metodi sono di tipo iterativo per approssimazioni successive. Si possono poi verificare situazioni particolari che rendono l'individuazione degli zeri non banale. Ad esempio nel caso di radici multiple e, ancor di più, multiple con molteplicità pari. In ogni caso il successo di un qualunque metodo di individuazione degli zeri si avvantaggia molto anche solo di una conoscenza approssimata del comportamento della funzione.

Metodo del bracketing e bisezione

Questo metodo si basa sul fatto di trovare due punti a e b in cui la funzione assume valori di segno opposto. Se la funzione è continua ⇒ ci deve essere almeno uno zero. Altrimenti la funzione è discontinua; se è limitata allora ci sarà un punto con un salto. L'idea del metodo è quella di valutare la funzione f nel punto intermedio tra a e b e decidere in quale dei due sottointervalli si trovi lo zero. La procedura viene iterata fino a individuare lo zero con una accuratezza richiesta ϵ. Si vede, però, che nel caso di una funzione del tipo

$$f(x) = \frac{1}{x-c}$$

Il metodo individuerà $x=c$ come uno zero. La sostituzione in $f(x)$ mostrerà però che si tratta di un falso zero. ~~Un~~ Come detto sopra, una situazione in cui questo metodo si rivela anche inefficace è del tipo



Metodo di Newton-Raphson

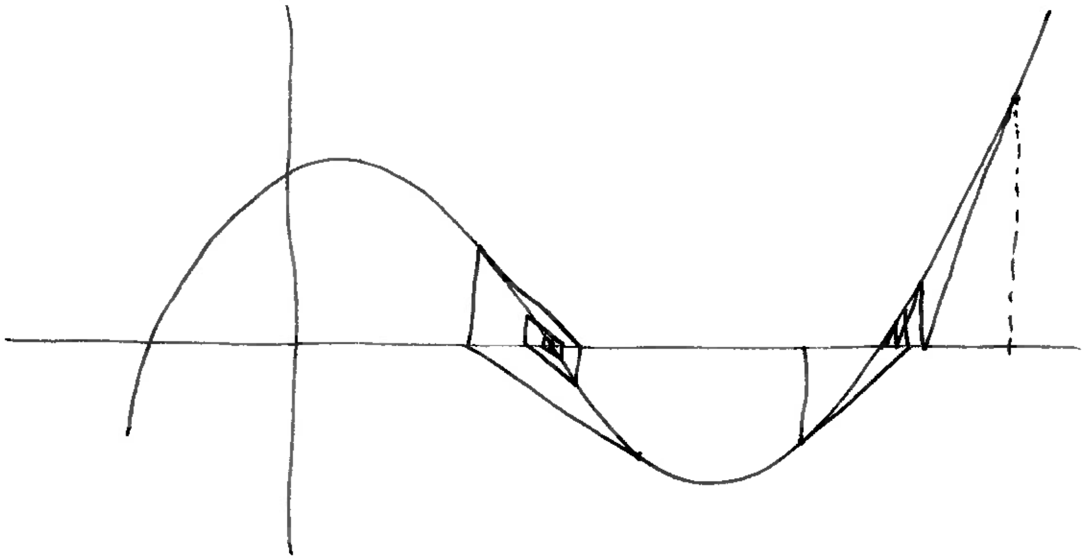
Un metodo molto noto e diffuso è quello di Newton-Raphson. Esso richiede però di conoscere anche la derivata prima della funzione f: Se x è un punto vicino allo zero, allora possiamo scrivere

$$f(x+\delta) = f(x) + \delta f'(x) + \dots$$

Così per avere $f(x+\delta)=0$ una buona scelta è:

$$\delta = -\frac{f(x)}{f'(x)}$$

Il metodo è anch'esso iterativo. Il nuovo punto diviene $x+\delta$ e così via. Il metodo ha anche una semplice interpretazione geometrica:



Ci sono però dei casi in cui il metodo fallisce. Principalmente sono i due seguenti: