

TEORIA CALCOLO NUMERICO

rappresentazione floating point

$$(-1)^s p N^q$$

$$s \in \{0, 1\}$$

p reale positivo, $1/N \leq p < 1$ implica che la prima cifra della mantissa è 1, se no troppo piccolo

q intero, limitato da minimo L e massimo U

N è base del sistema di numerazione ($N=2$)

underflow

$$0,1000\dots N^L$$

overflow

$$0,1111\dots N^U$$

round to even

quando l'ultima cifra decimale è 5, se la cifra che precede è pari la si lascia, se la cifra che precede è dispari si trasforma nel pari successivo:

$$(0,440) \quad 0,445 \quad (0,450)$$

$$(0,850) \quad 0,855 \quad (0,860)$$

cancellazione numerica

sottrazione tra due numeri dello stesso segno, dello stesso ordine di grandezza, la mantissa approssimabile, almeno in uno dei due numeri è presente errore di arrotondamento

$$s_1 = s_2$$

$$q_1 = q_2$$

$$p_1 \approx p_2$$

$\epsilon = N^{1-t}$ dove t è il numero di cifre della mantissa

precisione di macchina = $\epsilon/2$

due quantità si definiscono **equivalenti** in aritmetica del

calcolatore se $|\overline{e}_1 - \overline{e}_2| / |\overline{e}_1|$ o $|\overline{e}_1 - \overline{e}_2| / |\overline{e}_2| \approx \epsilon/2$

polinomio interpolante

$$P_n(x) = \sum_{j=1}^{n+1} y_j l_j(x)$$

il polinomio $l_j(x)$ si costruisce così:

al numeratore il prodotto di tutti gli $x-x_i$ tranne il termine $x-x_j$ così si annullerà ogni volta che x vale x_i

al denominatore il prodotto di tutti gli x_j-x_i con gli stessi x_i del numeratore

così quando x vale x_j numeratore e denominatore si equivalgono e il polinomio in x_j vale 1

in questo modo ho ottenuto che l_j vale 0 in tutte le x dei nodi dati tranne quella del nodo x_j dove vale 1, se moltiplico l_j per y_j del nodo, ho imposto l'interpolazione del nodo se sommo tutti i polinomi l_j per le diverse x_j , moltiplicati per il rispettivo valore y_j , ottengo l'interpolazione di tutti i nodi, perché ogni nodo vale sempre 0 in tutti gli addendi tranne il proprio in cui vale y_j

ASSEGNATI N+1 PUNTI DI ASCISSE X DISTINTE, $\exists!$ POLINOMIO DI GRADO $\leq N$ INTERPOLANTE I DATI ASSEGNATI

INTERPOLAZIONE DI FUNZIONI: all'aumentare dei nodi, garantisce la convergenza uniforme...

	f continua	f di classe C^1 o superiore
qualsiasi scelta di nodi	spline di qualsiasi ordine, se il passo $h \rightarrow 0$	spline di qualsiasi ordine, se il passo $h \rightarrow 0$
Chebyshev/ Chebyshev-Lobatto	spline di qualsiasi ordine, se il passo $h \rightarrow 0$	polinomio interpolante spline di qualsiasi ordine, se il passo $h \rightarrow 0$

ORDINI DI CONVERGENZA

	POLINOMIO INTERPOLANTE	$S_1(x) \ h \rightarrow 0$	SPLINE CUBICA $h \rightarrow 0$
funzione	$O(\log(n)/n^k)$ $f \in C_k, k \geq 1$ nodi di Chebyshev/ Chebyshev-Lobatto	$O(h^2)$ massimo per $S_1(x)$ $f \in C_k([a; b]),$ $k > 2$	$O(h^2)$ se $f \in C^2$ $O(h^4)$ massimo per $S_3(x)$ $f \in C_k([a; b]),$ $k > 4$ (saturazione)
derivate			S_3 knot-a-knot o vincolate, se $f \in C^4$ $O(h^{4-p})$ p è grado della derivata

SPLINE CUBICHE

devo determinare $4n$ coefficienti (ogni polinomio ha coefficiente su termine di grado 3,2,1,0 ovvero quattro coefficienti, e i polinomi sono n perché i nodi sono $n+1$) risolvendo per $4n-2$ condizioni:

$n+1$ condizioni per l'interpolazione di $n+1$ nodi: $n+1$

$n-1$ condizioni di continuità per ordini di derivate fino alla seconda: $3(n-1)$

servono 2 condizioni aggiuntive se non ho un sistema indeterminato ($4n-2$ equazioni e $4n$ incognite)

spline naturali: nel primo e nell'ultimo nodo (estremi) la derivata seconda vale 0

spline knot-a-knot: nel secondo e nel penultimo nodo la derivata terza è continua
viene restituita da MATLAB se i vettori x e y hanno lo stesso numero di elementi

spline vincolare: nel primo e nell'ultimo nodo la derivata prima della spline ha lo stesso valore della derivata prima della funzione da approssimare
viene restituita da MATLAB se il vettore y ha esattamente 2 elementi in più del vettore x ; il primo elemento di y sarà interpretato come il valore della derivata prima nel primo nodo, l'ultimo elemento di y sarà interpretato come il valore della derivata prima nell'ultimo nodo

norma 1 $\max(\sum |x_i|)$ sommo i valori assoluti delle entrate di ciascuna colonna
 $\text{norm}(A, 1)$

norma infinito $\max(\sum |x_j|)$ sommo i valori assoluti delle entrate di ciascuna riga
 $\text{norm}(A, \text{inf})$

$\|A\|_{\infty} = \|A^T\|_1$

norma spettrale

$$\|A\|_2 = \sqrt{\rho(A^T A)}$$

Teorema

$$\text{er} = 2K(A) \quad (\text{er}(A) + \text{er}(b))$$

$$K(A) \geq 1$$

l'errore iniziale è almeno dell'ordine della precisione di macchina 10^{-16} , e si espande a più del doppio di $K(A)$; se espando di 10^{16} ottengo un errore pari circa all'unità: $10^{-16} \cdot 10^{16} = 1$

matrice ortogonale A è ortogonale se la sua trasposta è anche la sua inversa

matrice di permutazione identità a cui scambio ordine delle righe

matrice simmetrica definita positiva

matrice diagonale dominante per righe per tutte le righe il valore assoluto dell'elemento diagonale della riga è $>$ (strettamente) della somma del valore assoluto degli altri elementi della riga

matrice tridiagonale

costo computazionale

matrice triangolare $n^2/2$

matrice non triangolare, Gauss $n^3/3$

matrice simmetrica che non necessita di scambi $n^3/6$

matrice non triangolare, cramer $(n+1)!$

il numero di cicli for annidati dà il grado di n nel costo computazionale

operare scambi se elemento $a_{kk}=0$; scambi non comportano costo computazionale, quindi sempre conveniente operare scambi se $|a_{kk}| \approx 0$ perché andando al denominatore un denominatore prossimo a 0 fa

schizzare la frazione a valori molto alti, si rischia di propagare l'errore. **pivoting parziale**: sostituire a l'elemento pivot a_{kk} con il massimo valore assoluto, tra gli elementi sotto a_{kk} nella stessa colonna

Gauss senza scambi garantito se **A diagonale dominante per righe** o **A diagonale dominante per colonne** o **A simmetrica definita positiva**

fattorizzazione LU (per Gauss senza scambi)

$$L = \text{tril}(A, -1) + \text{eye}(n)$$

$$U = \text{triu}(A)$$

risoluzione sistema con PA=LU

$$PAx = Pb$$

$$LUx = Pb$$

$$1) Ly = Pb$$

$$2) Ux = y$$

trovo vettore x

$$PA = LU \rightarrow \det(P)\det(A) = \det(L)\det(U)$$

$\det(L)=1$ poiché triangolare inferiore con tutti "1" sulla diagonale principale

$\det(U) = \prod_{i=1}^n u_{ii}$ poiché triangolare superiore con entrate diagonali che non conosciamo a priori

$\det(P) = (-1)^s$ con s numero di scambi di righe effettuati

$$\det(A) = (-1)^s \prod_{i=1}^n u_{ii}$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

Cholesky per matrice A simmetrica definita positiva

$$A = R^T R$$

$$A^{-1} = (R^T R)^{-1} = R^{-1} (R^T)^{-1} = R^{-1} (R^{-1})^T$$

inversa della trasposta = trasposta dell'inversa

in questo modo, conoscendo solo R^{-1} e non $(R^T)^{-1}$ (nel penultimo passaggio compare l'inversa della trasposta, ma grazie alla proprietà possiamo non conoscerla e conoscere soltanto l'inversa di R, che infatti compare due volte nell'ultimo passaggio), è possibile trovare l'inversa di A

matrice dei coefficienti $a_{ij} = \cos((j-1)\theta_i)$ con $\theta_i = \frac{2i-1}{2n}\pi$

vettore dei termini noti $b_i = \sum_{j=1}^n a_{ij}$

proprietà: la trasposta del prodotto è uguale al prodotto delle trasposte cambiate di ordine

$$(AB)^T = B^T A^T$$

il problema dei minimi quadrati ammette sempre almeno una soluzione
vogliamo minimizzare il quadrato della norma 2

$$\|Ay - b\|_2^2$$

$$\|Q^T z\|_2^2 = (Q^T z)^T Q^T z = z^T Q Q^T z = \|z\|_2^2$$

il quadrato della norma 2 equivale al prodotto scalare

il prodotto scalare è $AA^T = A^T A$

la trasposta del prodotto è uguale al prodotto delle trasposte cambiate di ordine $(AB)^T = B^T A^T$

$QQ^T = I$ poiché Q è ortogonale

MIGLIORE FATTORIZZAZIONE

	PA = LU	A = QR più impegnativa di PA=LU	A = R ^T R Cholesky metà costo di PA=LU	A = USV ^T scomposizione a valori singolari U,V ortogonali non univocamente determinate S univocamente determinata
matrice	A quadrata non singolare generica	A non quadrata, righe > colonne rg(A)= numero colonne (A quadrata non singolare generica se si dispone già di Q e R, se no PA=LU)	A quadrata, non singolare, simmetrica definita positiva	sistemi sovradimensionati di rango massimo o sottodimensionati

matrice singolare matrice quadrata con det=0 o rango non massimo

raggio spettrale max(abs(eig(A)))

norma spettrale o norma 2

$$\|A\|_2 = \sqrt{\rho(A^T A)}$$

vale 0 se e solo se A è la matrice nulla

se A è invertibile, ha tutti gli autovalori $\neq 0$ (è iniettiva)

se A è simmetrica (quadrata) $\|A\|_2 = \sqrt{\rho(A^2)} = \sqrt{\rho^2(A)} = \rho(A)$

poiché $A^T A = A A$ essendo simmetrica

se A è ortogonale $A^T A = I$, $\rho(A) = 1$

se A è simmetrica definita positiva ha tutti gli autovalori > 0

se A è simmetrica, è diagonalizzabile con S ortogonale; se S è

ortogonale $\|S\|_2 \|S^T\|_2 = \sqrt{\rho(S^T S)} = 1$ il calcolo degli autovalori per una matrice A simmetrica è ben condizionato

cerchi di Gershgorin

il centro è l'elemento diagonale della riga i -esima a_{ii} , il raggio è la somma dei moduli degli elementi della stessa riga tranne a_{ii} analogamente si costruiscono i cerchi delle colonne

Teorema 1 tutti gli autovalori della matrice A appartengono all'intersezione dei cerchi riga con i cerchi colonna

Teorema 2 denominata R_1 l'unione di k cerchi riga e R_2 l'unione dei restanti cerchi riga, con R_1 e R_2 disgiunte, k autovalori apparterranno a R_1 , $n-k$ autovalori apparterranno a R_2

Teorema di Bauer-Fike per qualsiasi norma

sia A una matrice diagonalizzabile e sia S invertibile tale che

$S^{-1} A S = D$ dove D è diagonale. \underline{A} è perturbazione della matrice A , $\underline{\lambda}$ è autovalore di \underline{A}

$\min |\underline{\lambda} - \lambda_i| \leq K(S) \|A - \underline{A}\|$

dipende dal numero di condizionamento della **matrice di similitudine**

se A possiede n autovettori linearmente indipendenti $\rightarrow A$ è diagonalizzabile

se tutti gli autovalori di A sono distinti $\rightarrow A$ è diagonalizzabile

metodo delle potenze (converge se si individua un unico autovalore λ di modulo massimo)

suppongo che ci sia un solo autovalore λ di A di modulo massimo (ci possono essere autovalori di A tra loro con lo stesso modulo, strettamente $< |\lambda|$) $\rightarrow \lambda$ è reale perché A è a coefficienti reali e se fosse complesso avrebbe anche il coniugato che avrebbe lo stesso modulo allora pareggerebbe col massimo

quoziente di Rayleigh $\rho = \frac{x^T A x}{x^T x}$

svd(A) restituisce s_1, s_2, \dots, s_n

$K_2(A) = s_1/s_n$

$\|A\|_2 = s_1$

$c = [c_n, \dots, c_0]$

polyval(x,c) restituisce il valore di $c_n x^n + \dots + c_0$

il possibile K è $\geq |y-y|/|x-x|$

nodo i -esimo x_i di Chebyshev nell'intervallo $[a,b]$ per polinomio interpolante di grado n

$$x_i = \frac{b-a}{2} z_i + \frac{b+a}{2}$$

$$z_i = -\cos\left(\frac{(2i-1)\pi}{2(n+1)}\right) \text{ dato dal problema; se } [a,b]=[-1,1] \rightarrow \mathbf{x_i=z_i}$$