

**DOMANDE E RISPOSTE  
SUI SISTEMI ARITMETICI A PRECISIONE FINITA  
E  
CONDIZIONAMENTO**

**1) FORNIRE UN ESEMPIO DI PROBLEMA BEN CONDIZIONATO E MAL CONDIZIONATO:**

Un problema ben condizionato potrebbe essere la risoluzione di un sistema triangolare superiore, invece uno mal condizionato le metriche di Hilbert e Vandermonde.

**2) E' POSSIBILE RENDERE BEN CONDIZIONATO UN PROBLEMA MAL CONDIZIONATO E PERCHE'?**

No perchè il condizionamento dipende dalla natura del problema stesso, indipendentemente dall'algoritmo usato.

**3) COSA SI INTENDE DIRE CHE IL SISTEMA ARITMETICO E' CARATTERIZZATO DALLA MASSIMA ACCURATEZZA DINAMICA?**

F sistema aritmetico f.p ha massima accuratezza dinamica se

presi  $x, y$  appartenente a  $F \rightarrow x \neq y \Rightarrow fl(x \# y) = fl(x) \# fl(y)$  e quindi le rappresentazioni coincidono perfettamente.

Quindi esso garantisce che il risultato di qualsiasi operazione floating point differisca dal risultato dell'operazione in  $R$  corrispondente (il risultato esatto), di una quantità che è il solo errore di rappresentazione di  $r$ , e viene perciò detto sistema a massima accuratezza dinamica.

**Il risultato  $r'$  dell'operazione f.p. che è un numero macchina, è esattamente la rappresentazione f.p.n. (floating point normalizzato) a precisione finita del risultato  $r$  della corrispondente operazione in  $R$  (reali).** Un tale sistema aritmetico garantisce dunque che il risultato di qualsiasi operazione f.p. differisca dal risultato dell'operazione in  $R$  corrispondente (il risultato esatto), di una quantità che è il solo errore di rappresentazione di  $r$ , e viene perciò detto sistema a massima rappresentazione dinamica.

**4) FORNIRE UN ESEMPIO DI ALGORITMO INSTABILE ED UNA TECNICA CON CUI POTERLO RENDERE STABILE:**

Un esempio di algoritmo instabile è l'algoritmo di Gauss che con la tecnica del pivoting diventa praticamente stabile.

**5) ILLUSTRARE BREVEMENTE IL SIGNIFICATO DI CIASCUNA GRANDEZZA, IN UN SISTEMA ARITMETICO A PRECISIONE FINITA:**

$\epsilon$  ossia l'epsilon macchina, in un sistema aritmetico f.p è il più piccolo numero appartenente ad  $F$  tale che:

$$1 + \epsilon = fl(1 + \epsilon) > 1$$

$\mu$  La massima accuratezza relativa in un sistema f.p è il massimo errore che si commette nella rappresentazione di un numero  $x$ :

$$u = \max \frac{|fl(x) - x|}{|x|} = \frac{1}{2} \beta^{1-t} \text{ nel caso dell'arrotondamento}$$

$$u = \max \frac{|fl(x) - x|}{|x|} = \beta^{1-t} \text{ nel caso del troncamento}$$

**6) SI CONFRONTINO I RISULTATI. SI SPIEGHI QUAL E' LA PROPRIETA' DELLA SOMMA CHE NON VALE PIU' LAVORANDO IN UN SISTEMA ARITMETICO A PRECISIONE FINITA:**

$(x (+) y) (+) z \neq x (+) (y (+) z)$  in f.p.

La proprietà associativa per la somma non vale per l'errore di Roundoff sull'operazione aritmetica.

**7) SI OSSERVI IN PARTICOLARE QUALE FENOMENO INCIDE SULLA PERDITA DELLE CIFRE SIGNIFICATIVE:**

Sulle cifre significative incide il fenomeno della cancellazione che è la manifestazione del mal condizionamento che si manifesta con la sottrazione di due numeri vicini.

$$\text{Dim: } C(f, x, y) = \frac{|x| + |y|}{|x - y|} \rightarrow 0 \quad \text{se } |x - y| \rightarrow 0$$

**8) SPIEGARE LA DIFFERENZA TRA PROBLEMA BEN CONDIZIONATO E MAL CONDIZIONATO:**

Un problema si dice ben condizionato se l'errore relativo (assoluto) nella soluzione ha al più lo stesso ordine di grandezza dell'errore relativo (assoluto) nei dati. Mentre un problema in cui l'errore relativo nella soluzione ha ordine di grandezza maggiore rispetto all'errore relativo nei dati si dice mal condizionato.

**9) SPIEGARE LA DIFFERENZA TRA ALGORITMO STABILE E INSTABILE:**

Un algoritmo si dice instabile, se gli errori di roundoff introdotti nei dati si propagano amplificandosi in maniera tale che i risultati siano inaccettabili.

**CHE COS'E' L'INDICE DI CONDIZIONAMENTO?**

Detto  $\delta$  l'errore nei dati e  $\bar{\sigma}$  l'errore corrispondente nella soluzione, e posto

$$\bar{\sigma} = \mu \cdot \delta$$

$\mu$  è detto indice di condizionamento del problema, ed inoltre risulta che:

se  $\mu \leq 1$  il problema è ben condizionato.

se  $\mu > 1$  il problema è mal condizionato.

**10) CHE COS'E' L'ERRORE DI ROUND OFF?**

Cosiderato un sistema aritmetico f.p a precisione finita F, sia x un numero reale appartenente all'insieme di rappresentabilità F e sia fl(x) la sua rappresentazione in F si dice errore assoluto di Roundoff il numero:

$$| \text{fl}(x) - x |$$

e si dice errore relativo di Roundoff il numero:

$$\frac{| \text{fl}(x) - x |}{| x |}$$

Inoltre l'errore di Roundoff si genera sia quando il numero reale viene rappresentato in F, sia nell'esecuzione di ogni operazione floating point.

**DOMANDE E RISPOSTE  
SUL CALCOLO DI  
PIGRECO**

**1) L'ALGORITMO DESCRITTO CHE IMPLEMENTA IL METODO DI ARCHIMEDE PER IL CALCOLO DI PIGRECO E' STABILE? PERCHE'?**

L'algoritmo è instabile, infatti si ha in base ai risultati precedenti che ad aumentare delle iterazioni l'errore relativo esplode.

**2) L'ALGORITMO DESCRITTO CHE IMPLEMENTA IL METODO DI LEIBNIZ PER IL CALCOLO DI PIGRECO E' STABILE? PERCHE'?**

E' stabile perchè l'errore di Roundoff cresce linearmente con n, ovvero il numero degli addendi, e quindi gli errori introdotti nei dati non si propagano amplificandosi in maniera tale che il risultato sia inaccettabile.

**3) E' NOTO UN METODO NUMERICO IL CUI ALGORITMO RISULTI COMPUTAZIONALMENTE PIU' O MENO VANTAGGIOSO RISPETTO AL METODO DI ARCHIMEDE PER IL CALCOLO DI PIGRECO, MOTIVARE LA RISPOSTA:**

Il metodo di Viete per il calcolo di Pigreco permette di ottenere risultati simili in meno iterazioni ed è anche stabile. Risulta quindi computazionalmente più vantaggioso rispetto al metodo di Archimede.

**4) E' NOTO UN METODO NUMERICO IL CUI ALGORITMO RISULTI COMPUTAZIONALMENTE VANTAGGIOSO RISPETTO AL METODO DI LEIBNIZ PER IL CALCOLO DI PIGRECO, MOTIVARE LA RISPOSTA:**

L'algoritmo di Archimede è computazionalmente più vantaggioso. Infatti con tale metodo si ottengono 3 cifre significative corrette con un poligono di 5 lati, mentre per ottenere lo stesso risultato con il metodo di Leibniz sono necessarie più 300 addendi a causa della lenta convergenza dello sviluppo in serie di Taylor.

**5) RICORDANDO L'ORDINE DI GRANDEZZA DELL'ERRORE DI TRONCAMENTO ANALITICO,  $|E_n|$ , ARRESTANDO LA FORMULA RICORRENTE DOPO N TERMINI, E DAI RISULTATI OTTENUTI, COSA SI PUO' AFFERMARE SULLA VELOCITA' DI CONVERGENZA DEL METODO DI LEIBNIZ?**

Se si utilizza il metodo di Leibniz, sebbene l'algoritmo sia stabile, a causa della lenta convergenza dello sviluppo in serie di McLaurin, solo per  $n=500000$  si ottiene il risultato accurato a 6 cifre significative.

**DOMANDE E RISPOSTE  
SULLA  
QUADRATURA**

**1) DARE LA DEFINIZIONE DI FORMULA DI QUADRATURA:**

Fissata una funzione integrabile nel senso di Riemann ed un interno n, dati n punti detti pesi  $X_i$  appartenenti a  $[a,b]$  detti nodi ed n valori  $A_i$  detti pesi, la combinazione lineare:

$$Q[f] = A_1 f(x_1) + \dots + A_n f(x_n)$$

è detta formula di quadratura.

**2) CHE COS E' L'ERRORE DI DISCRETIZZAZIONE?**

La differenza tra  $E[f] = I[f] - Q[f] = \int_a^b f(x) dx - \sum_{i=1}^n A_i f(x_i)$  è l'errore di discretizzazione della formula di quadratura  $Q[f]$ .

**3) QUANDO LA FORMULA DI QUADRATURA SI DICE ESATTA?**

I nodi e i pesi sono scelti in modo da minimizzare l'errore ( discretizzazione), una misura di tale errore è dato dal grado di precisione. Un modo pratico di calcolarlo è determinare una classe di funzioni per la quale la formula risulti esatta. Generalmente tale classe è quella dei polinomi per cui la formula si dice esatta di grado k.

**4) SCRIVERE L'ESPRESSIONE DELLA FORMULA TRAPEZOIDALE COMPOSITA,  $T_m[f]$ , SU M SOTTOINTERVALLI, PER IL CALCOLO DELL'INTEGRALE DEFINITO DI F:**

$$T_m[f] = h \left( \frac{f(x_0)}{2} + f(x_1) + f(x_2) + \dots + f(x_{m-1}) + \frac{f(x_m)}{2} \right) \quad \text{con } h = \frac{b-a}{m} \quad x_0 = a, x_m = b, x_i = a + i \cdot h$$

**5) FORNIRE UNA STIMA CALCOLABILE DELL'ERRORE DI DISCRETIZZAZIONE  $E_4[f]$  DELLA FORMULA TRAPEZOIDALE COMPOSITA  $T_4[f]$ :**

$$E_m[f] = T_m[f] - I[f] = E_{2m}[f] = T_{2m}[f] - I[f]$$

$$T_{2m}[f] = 2T_m[f] - T_m[f]$$

**6) QUANDO DUE FORMULE DI QUADRATURA SONO DI TIPO INNESTATO?**

Due formule  $Q1[f]$  e  $Q2[f]$  relative ad uno stesso intervallo, tali che l'insieme dei nodi di  $Q1[f]$  è contenuto nell'insieme dei nodi di  $Q2[f]$ , costituiscono una coppia di formule innestate.

**7) CHE VANTAGGIO COMPORTANO LE FORMULE DI TIPO INNESTATO?**

La disponibilità di formule di tipo innestato è particolarmente utile per ridurre la complessità computazionale delle formule di quadratura. Infatti essa è determinata attraverso il numero di valutazioni della funzione integranda dei nodi  $X_i$ , perchè il numero di operazioni relative a tale calcolo è predominante.

**8) PER QUALI FORMULE DI QUADRATURA E' POSSIBILE FORNIRE UNA STIMA CALCOLABILE DELL'ERRORE DI DISCRETIZZAZIONE? COSA RAPPRESENTA TALE STIMA?**

Per le formule di quadratura di tipo innestato.

**9) DESCRIVERE PERCHE' NELLO SVILUPPO SOFTWARE MATEMATICO PER LA QUADRATURA, CONVIENE UTILIZZARE FAMIGLIE DI FORMULE DI QUADRATURA COMPOSITA, DI TIPO INNESTATO:**

Poiché l'efficienza maggiore si raggiunge quando l'algoritmo calcola un risultato soddisfacente i requisiti di accuratezza con il minor numero di valutazioni di  $f(x)$ , le formule composite utilizzate in algoritmi di quadratura sono tali che, ad ogni passo dell'algoritmo, è possibile riutilizzare, tutte o quasi, le informazioni ottenute nei passi precedenti in modo tale da ridurre il numero di valutazioni della funzione integranda.

In particolare le formule più utilizzate sono quelle di tipo innestato come ad esempio le formule trapezoidali composite ottenute raddoppiando ad ogni passo il numero di intervalli.

**10) LE FORMULE SONO DI TIPO INNESTATO?**

Sì. Le formule composite utilizzate in algoritmi di quadratura sono tali che, ad ogni passo dell'algoritmo, è possibile riutilizzare, tutte o quasi, le informazioni ottenute nei passi precedenti in modo tale da ridurre il numero di valutazioni della funzione integranda.

In particolare le formule più utilizzate sono quelle di tipo innestato come ad esempio le formule trapezoidali composite ottenute raddoppiando ad ogni passo il numero di intervalli.

**11) SI DESCRIVA LA DIFFERENZA TRA STRATEGIA ADATTIVA E NON ADATTIVA:**

Un algoritmo adattivo è un algoritmo che sceglie dinamicamente la distribuzione dei nodi, in maniera da adottare il partizionamento dell'intervallo dell'integrazione al particolare andamento della funzione integranda. Mentre un algoritmo in cui l'insieme dei nodi è scelto secondo uno schema fissato indipendente dalla funzione integranda è detto algoritmo non adattivo.

**12) SE SI UTILIZZA UNA STRATEGIA ADATTIVA (LOCALE O GLOBALE) PER IL CALCOLO NUMERICO DI UN INTEGRALE DEFINITO, CON QUALE STRUTTURA DINAMICA SI DEVE IMPLEMENTARE TALE STRATEGIA ED IN QUALE ORDINE DEVONO ESSERE MEMORIZZATI, NELLA STRUTTURA DINAMICA, GLI INTERVALLI ESAMINATI SECONDO LA STRATEGIA IMPLEMENTATA?**

Se si vuole implementare una strategia adattiva globale si deve usare una lista ordinata, dove vengono conservate le informazioni relative a tutti gli intervalli esaminati, ed i cui intervalli sono memorizzati in ordine decrescente in base alla grandezza dell'errore di discretizzazione( e l'intervallo da dividere si trova sempre in testa alla lista).

Viceversa l'implementazione di una strategia adattiva locale si può realizzare mediante una pila costruita inserendo durante ogni suddivisione i due intervalli così ottenuti nella testa della pila, prima quello di destra e poi quello di sinistra (e l'intervallo da esaminare si trova sempre nella testa della pila).

**13) CONFRONTARE LA STRATEGIA LOCALE CON LA STRATEGIA GLOBALE IN TERMINI DI EFFICIENZA E DI OCCUPAZIONE DI MEMORIA RICHIESTA:**

La strategia Globale ha un vantaggio perchè fa poche valutazioni della funzione integranda, ma come svantaggio sfrutta molta memoria, mentre quella Locale esegue molte valutazioni e quindi è pesante in termini di efficienza computazionale, anche se occupa poca memoria perchè conserva informazioni solo quando la tolleranza è soddisfatta.

**14) DA CHE COSA DIPENDE LA COMPLESSITA' DI TEMPO DI UN ALGORITMO PER IL CALCOLO DI UNA FORMULA DI QUADRATURA?**

Il tempo dipende dalla tolleranza da dover soddisfare e anche dal massimo numero di valutazioni date.

## DEI DATI

### 1) CHE COSA SI INTENDE PER INTERPOLAZIONE POLINOMIALE?

L'interpolazione polinomiale è l'interpolazione di una serie di valori con una funzione polinomiale che passa per i punti. In particolare, un qualsiasi di  $n+1$  punti distinti può essere sempre interpolato da un polinomio di grado  $n$  che assume esattamente il valore dato in corrispondenza dei punti iniziali.

### 2) PER QUALI CASI SI ADOPERA L'INTERPOLAZIONE A TRATTI?

Se il numero dei nodi di interpolazione è elevato, il polinomio interpolante non fornisce in generale un modello accettabile. Infatti, al crescere del numero di punti aumenta il grado del polinomio interpolante ed aumentano anche le oscillazioni del polinomio corrispondente (unico problema gli spigoli). Ecco perché risolviamo tale problema con la SPLINE.

### 3) CHE COS'E' IL PROBLEMA DELL'APPROSSIMAZIONE?

Dati  $n$  valori distinti detti nodi ed  $n$  valori corrispondenti si vuole determinare una funzione  $f$  (detta funzione approssimante) la cui distanza nei nodi dai valori sia minima, la scelta della misura di tale distanza qualifica il problema di approssimazione qualificato.

### 4) IN GENERALE QUALE MODELLO SI PREFERISCE UTILIZZARE PER IL FITTING DI DATI AFFETTI DA ERRORE NON TRASCURABILE? PERCHE'?

Per errore non trascurabile è meglio usare un modello approssimante, perché la funzione ha valori la cui distanza dai nodi sia opportunamente piccola.

### 5) IL POLINOMIO INTERPOLANTE RISULTA SEMPRE IL PIU' ATTENDIBILE PER IL FITTING DI DATI, RISPETTO AD UN MODELLO APPROSSIMANTE? MOTIVARE LA RISPOSTA, DESCRIVENDO BREVEMENTE ALMENO UN ALTRO MODELLO INTERPOLANTE ED UNO APPROSSIMANTE ED IN QUALI IPOTESI POTREBBE ESSERE NECESSARIO UTILIZZARE L'UNO O L'ALTRO.

No, la scelta del modello dipende dal problema, in un caso in cui l'errore sui dati non è trascurabile e non è necessario passare per tutti i punti, come l'approssimazione dei minimi quadrati (costante elastica molla). Invece si deve usare il modello interpolante quando l'errore è trascurabile come nella descrizione di un profilo di un catamarano.

### 6) QUAL E' LA COMPLESSITA' DI TEMPO ASINTOTICO RICHIESTA DAL CALCOLO DEI COEFFICIENTI DELLA FORMULA DI NEWTON PER IL POLINOMIO INTERPOLANTE DI LAGRANGE?

$T(n) = n(n-1)/2 \text{ flop} \rightarrow T(n) = O(n^2) \text{ flop}$ .

### 7) ENUNCIARE LA FORMULAZIONE GENERALE DEL PROBLEMA DI INTERPOLAZIONE DI LAGRANGE, PARTICOLARIZZARE POI, LE CONDIZIONI DI INTERPOLAZIONE SULL'INSIEME DEI NODI.

Assegnati  $n$  punti  $P_n = (X_n, Y_n)$  nel piano diversi tra loro, l'interpolazione di Lagrange determina per quali valori di  $m$  esiste ed è unico il polinomio  $p(x) = a_m x^m + a_{m-1} x^{m-1} + \dots + a_0$  di grado al più  $m$ , e dati  $n$  nodi distinti  $(X_i)$   $i=1\dots n$  ed  $n$  valori  $(Y_i)$   $i=1\dots n$ , le condizioni di interpolazione di Lagrange determinano una funzione tale che  $f(x_i) = y_i$ ,  $i=1\dots n$ .

### 8) SOTTO QUALI CONDIZIONI IL POLINOMIO DI LAGRANGE INTERPOLANTE I NODI ASSEGNATI E' UNICO?

Dati  $n$  nodi  $X_1 \dots X_n$  distinti ed  $n$  valori corrispondenti  $Y_1 \dots Y_n$  il polinomio  $p$  di grado al più  $m$  tale che  $p(x_i) = y_i$ ,  $i = 1 \dots n$  è unico se  $m \leq n-1$ .

### 9) SCRIVERE L'ESPRESSIONE DEL POLINOMIO, RAPPRESENTANDOLO ATTRAVERO LA FORMULA DI NEWTON (ESPRIMENDO I COEFFICIENTI IN FUNZIONE DI DIFFERENZE DIVISE).

Considerato il polinomio  $Q_k = y[x_1, x_2, \dots, x_{k+1}]$  esprimendo i coefficienti  $K$ -ima  $a_k$  in funzione delle differenze divise si ha:  
 $p(x) = y[x_1] + y[x_1, x_2](x - x_2) + \dots + y[x_1, \dots, x_n](x - x_1)(x - x_2) \dots (x - x_{n-1})$

### 10) ESPRIMERE IL POLINOMIO DI LAGRANGE CON L'AGGIUNTA DI UN PUNTO A PARTIRE DAL POLINOMIO PRECEDENTE.

$P(x) = y[x_1] + y[x_1, x_2](x_1 - x_2) + y[x_1, x_2, x_3](x - x_1)(x - x_2)$ .  
 $Q(x) = p(x) + a_4(x - x_1)(x - x_2)(x - x_3)$ .

**11) QUAL E' LA COMPLESSITA' DI TEMPO DELL'ALGORITMO DI HORNER?**

**$T(n) = O(n)$  Flop.**

**13) DESCRIVERE ALMENO UNO SCHEMA DI COSTRUZIONE DELLA TABELLA DELLE DIFFERENZE DIVISE PER IL CALCOLO DEI COEFFICIENTI DELLA FORMULA DI NEWTON PER IL POLINOMIO INTERPOLANTE DI LAGRANGE.**

Schema di AITKEN:

Differenze Divise:

$$y[x_1, x_2] = y_2 - y_1 / x_2 - x_1$$

$$y[x_2, x_3] = y_3 - y_2 / x_3 - x_2$$

$$y[x_1, x_2, x_3] = y[x_2, x_3] - y[x_1, x_2] / x_3 - x_1$$

$$y[x_2, x_3, x_4] = y[x_3, x_4] - y[x_2, x_3] / x_4 - x_2$$

$$y[x_1, x_2, x_3, x_4] = y[x_2, x_3, x_4] - y[x_1, x_2, x_3] / x_4 - x_1$$

**14) CHE COS'E' LO SPLINE?**

Una SPLINE è una funzione, costituita da un insieme di polinomi raccordati tra loro, il cui scopo è interpolare in un intervallo un insieme di punti (detti nodi della SPLINE), in modo da essere continua (almeno fino ad un dato ordine di derivate) in ogni punto dell'intervallo.

Definizione (incompleta):

Sia data  $A = \{a = x_0 < x_1 < \dots < x_n = b\}$  una suddivisione dell'intervallo chiuso  $[a, b]$ . Una funzione SPLINE di grado  $p$  con nodi nei punti  $x_i$  con  $i = 0, 1, 2, \dots, n$  è una funzione su  $[a, b]$  indicata con  $sp(x)$  tale che nell'intervallo  $[a, b]$  si abbia:

1) In ogni sottointervallo  $[x_i, x_{i+1}]$  con  $i = 0, 1, 2, \dots, n-1$  la funzione  $sp(x)$  è un polinomio di grado  $p$ .

2) La funzione  $sp(x)$  e le sue prime  $p-1$  derivate sono continue.

**15) DESCRIVERE BREVEMENTE L'ALGORITMO IMPLEMENTATO PER LA COSTRUZIONE DELLA SPINE CUBICA NATURALE INTERPOLANTE I NODI ASSEGNATI E LA SUA COMPLESSITA' DI TEMPO:**

- Ordinamento dei nodi e dei valori corrispondenti.
- Costruzione di  $A$  e del vettore dei termini noti con  $A$  tridiagonale, ben condizionata.
- Risoluzione del sistema in tempo lineare  $T(n) = O(n)$  flop.

**DOMANDE E RISPOSTE  
RISOLUZIONE DI SISTEMI LINEARI  
E  
CALCOLO MATRICIALE**

**1) CHE COSA SI INTENDE PER MATRICE SINGOLARE?**

Una matrice singolare è una matrice quadrata con determinante pari a zero.

**2) QUAL E' LA COMPLESSITA' DI TEMPO E DI SPAZIO DELL'ALGORITMO DI BACK-SUBSTITUTION?**

$$T_{bs}(n) = n^2 \rightarrow O(n^2)$$

$$S_{bs}(n) = n^2 + 2n \rightarrow O(n^2)$$

**3) QUAL E' LA COMPLESSITA' DI TEMPO E DI SPAZIO ASINTOTICA DEL METODO DI ELIMINAZIONE DI GAUSS?**

$$T_{gauss}(n) = 2(n^3 - n)/3 \text{ flop} + (n^2 - n)/2 \text{ flop} = O(n^3)$$

(sostituire ad  $n$  la dimensione della matrice per ricavare il numero di Flop necessari).

$$S_{gauss}(n) = n(n+1) = O(n^2)$$

**4) DESCRIVERE IN COSA CONSISTE LA TECNICA DEL PIVOTING PARZIALE:**

La strategia di pivoting parziale, seleziona l'elemento pivot al passo  $k$ , l'elemento di massimo modulo tra quelli della  $k$ -ma colonna a partire da quello diagonale e lo scambia.

**5) DESCRIVERE COME INCIDE IL PIVOTING PARZIALE SULLA STABILITA' DELL'ALGORITMO DI GAUSS:**

- Il pivoting parziale rende l'algoritmo di Gauss praticamente stabile.

**6) QUAL E' LA COMPLESSITA' DI TEMPO ASINTOTICA DEL METODO DI ELIMINAZIONE DI GAUSS CON PIVOTING PARZIALE?**

$$T_{\text{gaussPv}}(n) = O(n^2/2)$$

**7) DIRE QUAL E' L'ALGORITMO PIU' EFFICIENTE PER FATTORIZZARE LA MATRICE NEL PRODOTTO DI DUE MATRICI TRIANGOLARI E COME SI ESPRIME TALE FATTORIZZAZIONE:**

E' l'algoritmo di fattorizzazione LU, dove L è la matrice triangolare inferiore composta dai moltiplicatori e con tutti 1 sulla diagonale principale, ed U è la matrice triangolare superiore ottenuta all'ultimo passo dell'algoritmo di eliminazione di Gauss.

Si ha quindi che  $LU=PA$ .

**8) QUAL E' LA COMPLESSITA' DI TEMPO ASINTOTICA DEL METODO DI FATTORIZZAZIONE LU?**

$$TLU(n) = (n^3/3 + mn^2)$$

**9) SCRIVERE LA COMPLESSITA' DI TEMPO ASINTOTICA DELL'ALGORITMO UTILIZZATO PER IL CALCOLO DELLA MATRICE INVERSA E PER IL CALCOLO DEL DETERMINANTE:**

$$T^{-1}(n) = O(n^3/3) + 2nO(n^2/2) = O(4/3n^3)$$

$T(n)$  determinante =  $O(n!)$  ed utilizzando la fattorizzazione LU il costo totale diventa =  $O(n^3/3)$

**10) DESCRIVERE BREVEMENTE L'ALGORITMO IMPLEMENTATO PER IL CALCOLO DELL'INVERSA DI UNA MATRICE QUADRATA, DI DIMENSIONE N TRIANGOLARE SUPERIORE:**

Per calcolare l'inversa di una assegnata matrice A di dimensione n, triangolare superiore occorre effettuare n volte l'algoritmo di back-substitution.

(L'algoritmo implementato sfrutta la fattorizzazione LU per il calcolo della matrice inversa, ma siccome nel nostro caso la matrice è già triangolare superiore l'algoritmo esegue solo n volte la back-substitution).

**11) ENUNCIARE IL TEOREMA DEL CONDIZIONAMENTO:**

Sia  $\| \cdot \|$  una norma matriciale submoltiplicativa (cioè  $\|AB\| \leq \|A\| \|B\|$ ) e compatibile con una norma vettoriale (cioè  $\|Ax\| \leq \|A\| \|x\|$  per ogni x diverso da 0). Sia inoltre il sistema  $Ax = b$  (con A non singolare) e si consideri il sistema perturbato:

$$(A + \Delta A)(x + \Delta x) = (b + \Delta b)$$

Se

$$\|\Delta A\| < 1 / \|A^{-1}\|$$

posto  $\mu(A) = \|A\| \|A^{-1}\|$  si ha:

$$\|\Delta x\| / \|x\| \leq \mu(A) / (1 - \mu(A) \|\Delta A\| / \|A\|) (\|\Delta A\| / \|A\| + \|\Delta b\| / \|b\|)$$

**12) ENUNCIARE IL CRITERIO DI SYLVESTER:**

Sia A una matrice simmetrica reale  $n \times n$  (i cui valori sono numeri reali). Per  $i=1, \dots, n$ , sia  $D_i$  il determinante del minore ottenuto cancellando da A le ultime  $n-i$  righe e le ultime  $n-i$  colonne. La matrice A è definita positiva se e solo se  $D_i > 0$  per ogni i.

**13) QUANDO UNA MATRICE SI DEFINISCE POSITIVA?**

Una matrice A simmetrica si definisce positiva se per ogni x, con x diverso da zero, risulta che:

$$x^T A x > 0$$

**14) SPIEGARE IN COSA CONSISTE LA FATTORIZZAZIONE DI CHOLSKY:**

Sia A appartenente ad  $R^{n \times n}$  una matrice simmetrica definita positiva.

Allora esiste una ed una sola matrice triangolare inferiore L appartenente ad  $R^{n \times n}$ ,

con  $l_{k,k} > 0$  per  $k = 1..n$ , tale che:

$$A = LL^T$$

**15) DIRE QUAL E' L'ALGORITMO PIU' EFFICIENTE PER LA FATTORIZZARE LA MATRICE SIMMETRICA S (NON POSITIVA), NEL PRODOTTO DI DUE MATRICI TRIANGOLARI, E COME SI ESPRIME TALE FATTORIZZAZIONE?**

Sia  $A \in \mathbb{R}^{n \times n}$  una matrice simmetrica tale che tutte le sue sottomatrici principali sono non singolari. Esistono allora una ed una sola matrice  $L \in \mathbb{R}^{n \times n}$  triangolare unitaria inferiore ed una ed una sola matrice  $D \in \mathbb{R}^{n \times n}$  diagonale tali che

$$A = LDL^T$$

Inoltre,  $L$  coincide con la matrice triangolare inferiore della fattorizzazione LU di  $A$  e gli elementi diagonali di  $D$  coincidono con i corrispondenti elementi diagonali della matrice triangolare superiore della fattorizzazione LU di  $A$ , ovvero  $L^T = D^{-1}U$ .

I risultati precedenti suggeriscono che il calcolo della fattorizzazione  $LDL^T$  si può eseguire modificando opportunamente l'algoritmo di eliminazione di Gauss, utilizzato per ottenere la fattorizzazione LU. Inoltre, grazie alla simmetria della matrice  $A$  e delle sottomatrici attive, ad ogni passo, dopo avere calcolato i moltiplicatori, si può operare solo sul triangolo superiore della sottomatrice attiva corrente.

L'algoritmo si può eseguire in place, memorizzando gli elementi del triangolo superiore di  $A$  al posto

dei corrispondenti elementi di  $L^T$  e gli elementi diagonali di  $D$  al posto di quelli di  $A$ ;

gli elementi diagonali di  $L$ , essendo tutti uguali ad 1, non sono memorizzati. Una descrizione di tale algoritmo è fornita di seguito.

**16) SI DESCRIVA LA COMPLESSITA' DI TEMPO E DI SPAZIO ASINTOTICA DELL'ALGORITMO DI FATTORIZZAZIONE LU SENZA PIVOTING SPECIFICO PER UNA MATRICE A BANDA.**

$TLU_{band}(n,p,q) = O(npq)$  flop

$SLU_{band}(n,p,q) = (p+q+1)n = O((p+q+1)n)$

**17) DESCRIVERE UNO SCHEMA DI MEMORIZZAZIONE DI UNA MATRICE A BANDA CHE CONSENTA UN RISPARMIO IN TERMINI DI COMPLESSITA' DI SPAZIO RISPETTO ALLA MEMORIZZAZIONE DI TUTTI GLI ELEMENTI DELLA MATRICE:**

In generale, per memorizzare una generica matrice a banda  $A$ , di dimensione  $n \times n$ , con ampiezza di banda  $p$  e  $q$ , si utilizza un array bidimensionale di dimensione  $(p+q+1) \times n$ , nel modo seguente:

$AB(q+1+i-j, j) = a_{ij}, \max(1, j-q) \leq i \leq \min(n, j+p)$ ,

Tale schema di memorizzazione è detto a banda ed ha una complessità di spazio pari a  $(p+q+1)n$ .

E' chiaro che la memorizzazione a banda è significativamente vantaggiosa, rispetto alla usuale memorizzazione mediante un array bidimensionale di dimensione  $n \times n$ , se  $p, q \ll n/2$ .

**18) DESCRIVERE UNO SCHEMA DI MEMORIZZAZIONE DI UNA MATRICE TRIDIAGONALE CHE CONSENTA UN RISPARMIO IN TERMINI DI COMPLESSITA' DI SPAZIO RISPETTO ALLA MEMORIZZAZIONE DI TUTTI GLI ELEMENTI DELLA MATRICE:**

In generale, data una matrice tridiagonale  $A$  di dimensioni  $n \times n$  si utilizza un array monodimensionale di lunghezza  $n$  per memorizzare la diagonale principale e due array monodimensionali di lunghezza  $n-1$  per memorizzare la diagonale inferiore e quella superiore. Tale schema di memorizzazione ha una complessità di spazio pari a  $3n-2$ , cioè  $O(n)$ , di un ordine di grandezza inferiore rispetto alla usuale memorizzazione mediante un array bidimensionale.

**19) SI DESCRIVA LA COMPLESSITA' DI TEMPO E DI SPAZIO ASINTOTICA DELL'ALGORITMO DI FATTORIZZAZIONE LU SENZA PIVOTING SPECIFICO PER UNA MATRICE TRIDIAGONALE.**

$TLU_{trid}(n) = 3n - 3 = O(n)$  flop

$SLU_{trid}(n) = 3n - 2 = O(n)$

**20) SPIEGARE BREVEMENTE I VANTAGGI COMPUTAZIONALI CHE SI OTTENGONO SCAMBIANDO VIRTUALMENTE INVECE CHE EFFETTIVAMENTE LE RIGHE DELLA MATRICE:**

Utilizzando il vettore  $I_{piv}$  per non scambiare fisicamente le righe della matrice e quindi evitare di costruire la matrice per gli scambi, e nella forward-substitution tramite  $I_{piv}$  già si conoscono gli scambi da effettuare. Tutto a favore di minore memoria occupata e di vantaggi computazionali sui calcoli.

**21) SI DESCRIVA UN ALTRO MODO IN CUI L'UTILIZZO DELLA FATTORIZZAZIONE LU CONSENTE UN RISPARMIO, IN TERMINI DI COSTO COMPUTAZIONALE, RISPETTO ALL'UTILIZZO DELL'ALGORITMO DI GAUSS.**



La fattorizzazione LU oltre a calcolare l'inversa può risolvere il problema del calcolo del determinante.

**15) DIRE QUAL E' L'ALGORITMO PIU' EFFICIENTE PER RISOLVERE UN SISTEMA LINEARE CON MATRICE SIMMETRICA DEFINITA POSITIVA? QUAL E' LA SUA COMPLESSITA' DI TEMPO E DI SPAZIO?**

E' la fattorizzazione di Cholesky, ossia:

sia A appartenente ad  $R^{n \times n}$  una matrice simmetrica definita positiva.

Allora esiste una ed una sola matrice triangolare inferiore L appartenente ad  $R^{n \times n}$ ,

con  $l_{k,k} > 0$  per  $k = 1..n$ , tale che:

$$A = LL^T$$

$TChol(n) = O(n^{3/2})$  flp.

$SChol(n) = O(n^2)$  mentre se si utilizza uno schema di memorizzazione packed risulta essere  $O(n^2/2)$  e quindi risulta dimezzata.

**Metodo di Dekker Brent:**

Il metodo dekker-brent è un metodo ibrido che rappresenta una combinazione dei metodi di bisezione e secanti. Nel metodo dekker-brent a fine di fattorizzare una sufficiente velocità di convergenza a partire da un intervallo di ricerca [a,b] la scelta dell'approssimazione successiva avviene mediante il metodo delle secanti. Se il valore così calcolato cade all'esterno dell'intervallo di ricerca si utilizza il metodo di bisezione al fine di garantire la convergenza di individuare un intervallo contenuto nel precedente come per il metodo di bisezione.

**L'algoritmo di Leibnitz è stabile? Perché?**

L'algoritmo di Leibnitz è stabile perché l'errore di round off introdotto nei dati non si propaga amplificandosi in maniera tale che il risultato sia inaccettabile

**E' noto un metodo numerico il cui algoritmo risulti computazionalmente vantaggioso rispetto al metodo di Leibnitz, per il calcolo di pi-greco? Motivare la risposta**

Il metodo di Viete per il calcolo di pi-greco

**Fornire un esempio di problema ben condizionato e di uno mal condizionato**

Un problema ben condizionato è la risoluzione di un sistema triangolare superiore mentre mal condizionato la risoluzione della matrice di hilbert o di vandermonde

**E' possibile rendere ben condizionato il problema mal condizionato descritto? Perché?**

No, perché il condizionamento di un problema dipende dalla natura del problema stesso

**Espressione dell'indice di condizionamento**

$|F'(x)|x / |f(x)|$  e poi svolgi

**Formula Trapeziodale composta**

$T_m[f] = h( f(x_0)/2 + f(x_1) + f(x_2) + \dots + f(x_{m-1}) + f(x_m)/2 )$  con  $h = (b-a)/m$   $x_0 = a, x_m = b, x_i = a + i \cdot h$

**tavola delle differenze divise per il calcolo dei coefficienti della formula di newton per il polinomio interpolare di Lagrange**

$x_i$	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$
$x_0$	$f(x_0)$			
$x_1$	$f(x_1)$	$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$		
$x_2$	$f(x_2)$	$f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$	$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$	
$x_3$	$f(x_3)$	$f[x_2, x_3] = \frac{f(x_3) - f(x_2)}{x_3 - x_2}$	$f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1}$	$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$

**La strategia adattativa globale è più o meno efficiente della strategia adattativa locale? La strategia adattativa globale richiede una maggiore o minore occupazione di memoria rispetto alla strategia adattativa locale?**

La strategia adattativa globale è più efficiente perché fa poche valutazioni della funzione ma il suo svantaggio è che sfrutta molta memoria, quella locale invece esegue molte valutazioni di conseguenza risulta più pesante in termini di efficienza computazionale, ma occupa poca memoria perché conserva informazioni solo quando la tolleranza sull'errore è soddisfatta

**Si descriva almeno un altro problema in cui l'utilizzo della fattorizzazione LU consente un risparmio, un termine di costo computazionale rispetto all'utilizzo dell'algoritmo di Gauss motivare la risposta attraverso il calcolo della complessità di tempo richiesta dall'algoritmo in entrambi i casi**

La fattorizzazione LU oltre a calcolare l'inversa può risolvere il problema del calcolo del determinante

**Complessità di tempo asintotica dell'algoritmo di Horner per la valutazione del polinomio interpolante di Lagrange rappresentato attraverso la formula di Lagrange?**

$T(n) = O(n^2)$  flops

**Complessità di tempo asintotica dell'algoritmo di Horner per la valutazione del polinomio interpolante di Lagrange rappresentato attraverso la formula di Newton?**

$T(n) = O(n)$  flops

**Cosa si intende con il dire che il sistema aritmetico è caratterizzato dalla massima accuratezza dinamica?**

Un sistema aritmetico garantisce che il risultato di qualsiasi operazione f.p. differisca dal risultato dell'operazione in R corrispondente (il risultato esatto), di una quantità che è il solo errore di rappresentazione di r, e viene perciò detto sistema a massima accuratezza dinamica.

**Le formule di quadratura  $T_2[f]$  e  $T_4[f]$  sono di tipo innestato?**

Sì

**Descrivere perché, nello sviluppo di software matematico per la quadratura, conviene utilizzare famiglie di formule di quadratura composite, di tipo innestato.**

Poiché l'efficienza maggiore si raggiunge quando l'algoritmo calcola un risultato soddisfacente i requisiti di accuratezza con il minor numero di valutazioni di  $f(x)$ , le formule composite utilizzate in algoritmi di quadratura sono tali che, ad ogni passo dell'algoritmo, è possibile riutilizzare tutte, o quasi, le informazioni ottenute nei passi precedenti, in modo da ridurre il numero di valutazioni della funzione integranda. In particolare le formule più utilizzate sono quelle di tipo innestato come, ad esempio, le formule trapezoidali composite  $\{T_m[f]\}_{m=2k}$ ,  $k = 0, 1, 2, \dots$  ( $T_1[f]$   $T[f]$ ) ottenute raddoppiando ad ogni passo il numero di intervalli.

**Da cosa dipende la complessità di tempo di un algoritmo per il calcolo di una formula di quadratura?**

Considerando la complessità di tempo di un algoritmo (ad es.):  $TLU_{band}(n, p, q) = O(npq)$  flop.

Si può notare che al crescere di p e q la complessità di tempo si avvicina a  $O(n^3)$  flop, cioè a quella dell'algoritmo

per una matrice di dimensione  $n \times n$ , e che per p, q n, essa è proprio  $O(n^3)$  flop. L'algoritmo è dunque "significativamente vantaggioso" se  $p, q \ll n$ .

**Descrivere l'algoritmo per la costruzione e valutazione della spline cubica naturale interpolante i nodi assegnati.**

Passo 1): costruzione della matrice A e del vettore dei termini noti  $3B f$ ;

Passo 2): risoluzione del sistema  $A = 3B f$ ;

Passo 3): determinazione dell'intervallo  $[x_i, x_{i+1}]$  a cui x appartiene;

Passo 4): calcolo dei coefficienti del polinomio interpolante di Hermite, nell'intervallo  $[x_i, x_{i+1}]$ ;

Passo 5): valutazione in x.

**Dire qual'è l'algoritmo più efficiente per fattorizzare la matrice nel prodotto di due matrici triangolari e come si esprime tale fattorizzazione.**

E' l'algoritmo di fattorizzazione LU il quale tramite l'algoritmo di Gauss trasforma il sistema, in un sistema

triangolare superiore, l'altro fattore consiste in una matrice triangolare inferiore composta da i moltiplicatori e dalla diagonale principale formata da tutti 1.

**Fornire una stima calcolabile dell'errore di discretizzazione,  $E_4[f]$ , per la formula trapezoidale composta  $T_4[f]$ .**

La differenza:

$$E[f] = I[f] - Q[f] = a$$

$$b \int_a^b f(x) dx - n$$

$$i=1 \sum_{i=1}^n A_i f(x_i)$$

è l'errore di discretizzazione della formula di quadratura  $Q[f]$ .

**Enunciato criterio di Sylvester**

Una matrice  $A$  è definita positiva se il determinante di ciascuna sottomatrice principale di  $A$  è  $> 0$ .

**Sistema a massima accuratezza relativa:**

È un sistema aritmetico che garantisce che il risultato di una qualsiasi operazione floating point differisca dal risultato dell'operazione in  $R$  corrispondente (il risultato esatto), di una quantità che è il solo errore di rappresentazione in  $r$ .

Indice di condizionamento:

Detto l'errore nei dati e l'errore corrispondente nella soluzione, e posto

$$\kappa = \frac{\epsilon_{\text{sol}}}{\epsilon_{\text{dati}}}$$

è detto indice di condizionamento del problema; risulta, quindi, che:

se  $\kappa \approx 1$  il problema è ben condizionato;

se  $\kappa \gg 1$  il problema è mal condizionato.

L'**interpolazione polinomiale** è l'interpolazione di una serie di valori (es. dei dati sperimentali) con una funzione polinomiale che passa per i punti dati. In particolare, un qualsiasi insieme di  $n+1$  punti distinti può essere sempre interpolato da un polinomio di grado  $n$  che assume esattamente il valore dato in corrispondenza dei punti iniziali.

**Una spline** è una funzione, costituita da un insieme di polinomi raccordati tra loro, il cui scopo è interpolare in un intervallo un insieme di punti (detti *nod*i della spline), in modo da essere continua (almeno fino ad un dato ordine di derivate) in ogni punto dell'intervallo.

**Definizione:**

Sia data una suddivisione dell'intervallo chiuso  $[a, b]$ . Una funzione *spline* di

grado  $p$  con *nod*i nei punti  $x_i$  con  $i = 1, \dots, n$  è una funzione su  $[a, b]$  indicata con  $sp(x)$  tale che, nell'intervallo  $[a, b]$  si abbia:

1. in ogni sottointervallo  $[x_i, x_{i+1}]$  con la funzione  $sp(x)$  è un polinomio di grado  $p$ .

2. la funzione  $sp(x)$  e le sue prime  $p-1$  derivate sono continue.

**Quando la formula di quadratura si dice esatta?**

I nodi e i pesi sono scelti in modo da minimizzare l'errore (di discrezione). Una misura di tale errore è dato dal grado di precisione. Un modo pratico di calcolarlo è determinare una classe di funzioni per la quale la formula risulti esatta. Generalmente tale classe è quella dei polinomi per cui una formula si dice esatta di grado  $k$ .

Il metodo di Dekker-Brent rappresenta una combinazione dei metodi di Bisezione e Secanti.

In questo metodo, a partire da un intervallo di ricerca  $[a, b]$  la scelta dell'approssimazione successiva avviene mediante metodo delle secanti, se il valore così calcolato cade all'esterno dell'intervallo di ricerca, si utilizza il metodo di bisezione.

Al fine di garantire la convergenza si individua un intervallo contenuto nel precedente come per il metodo di bisezione.

**Cardinalità dei numeri macchina di  $F$ :** sono i numeri reali che appartengono all'insieme

$$[r_{\min}, r_{\max}] \setminus \{0\} \cup [r_{\min}, r_{\max}]$$

Definizione 4.2. (Errore di discretizzazione)

La differenza:

$$E[f] = I[f] - Q[f]$$

E' l'errore di discretizzazione della formula di quadratura  $Q[f]$ .

### **Le formule sono di tipo innestato?**

Si. Le formule composite utilizzate in algoritmi di quadratura sono tali che, ad ogni passo dell'algoritmo, `e possibile

riutilizzare tutte, o quasi, le informazioni ottenute nei passi precedenti, in modo da ridurre il numero di valutazioni della funzione integranda. In particolare le formule più utilizzate sono quelle di tipo innestato come, ad esempio, le

formule trapezoidali composite ottenute raddoppiando ad ogni passo il numero di intervalli.

Problema ben condizionato: risoluzione di un sistema triangolare superiore

Problema mal condizionato: risoluzione della matrice di Hilbert e Vendermonde

Un problema mal condizionato non può diventare ben condizionato perché il condizionamento di un problema dipende dalla natura del problema stesso.

Complessità asintotica dell'algoritmo per il calcolo della matrice inversa  $A^{-1}$

$$A^{-1} = LU + n \text{ (back+forward)}$$

$$T(n) = O(n^3) + n(O(n) + O(n)) = O(n^2)T(n) = O(4/5 n^3)$$

La fattorizzazione LU risolve il problema del calcolo del determinante.

La complessità di tempo asintotica per il calcolo dell'inversa di una matrice tridiagonale:

$$T(n) = 3n = O(n) \text{ flops}$$

Complessità del metodo di Gauss (tempo):

$$T(\text{Gauss}) = 2(n^3 - n)/3 + (n^2 - n)/2 = O(n^3)$$

Complessità del metodo di Gauss (tempo):

$$S(\text{Gauss}) = n(n + 1) = O(n^2)$$