

AN410 - Analisi numerica 1

Filippo Rossi

A.A. 2025/2026

Per segnalare errori: rf.rossi.filippo@gmail.com

Appunti presi dalle lezioni e dalle note del corso tenuto dal professore Roberto Ferretti

Indice

1	Prime definizioni e risultati utili	3
1.1	Norma di vettore	3
1.2	Norma di matrice	3
1.3	Numero di condizionamento	3
1.4	Inversa di una matrice 2x2	3
1.5	Costante di Lipschitz	3
1.6	Teorema di Rolle	3
1.7	Teorema del punto fisso di Banach	4
2	Perturbazione sistemi lineari	5
3	Eliminazione di Gauss	7
3.1	Algoritmo	7
3.2	Stabilità	8
3.3	Complessità	8
4	Fattorizzazione LU	10
4.1	Esistenza della fattorizzazione	10
4.2	Algoritmo Doolittle	12
4.3	Algoritmo Cholesky	13
4.4	Matrici inverse	14
4.5	Determinante	14
4.6	Complessità	14

5	Metodi iterativi per sistemi lineari	16
5.1	Introduzione ai metodi iterativi per sistemi lineari	16
5.2	Metodo di Jacobi	16
5.2.1	Algoritmo	16
5.2.2	Complessità	17
5.3	Metodo di Gauss-Seidel	17
5.4	Metodo SOR	18
5.5	Metodo di Richardson	18
6	Metodi iterativi per equazioni nonlineari	20
6.1	Introduzione ai metodi iterativi per equazioni non lineari	20
6.2	Il metodo di bisezione	21
6.3	Metodo di Newton	22
6.4	Metodo delle secanti	23
6.5	Metodo delle corde	23
6.6	Metodo di Steffensen	24
7	Residui, ordini e velocità di convergenza, criteri d'arresto e errori	26
7.1	Residuo	26
7.2	Velocità e ordine di convergenza	26
7.3	Criteri d'arresto	26
7.4	Errori	26
7.4.1	Numero di iterazioni necessarie per errore minore di ϵ	26
7.4.2	Numero di iterazioni necessarie per ottenere una riduzione dell'errore iniziale di un fattore e	27
7.4.3	Stima a priori e fattore $\frac{1}{1-L}$	27
8	Approssimazioni di funzioni di una variabile	28
8.1	Teoria generale dell'interpolazione	28
8.2	Polinomio interpolatore di Lagrange	30
8.3	Polinomio interpolatore di Newton	31
8.4	Strategie di infittimento dei nodi	33
8.4.1	Nodi Chebishev	33
8.4.2	Interpolazioni composite	33
8.5	Convergenza uniforme del polinomio interpolatore	33
8.6	Interpolazione di Hermite	35
8.7	Errore quadratico minimo	36
9	Integrazione numerica	39
9.1	Teoria generale sull'integrazione numerica	39
9.2	Formule di quadratura di Newton-Cotes	41
9.2.1	Formule chiuse	43
9.2.2	Formule aperte	44
9.2.3	Grado di precisione, errore e convergenza	45
9.3	Quadrature Gaussianhe	48

1 Prime definizioni e risultati utili

1.1 Norma di vettore

Sia x un vettore tale che $x = (x_1, \dots, x_n)$.

Norma di Hölder: $\|x\|_\alpha = (\sum_{i=1}^n |x_i|^\alpha)^{\frac{1}{\alpha}}$

Norma 1: $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$

Norma euclidea: $\|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2}$

Norma infinito: $\|x\|_\infty = \max_i |x_i|$

Su \mathbb{R}^n sono tutte equivalenti.

1.2 Norma di matrice

Sia A una matrice.

Norma 1: $\|A\|_1 = \max_j \sum_i |a_{ij}|$

Norma 2: $\|A\|_2 = \rho(A^t A)^{\frac{1}{2}}$ con $\rho = (\text{raggio spettrale}) = \max(|\lambda_i(B)|)$ ¹

Norma infinito: $\|A\|_\infty = \max_i \sum_j |a_{ij}|$

1.3 Numero di condizionamento

Sia A una matrice. Il numero di condizionamento è denotato con $k(A)$ ed è definito come:

$$k(A) = \|A\| \|A^{-1}\|.$$

1.4 Inversa di una matrice 2x2

Sia

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Il determinante di A è:

$$\det(A) = ad - bc$$

Se $\det(A) \neq 0$, allora A è invertibile e la sua inversa è:

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

1.5 Costante di Lipschitz

Per $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ vale che la sua costante di Lipschitz è la norma infinito di J_F : $\|J_F\|_\infty$.

1.6 Teorema di Rolle

Sia $f \in C([a, b])$, derivabile in (a, b) e tale che $f(a) = f(b)$; allora esiste almeno un punto $c \in (a, b)$ tale che $f'(c) = 0$.

¹Se A è simmetrica allora $\|A\|_2 = \rho(A)$.

1.7 Teorema del punto fisso di Banach

Sia (X, d) uno spazio metrico completo e sia $T : X \rightarrow X$ un'applicazione contrattiva, cioè esiste una costante $L \in [0, 1)$ tale che

$$d(T(x), T(y)) \leq L \cdot d(x, y) \quad \forall x, y \in X.$$

Allora T ammette un unico punto fisso $x^* \in X$, ovvero esiste un unico $x^* \in X$ tale che $T(x^*) = x^*$. Inoltre, per ogni $x_0 \in X$, la successione definita iterativamente da

$$x_{n+1} = T(x_n), \quad n = 0, 1, 2, \dots$$

converge a x^* , e vale la stima dell'errore

$$d(x_n, x^*) \leq \frac{L^n}{1 - L} d(x_1, x_0).$$

2 Perturbazione sistemi lineari

Consideriamo un sistema lineare:

$$Ax = b$$

dove $A \in \mathbb{R}^{n \times n}$ è una matrice non singolare, $x \in \mathbb{R}^n$ è il vettore incognito e $b \in \mathbb{R}^n$ è il termine noto.

Teorema 2.1 (Condizionamento rispetto al termine noto). Sia x la soluzione del sistema $Ax = b$ e sia \tilde{x} la soluzione del sistema perturbato $A\tilde{x} = b + \delta b$, dove δb rappresenta una perturbazione del termine noto. Allora vale la seguente stima dell'errore relativo:

$$\frac{\|\delta x\|}{\|x\|} \leq K(A) \frac{\|\delta b\|}{\|b\|}$$

dove:

- $\delta x = \tilde{x} - x$ è l'errore sulla soluzione
- $K(A) = \|A\| \cdot \|A^{-1}\|$ è il numero di condizionamento della matrice A
- $\|\cdot\|$ denota una qualsiasi norma matriciale subordinata a una norma vettoriale

Dimostrazione. Dati:

$$Ax = b \quad (\text{sistema originale})$$

$$A\tilde{x} = b + \delta b \quad (\text{sistema perturbato})$$

Passo 1: Sottraiamo le due equazioni

$$A\tilde{x} - Ax = (b + \delta b) - b$$

$$A(\tilde{x} - x) = \delta b$$

$$A\delta x = \delta b$$

Passo 2: Ricaviamo δx

$$\delta x = A^{-1}\delta b$$

Passo 3: Prendiamo la norma e applichiamo la disuguaglianza triangolare

$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\| \cdot \|\delta b\|$$

Passo 4: Dal sistema originale $Ax = b$ abbiamo:

$$\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|$$

Quindi:

$$\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}$$

Passo 5: Combiniamo i risultati del Passo 3 e del Passo 4

$$\begin{aligned}\frac{\|\delta x\|}{\|x\|} &\leq \|A^{-1}\| \cdot \|\delta b\| \cdot \frac{\|A\|}{\|b\|} \\ &= \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\delta b\|}{\|b\|} \\ &= K(A) \frac{\|\delta b\|}{\|b\|}\end{aligned}$$

□

Il numero di condizionamento $K(A)$ rappresenta il **fattore di amplificazione** dell'errore relativo:

- Se $K(A)$ è piccolo (vicino a 1), il sistema è **ben condizionato**: piccole perturbazioni su b producono piccole variazioni sulla soluzione
- Se $K(A)$ è grande, il sistema è **mal condizionato**: piccole perturbazioni su b possono causare grandi variazioni sulla soluzione

Possiamo fare le seguenti osservazioni:

1. Si ha sempre $K(A) \geq 1$ per ogni matrice non singolare
2. Il risultato vale per qualsiasi norma matriciale subordinata
3. La stima è ottimale: esistono perturbazioni che raggiungono l'uguaglianza
4. Il condizionamento è una proprietà intrinseca della matrice A , indipendente dal termine noto b

3 Eliminazione di Gauss

3.1 Algoritmo

La logica del metodo di eliminazione è di riportare un generico sistema quadrato nella di un sistema triangolare. Per mettere in evidenza i vari passi di eliminazione delle variabili, riscriviamo il sistema come $A^{(1)}x = b^{(1)}$, o per esteso:

$$\begin{cases} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \cdots + a_{1n}^{(1)}x_n = b_1^{(1)} \\ a_{21}^{(1)}x_1 + a_{22}^{(1)}x_2 + \cdots + a_{2n}^{(1)}x_n = b_2^{(1)} \\ \vdots \\ a_{n1}^{(1)}x_1 + a_{n2}^{(1)}x_2 + \cdots + a_{nn}^{(1)}x_n = b_n^{(1)} \end{cases} \quad (1)$$

Si parte dalla eliminazione della variabile x_1 sottraendo alla riga k -esima la prima moltiplicata per il cosiddetto moltiplicatore $m_{k1} = \frac{a_{k1}^{(1)}}{a_{11}^{(1)}}$. Alla fine di $n - 1$ combinazioni lineari così costruite, il sistema sarà nella forma

$$\begin{cases} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \cdots + a_{1n}^{(1)}x_n = b_1^{(1)} \\ a_{22}^{(2)}x_2 + \cdots + a_{2n}^{(2)}x_n = b_2^{(2)} \\ \vdots \\ a_{n2}^{(2)}x_2 + \cdots + a_{nn}^{(2)}x_n = b_n^{(2)} \end{cases} \quad (2)$$

e la variabile x_1 sarà presente solo nella prima equazione. Si riparte quindi dalla variabile x_2 con la stessa modalità (meno che per il fatto che le combinazioni lineari si effettuano sulle righe dalla terza in poi con i moltiplicatori $m_{ki} = \frac{a_{ki}^{(i)}}{a_{ii}^{(i)}}$), e così via. Il risultato è una sequenza di sistemi equivalenti

$$A^{(1)}x = b^{(1)}, \quad A^{(2)}x = b^{(2)}, \quad \dots, \quad A^{(n)}x = b^{(n)},$$

dove l'ultimo dei quali è un sistema triangolare

$$\begin{cases} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \cdots + a_{1n}^{(1)}x_n = b_1^{(1)} \\ a_{22}^{(2)}x_2 + \cdots + a_{2n}^{(2)}x_n = b_2^{(2)} \\ \vdots \\ a_{nn}^{(n)}x_n = b_n^{(n)} \end{cases} \quad (3)$$

di conseguenza possiamo trovare la soluzione usando il metodo delle sostituzioni all'indietro, ovvero:

$$\begin{aligned} x_n &= \frac{\beta_n}{\alpha_{nn}} \\ x_k &= \frac{1}{\alpha_{kk}} \left(\beta_k - \sum_{j=k+1}^n \alpha_{kj}x_j \right) \quad (k = n-1, \dots, 1) \end{aligned} \quad (4)$$

In tutto questo procedimento, si è implicitamente supposto che all' i -esimo passo l'elemento (detto pivot) $a_{ii}^{(i)}$ sia non nullo. Questo non è necessariamente vero; è vero però che se la matrice A è nonsingolare almeno uno tra gli elementi $a_{ki}^{(i)}$ (con $k \geq i$) è non nullo. L'algoritmo viene perciò definito a meno di permutazioni tra la riga i -esima ed una riga successiva.

3.2 Stabilità

È abbastanza intuibile che nel metodo di eliminazione la operazione di combinazione lineare tra righe porti a lungo andare alla perdita di cifre significative per sottrazione, e questo effetto porta alla impossibilità di applicare il metodo di eliminazione in dimensione alta. Moltiplicatori di modulo molto grande causano a loro volta una ulteriore amplificazione di questo errore, che si traduce nel fatto che il sistema che si ottiene dal procedimento di eliminazione è solo approssimativamente triangolare. D'altra parte, nella operazione di prodotto della riga i -esima per il moltiplicatore, gli elementi $a_{ij}^{(i)}$ per $j < i$, se non nulli, si propagano con modulo dipendente dal moltiplicatore. Per aumentare la stabilità dell'algoritmo l'idea è quindi quella di scegliere il pivot in modo che i moltiplicatori risultanti siano "piccoli".

Nella strategia di *pivoting parziale*, che è basata su permutazioni delle righe, al passo i -esimo di eliminazione viene portata in i -esima posizione l'equazione j -esima (con $j \geq i$), dove

$$|a_{ji}^{(i)}| = \max_{k \geq i} |a_{ki}^{(i)}|.$$

Nella strategia di *pivoting totale*, basata su permutazioni sia di righe che di colonne, al passo i -esimo di eliminazione viene portato in posizione di pivot l'elemento $a_{jl}^{(i)}$ (con $j, l \geq i$) per cui si abbia

$$|a_{jl}^{(i)}| = \max_{k, h \geq i} |a_{kh}^{(i)}|.$$

Questo significa portare in i -esima posizione l'equazione j -esima, e scambiare le variabili x_i e x_l (operazione quest'ultima di cui va tenuta opportunamente memoria).

3.3 Complessità

Per prima cosa calcoliamo la complessità della soluzione del sistema triangolare. Per calcolare l' n -esima entrata del vettore soluzione è necessaria una sola operazione (un prodotto), per la $(n-1)$ -esima sono necessarie invece $1+2$ operazioni (una differenza, un prodotto e una divisione). In generale quindi, per il calcolo della k -esima componente, sono necessarie $2n-2k-2$ operazioni ($n-k$ somme, $n-k+1$ prodotti e 1 divisione). Quindi, il numero di operazioni necessarie sono

$$\begin{aligned} C_n &= \sum_{k=1}^n 2n - 2k - 2 \\ &= 2n \sum_{k=1}^n 1 - 2 \sum_{k=1}^n k - 2 \sum_{k=1}^n 1 \\ &= 2n^2 - 2 \frac{n(n+1)}{2} - 2n = O(n^2) \end{aligned}$$

Ora calcoliamo la complessità dell'algoritmo per la triangolarizzazione del sistema. Consideriamo la riduzione della prima colonna, e contiamo le operazioni necessarie per ogni riga. Per la seconda riga sono necessarie $(n+1) + 1 + (n+1) = 2n+3$ operazioni ($n+1$ somme per le entrate della matrici e del vettore, un prodotto per calcolare il moltiplicatore e altre $n+1$ prodotti per le entrate della matrici e del vettore). Visto che questo numero non dipende dalla scelta della riga, allora per ridurre la prima colonna sono necessarie $(2n+3)(n-1) = 2n^2 - n - 3$. Generalizzando quanto detto prima, per ridurre una riga qualsiasi della k -esima colonna, sono necessarie $(n+1-k+1) + 1 + (n+1-k+1) = 2n-2k+5$ operazioni. E quindi per ridurre la k -esima colonna sono necessarie $(2n-2k+5)(n-k) = 2n^2 + 2k^2 - 4nk + 5n - 5k$ operazioni. Di conseguenza, il numero di operazioni per la riduzione di una matrice di ordine n è:

$$\begin{aligned}
C_n &= \sum_{k=1}^{n-1} (2n^2 + 2k^2 - 4nk + 5n - 5k) \\
&= O\left(2n^3 + 2\frac{n^2(2(n-1)+1)}{6} - 2n\frac{n(n+1)}{2}\right) \\
&= O\left(2n^3 + \frac{2}{3}n^3 - 2n^3\right) \\
&= O\left(\frac{2}{3}n^3\right)
\end{aligned}$$

Per quanto riguarda l'operazione di pivoting, la sua complessità è data dal numero di confronti necessari a determinare il nuovo pivot (mentre la complessità della operazione di scambio di righe/colonne è sempre lineare). Per ogni eliminazione, tale numero è lineare in n per il pivoting parziale e quadratico per il pivoting totale. Ne risulta un incremento di complessità che è quadratico (e quindi asintoticamente trascurabile) per il pivoting parziale, e cubico (quindi non trascurabile asintoticamente) per il pivoting totale.

Teorema 3.1 (Esistenza della soluzione per il metodo di eliminazione). Se la matrice A del sistema (1.5) è nonsingolare, esiste una permutazione delle equazioni per cui questo algoritmo può essere completato. Il vettore che si ottiene dalla sostituzione all'indietro (1.4) è la unica soluzione del sistema (1.5).

Dimostrazione. Basta osservare che ogni passo del processo di eliminazione trasforma il sistema in un sistema equivalente. D'altra parte, questa equivalenza non sussisterebbe se non fosse possibile trovare ad ogni passo un pivot non nullo (infatti, in questo caso da un sistema ben posto se ne sarebbe ottenuto un altro senza soluzione o con infinite soluzioni). \square

Teorema 3.2 (Necessità delle permutazioni di righe). Se tutti i minori principali di A sono nonsingolari, in particolare se la matrice A è a diagonale dominante o definita positiva, allora l'algoritmo di eliminazione può essere completato senza permutazione delle righe.

4 Fattorizzazione LU

4.1 Esistenza della fattorizzazione

Dal metodo di eliminazione discende la possibilità di fattorizzare (a meno di permutazioni di righe) la matrice A nel prodotto di due matrici triangolari, L triangolare inferiore e U triangolare superiore. Si può quindi scrivere, in modo chiaramente non unico,

$$A = LU$$

e porre il sistema (1.5) nella forma

$$Ax = LUx = b.$$

Introducendo una variabile ausiliaria z , la soluzione del sistema (1.5) si ottiene quindi dalla successiva soluzione dei due sistemi triangolari $Lz = b$ e $Ux = z$.

Nel caso in cui sia necessaria una permutazione P di righe, o si utilizzi una fattorizzazione pivotata, si ha

$$PA = LU$$

ed i due sistemi triangolari da risolvere sono $Lz = Pb$ e $Ux = z$. Di questa permutazione di righe occorre perciò tenere memoria nel caso in cui si risolvano più sistemi lineari con la stessa matrice A ma con diversi termini noti.

Teorema 4.1. Consideriamo il problema $Ax = b$, con $A \in \text{Mat}_{n \times n}(\mathbb{R})$ matrice non singolare e $b \in \mathbb{R}^n$. Allora esistono due matrici $L, U \in \text{Mat}_{n \times n}(\mathbb{R})$ con L triangolare inferiore, U triangolare superiore e una matrice di permutazione $P \in \text{Mat}_{n \times n}(\mathbb{R})$ tale che $PA = LU$. Inoltre, se $P = \mathbb{1}_n$, L assume la seguente forma:

$$l_{ki} = \begin{cases} \frac{a_{ki}^{(i)}}{a_{ii}^{(i)}} & \text{se } k > i \\ 1 & \text{se } k = i \\ 0 & \text{se } k < i \end{cases}$$

dove con $a_{lt}^{(i)}$ indichiamo l'elemento nel posto di riga l e colonna t della matrice $A(i)$, che é la matrice ottenuta dalla matrice A^{i-1} sostituendo la riga r_e con la riga $r_e - \frac{a_{ei}^{(i-1)}}{a_{ii}^{(i-1)}} r_i$ per ogni $e \in \{i+1, \dots, n\}$ e con passo base $A^{(0)} := A$.

Dimostrazione. La matrice di permutazione P viene applicata cosí da scambiare le righe in modo tale che, effettuando l'eliminazione di Gauss, si ottenga sempre un pivot non nullo. Quindi, a meno di sostituire A con se stessa con le righe permutate, possiamo cercare direttamente le matrici L, U tali che $A = LU$. Consideriamo quindi le seguenti matrici

trasformanti:

$$T_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & -\frac{a_{i+1,i}^{(i)}}{a_{i,i}^{(i)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{a_{n,i}^{(i)}}{a_{i,i}^{(i)}} & \cdots & 1 \end{pmatrix}$$

con $i \in \{1, \dots, n-1\}$. Si può osservare che, applicando le matrici trasformanti T_i alla matrice A , si ottiene una matrice triangolare superiore U tale che la soluzione di $Ax = b$ è uguale a $Ux = b$ (infatti applicare le tutte le matrici trasformanti equivale a ridurre a scala la matrice tramite il metodo di eliminazione di Gauss). Vale quindi:

$$\left(\prod_{i=1}^{n-1} T_{n-i} \right) A = U$$

per una matrice U triangolare superiore. Di conseguenza vale:

$$A = \left(\prod_{i=1}^{n-1} T_i \right) U$$

Definiamo quindi $L := \left(\prod_{i=1}^{n-1} T_i \right)$, e osservando che L è prodotto di matrici triangolari inferiori, abbiamo che L è triangolare inferiore.

Mostriamo ora che L ha la forma presente nell'enunciato. Definiamo

$$m_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{a_{i+1,i}^{(i)}}{a_{i,i}^{(i)}} \\ \vdots \\ \frac{a_{n,i}^{(i)}}{a_{i,i}^{(i)}} \end{pmatrix}$$

Per prima cosa osserviamo che vale $T_i^{-1} = \mathbb{1}_n + m_i e_i$, infatti:

$$\begin{aligned} T_i T_i^{-1} &= (\mathbb{1}_n - m_i e_i)(\mathbb{1}_n + m_i e_i) \\ &= \mathbb{1}_n - m_i e_i + m_i e_i - m_i e_i m_i e_i = \mathbb{1}_n \end{aligned}$$

dove abbiamo usato che $e_i m_i = 0_n$. Ora per induzione mostriamo che $\left(\prod_{i=1}^k T_i\right) = \mathbb{1}_n + \sum_{i=1}^k m_i e_i$: il passo base é ovvio, quindi supponiamo vero per $k-1$ e mostriamolo per k :

$$\begin{aligned}\prod_{i=1}^k T_i &= \left(\prod_{i=1}^{k-1} T_i\right) T_k \\ &= \left(\mathbb{1}_n + \sum_{i=1}^{k-1} m_i e_i\right) (\mathbb{1}_n + m_k e_k) \\ &= \mathbb{1}_n + \sum_{i=1}^k m_i e_i + \sum_{i=1}^k m_i e_i m_k e_k \\ &= \mathbb{1}_n + \sum_{i=1}^k m_i e_i\end{aligned}$$

dove abbiamo di nuovo che $e_i m_i = 0_n$. Questo conclude la dimostrazione. \square

4.2 Algoritmo Doolittle

Un'altra possibilità è di utilizzare la formula di prodotto tra L e U ,

$$a_{ij} = \sum_k l_{ik} u_{kj}$$

con le condizioni

$$\begin{aligned}l_{ik} &= 0 \quad (k > i) \\ u_{kj} &= 0 \quad (k > j) \\ l_{ii} &= 1\end{aligned}$$

(quest'ultima condizione porta ad un risultato univocamente determinato, che coincide con la cosiddetta fattorizzazione di Doolittle per la quale si ha $U = A^{(n)}$). Partendo dalla prima riga di A si ha:

$$a_{1j} = \sum_k l_{1k} u_{kj} = l_{11} u_{1j} = u_{1j}$$

da cui si ottiene $u_{1j} = a_{1j}$ e quindi tutta la prima riga di U . Passando poi alla prima colonna di A :

$$a_{i1} = \sum_k l_{ik} u_{k1} = l_{i1} u_{11}$$

e poiché l'elemento u_{11} è stato già calcolato in precedenza, si ottiene per $i \geq 2$:

$$l_{i1} = \frac{a_{i1}}{u_{11}}.$$

Dalla seconda riga di A si ha:

$$a_{2j} = \sum_k l_{2k} u_{kj} = l_{21} u_{1j} + l_{22} u_{2j} = l_{21} u_{1j} + u_{2j}$$

e quindi, per $j \geq 2$,

$$u_{2j} = a_{2j} - l_{21}u_{1j},$$

mentre considerando la seconda colonna di A si ha analogamente

$$a_{i2} = \sum_k l_{ik}u_{k2} = l_{i1}u_{12} + l_{i2}u_{22}$$

da cui si ottiene per $i \geq 3$:

$$l_{i2} = \frac{1}{u_{22}}(a_{i2} - l_{i1}u_{12}).$$

L'algoritmo continua in questo modo, alternativamente ottenendo per $j \geq p$ dalla p -esima riga di A :

$$u_{pj} = a_{pj} - \sum_{k < p} l_{pk}u_{kj}, \quad (5)$$

e per $i > q$ dalla q -esima colonna di A :

$$l_{iq} = \frac{1}{u_{qq}} \left(a_{iq} - \sum_{k < q} l_{ik}u_{kq} \right). \quad (6)$$

La necessità di riordinare le righe della matrice A appare nel caso in cui in (1.9) compaia un pivot u_{qq} nullo. Esistono varianti pivotate di questo algoritmo.

4.3 Algoritmo Cholesky

Analogamente a quanto si è fatto per la fattorizzazione di Doolittle ponendo $l_{ii} = 1$, si possono ottenere altri tipi di fattorizzazione fissando in modo diverso gli elementi sulle diagonal. Con la scelta $u_{ii} = 1$ si ottiene la cosiddetta fattorizzazione di Crout, mentre scegliendo $l_{ii} = u_{ii}$ (si dimostra che ciò è possibile nelle matrici definite positive) si ottiene la fattorizzazione di Cholesky, che pone A nella forma

$$A = LL^t.$$

Con l'idea di calcolare la matrice L per gradini (quindi in ordine $l_{11}, l_{21}, l_{22}, l_{31}, \dots$), si ottiene dalla p -esima riga di A :

$$l_{pp} = \left(a_{pp} - \sum_{k < p} l_{pk}^2 \right)^{\frac{1}{2}}, \quad (7)$$

e per $i > p$:

$$l_{ip} = \frac{1}{l_{pp}} \left(a_{ip} - \sum_{k < p} l_{ik}l_{pk} \right). \quad (8)$$

In questo caso non serve utilizzare le colonne di A , vista la sua simmetria.

$$l_{pp} = \left(a_{pp} - \sum_{k < p} l_{pk}^2 \right)^{\frac{1}{2}} \leq a_{pp}^{\frac{1}{2}}, \quad (9)$$

4.4 Matrici inverse

Tramite la fattorizzazione LU di una matrice PA è possibile calcolare la matrice inversa di A , infatti:

$$PA = LU \iff A = P^{-1}LU \iff A^{-1} = (P^{-1}LU)^{-1} = U^{-1}L^{-1}P$$

Nel caso in cui la matrice A non necessita della pivotazione, vale

$$A^{-1} = U^{-1}L^{-1}$$

4.5 Determinante

Anche per il determinante abbiamo una formula, infatti, ricordando che il determinante di una matrice triangolare superiore è il prodotto degli elementi sulla diagonale, vale:

$$\det(A) = \det(P^{-1}LU) = \det(P^{-1})\det(L)\det(U) = (-1)^s \det(L)\det(U) = (-1)^s \prod_k l_{kk} \prod_k u_{kk}$$

dove s è il numero di permutazioni associate alla matrice di permutazione P . Inoltre se stiamo procedendo con la fattorizzazione di Doolittle vale

$$\det(A) = (-1)^s \prod_k u_{kk}$$

4.6 Complessità

Per calcolare la complessità dell'algoritmo per la costruzione delle matrici L e U bisogna contare il numero di operazioni effettuate per costruire gli elementi delle matrici.

Cominciamo con U . Consideriamo la p -esima riga di U : per calcolare l'elemento u_{pj} sono necessarie $p-2$ somme e $p-1$ prodotti per calcolare la sommatoria, e infine bisogna fare un'altra somma per il termine di A fuori dalla sommatoria. Quindi per l'elemento u_{pj} abbiamo bisogno di $p-2 + p-1 + 1 = 2p-2 = 2(p-1)$ operazioni. Di conseguenza, visto che il numero di operazioni dipende solo dalla riga in cui si trova l'elemento e che nella p -esima riga ci sono $n-p+1$ elementi non nulli, il numero di operazioni necessarie per calcolare gli elementi della p -esima riga sono $2(p-1)(n-p+1)$. Di conseguenza, la complessità computazionale per calcolare U sarà

$$\begin{aligned} C_n &= \sum_{p=1}^n 2(p-1)(n-p+1) = 2 \sum_{p=1}^n (np - p^2 + p - n + p - 1) \\ &= 2n \sum_{p=1}^n p - 2 \sum_{p=1}^n p^2 + 4 \sum_{p=1}^n p - 2n \sum_{p=1}^n 1 - 2 \sum_{p=1}^n 1 \\ &= 2n \frac{n(n+1)}{2} - 2 \frac{n(n+1)(2n+1)}{6} + 4 \frac{n(n+1)}{2} - 2n^2 - 2n \\ &= O\left(n^3 - \frac{2}{3}n^3\right) \\ &= O\left(\frac{n^3}{3}\right) \end{aligned}$$

Ora, la complessità dell'algoritmo per calcolare gli elementi di L è asintoticamente uguale a quella calcolata per gli elementi di U . Quindi la complessità totale sarà:

$$O\left(\frac{n^3}{3}\right) + O\left(\frac{n^3}{3}\right) = O\left(\frac{2}{3}n^3\right)$$

Quindi avrà lo stesso costo del metodo di eliminazione di Gauss. D'altra parte, occorre ricordare che nel metodo di eliminazione questo costo è richiesto anche quando pur restando fissa la matrice A , viene cambiato il vettore dei termini noti. In questa eventualità, nel metodo di fattorizzazione LU non occorre fattorizzare di nuovo la matrice ed il costo computazionale resta solo quello della soluzione dei due sistemi triangolari (ovvero $O(2n^2)$). Quindi facendo il conto per due sistemi lineari con stessa matrice A avremo:

$$\begin{array}{ll} \text{per Gauss:} & 2O\left(\frac{2}{3}n^3\right) + 2O(n^2) = O\left(\frac{2}{6}n^3\right) \\ \text{per Doolittle:} & O\left(\frac{2}{3}n^3\right) + 2O(n^2) = O\left(\frac{2}{3}n^3\right) \end{array}$$

Quindi conviene la fattorizzazione LU .

Nel caso di sistemi con matrice definita positiva, il metodo di Cholesky (che ha complessità $O(n^3/3)$ poiché si deve calcolare una sola matrice) è invece conveniente anche nella soluzione di un singolo sistema lineare.

Infine il calcolo dell'inversa, richiedendo prima la fattorizzazione e poi la soluzione di $2n$ sistemi triangolari, ha una complessità globale di $O(8n^3/3)$ operazioni in virgola mobile.

5 Metodi iterativi per sistemi lineari

5.1 Introduzione ai metodi iterativi per sistemi lineari

Consideriamo sempre il problema $Ax = b$, risolvere il problema in modo iterativo consiste nel porre il problema in una equazione di punto fisso, ovvero considerare un problema della forma $x = Bx + c$ che abbia le stesse soluzioni del problema iniziale. La matrice B si chiama matrice di iterazione.

Ci sono vari metodi per i sistemi lineari, e partono tutti dalla seguente decomposizione della matrice A :

$$A = D + E + F$$

con

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$
$$E = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix}$$
$$F = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Osserviamo ora che vale il seguente criterio necessario e sufficiente per ogni metodo iterativo per sistemi lineari:

Teorema 5.1. Siano A e B matrici come sopra, allora per ogni x_0 la successione $x_{k+1} = Bx_k + c$ é convergente se e solo se il raggio spettrale soddisfa la condizione

$$\rho(B) < 1$$

5.2 Metodo di Jacobi

5.2.1 Algoritmo

Si può osservare che:

$$\begin{aligned} Ax = b &\iff (D + E + F)x = b \\ &\iff Dx = b - (E + F)x \\ &\iff x = D^{-1}b - D^{-1}(E + F)x \end{aligned}$$

Quindi si può considerare il sistema iterativo:

$$x_j^{(k)} = \frac{1}{a_{jj}} \left(b_j - \sum_{i \neq j} a_{ji} x_i^{(k-1)} \right)$$

che ha come matrice iterativa $B_J = D^{-1}(E + F)$.

Teorema 5.2. Consideriamo il problema $Ax = b$, con $A \in \text{Mat}_{n \times n}(\mathbb{R})$ matrice non singolare e $b \in \mathbb{R}^n$. Se A è strettamente a diagonale dominante, allora il metodo di Jacobi converge all'unica soluzione di $Ax = b$.

Dimostrazione. Lo schema iterativo di Jacobi ha la forma:

$$T(x) = D^{-1}b - D^{-1}(E + F)x$$

quindi vale che $T'(x) = J_T = -D^{-1}(E + F)$. Quindi per mostrare che lo schema converga alla soluzione basta mostrare che in una norma qualsiasi vale $\|J_T\| < 1$. Usando la norma infinito abbiamo:

$$\|J_T\|_\infty = \max_{i \in \{1, \dots, n\}} \left\{ \sum_{j \neq i}^n \left| \frac{a_{ij}}{a_{ii}} \right| \right\}$$

Ma visto che la matrice è strettamente a diagonale dominante vale per ogni $i \in \{1, \dots, n\}$ che $\sum_{j \neq i}^n \left| \frac{a_{ij}}{a_{ii}} \right| < 1$. Quindi abbiamo che $\|J_T\|_\infty < 1$ e la convergenza dello schema. \square

5.2.2 Complessità

Calcoliamo ora la complessità computazionale dell'algoritmo. Per calcolare la j -esima componente del vettore soluzione sono necessarie $(n - 2) + (n - 1) + 1 + 1 = 2n - 1$ operazioni ($n - 2$ somme dalla sommatoria, $n - 1$ prodotti dalla sommatoria, una somma con b_j e un prodotto con $\frac{1}{a_{jj}}$). Quindi la complessità computazionale dell'algoritmo è

$$C_n = \sum_{k=1}^n 2n - 1 = O(2n^2)$$

Il conto appena fatto è valido per tutti i metodi iterativi seguenti per sistemi lineari.

5.3 Metodo di Gauss-Seidel

Si può osservare che:

$$\begin{aligned} Ax = b &\iff (D + E + F)x = b \\ &\iff (D + E)x = b - Fx \\ &\iff x = (D + E)^{-1}b - (D + E)^{-1}Fx \end{aligned}$$

Abbiamo potuto invertire $D+E$ poiché triangolare inferiore, e quindi per forza non singolare. Ponendo il problema in questa forma otteniamo il seguente schema:

$$x_j^{(k+1)} = \frac{1}{a_{jj}} \left(b_j - \sum_{i>j} a_{ji}x_i^{(k)} - \sum_{i<j} a_{ji}x_i^{(k+1)} \right)$$

Da un punto di vista della programmazione, questo metodo é piú efficiente di quello di Jacobi poiché permette di lavorare con un unico vettore di soluzioni.

Teorema 5.3. Consideriamo il problema $Ax = b$, con $A \in \text{Mat}_{n \times n}(\mathbb{R})$ matrice non singolare e $b \in \mathbb{R}^n$. Se A é strettamente a diagonale dominante o definita positiva, allora il metodo di Gauss-Seidel converge all'unica soluzione di $Ax = b$.

5.4 Metodo SOR

Fissando un $w \in \mathbb{R}$ i può osservare che

$$\begin{aligned} Ax = b &\iff (D + E + F)x = b \\ &\iff w(D + E + F)x = wb \\ &\iff (wD + wE + wF)x = wb \\ &\iff (D - D + wD + wE + wF)x = wb \\ &\iff [D - D(1 - w) + wE + wF]x = wb \\ &\iff (D + wE)x = wb - [wF - D(1 - w)]x \\ &\iff x = (D + wE)^{-1}wb + (D + wE)^{-1}[D(1 - w) - wF]x \end{aligned}$$

Quindi si può considerare il sistema iterativo:

$$x^{(k+1)} = (1 - w)x^{(k)} + wx_{GS}^{(k+1)}$$

dove $x_{GS}^{(k+1)}$ é la soluzione $k+1$ -esima ottenuta con il metodo di Gauss-Seidel. Si può osservare che se $w = 0$ il metodo non fa aggiornamenti (quindi converge alla soluzione solamente se si parte dalla soluzione) mentre se $w = 1$ allora il metodo coincide con quello di Gauss-Seidel.

Teorema 5.4. Se una matrice A é definita positiva, allora la successione del meotdo SOR converge all'unica soluzione del sistema lineare.

5.5 Metodo di Richardson

Fissando un $\beta \in \mathbb{R}$ si può osservare che

$$\begin{aligned} Ax = b &\iff \beta Ax = \beta b \\ &\iff x - x + \beta Ax = \beta b \\ &\iff [x - (\mathbb{1}_n - \beta A)]x = \beta b \\ &\iff x = \beta b + (\mathbb{1}_n - \beta A)x \end{aligned}$$

Quindi si può considerare il sistema iterativo:

$$x^{(k+1)} = x - \beta (Ax^{(k)} - b)$$

Teorema 5.5. Consideriamo il problema $Ax = b$, con $A \in \text{Mat}_{n \times n}(\mathbb{R})$ matrice non singolare e $b \in \mathbb{R}^n$. Se A definita positiva e $\beta > 0$, allora il metodo di Richardson converge all'unica soluzione di $Ax = b$ se e solo se $\beta \in \left(0, \frac{2}{\lambda_{\max}(A)}\right)$, dove $\lambda_{\max}(A)$ è l'autovalore più grande di A .

Dimostrazione. Sappiamo che un metodo iterativo converge se e solo se $\rho(B) < 1$, dove B è la matrice di iterazione del sistema iterativo. Sia ora $\lambda(A)$ un autovalore di A , quindi possiamo trovare il corrispondente autovalore di B nel seguente modo:

$$\lambda(B) = \lambda(\mathbb{1}_n - \beta A) = 1 - \beta \lambda(A)$$

Ora, vogliamo risolvere $\rho(B) < 1$, e quindi dobbiamo trovare β tale che:

$$|1 - \beta \lambda(A)| < 1$$

che corrisponde a risolvere:

$$-1 < 1 - \beta \lambda(A) < 1$$

Visto che β è positivo, e che gli autovalori sono positivi (poiché la matrice è definita positiva) allora la disuguaglianza $1 - \beta \lambda(A) < 1$ è sempre soddisfatta. Per l'altra invece basta osservare che:

$$-1 < 1 - \beta \lambda(A) \iff \beta < \frac{2}{\lambda(A)}$$

Ora l'ultima condizione, è vera per $\lambda(A) = \lambda_{\max}(A)$ se e solo se è vera per ogni altro autovalore, quindi abbiamo ottenuto la tesi. \square

6 Metodi iterativi per equazioni nonlineari

6.1 Introduzione ai metodi iterativi per equazioni non lineari

Consideriamo una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$, vogliamo creare una successione reale che converga a una soluzione dell'equazione

$$f(x) = 0$$

Per farlo vogliamo trasformare il problema sopra in uno della forma

$$g(x) = x$$

Teorema 6.1. Sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione e sia $\bar{x} \in \mathbb{R}$ tale che $g(\bar{x}) = \bar{x}$. Supponiamo inoltre che esista un $m \in \mathbb{N}_{\geq 1}$ e un intorno $U \in N(\bar{x})$ tale che $g \in C^{m+1}(U)$ (o che sia prolungabile ad una con questa regolarità). Se $g^{(i)}(\bar{x}) = 0$ per ogni $i \in \{1, \dots, m\}$ e $g^{(m+1)} \neq 0$ allora esiste un $x_0 \in U$ tale che la successione definita iterativamente da

$$x_{k+1} := g(x_k)$$

converge a \bar{x} con ordine $m + 1$.

Dimostrazione. A meno di sostituire g con il suo prolungamento liscio, possiamo supporre che $g \in C^{m+1}(U)$. Visto che $g'(\bar{x}) = 0$ e che g' è una funzione continua, dal teorema di permanenza del segno abbiamo che esiste un intorno $W \in N(\bar{x})$ tale che $f'(x) < 1$ per ogni $x \in W$. Quindi, per ogni $x \in W \cap U$ vale che $f'(x) < 1$, ovvero è una contrazione. Usando ora il teorema di Banach e scegliendo $x_0 \in W \cap U$ abbiamo che il metodo iterativo converge a \bar{x} .

Ora mostriamo l'ordine di convergenza. Osserviamo che

$$x_{k+1} - \bar{x} = g(x_k) - \bar{x} = g(x_k) - g(\bar{x}) \quad (10)$$

Ora, sviluppando la funzione g in $\bar{x} \in W \cap U$ (che è un intorno di \bar{x} dove per ipotesi $g \in C^{m+1}$) usando il resto di Lagrange, per ogni $x \in W \cap U$ esisterà un $c_x \in W \cap U$ tale che:

$$g(x) = g(\bar{x}) + g'(\bar{x})(x - \bar{x}) + \dots + \frac{g^{(m)}(\bar{x})}{m!}(x - \bar{x})^m + \frac{g^{(m+1)}(c_x)}{(m+1)!}(x - \bar{x})^{m+1} = g(\bar{x}) + \frac{g^{(m+1)}(c_x)}{(m+1)!}(x - \bar{x})^{m+1}$$

Passando ai moduli l'equazione appena trovata, e valutando le funzioni in un generico x_k otteniamo

$$g(x_k) = g(\bar{x}) + \frac{g^{(m+1)}(c_x)}{(m+1)!}(x_k - \bar{x})^{m+1}$$

Inserendo ora quanto appena trovato nella (14) e passando ai moduli otteniamo

$$|x_{k+1} - \bar{x}| = \left| g(\bar{x}) + \frac{g^{(m+1)}(c_x)}{(m+1)!}(x_k - \bar{x})^{m+1} - g(\bar{x}) \right| = \left| \frac{g^{(m+1)}(c_x)}{(m+1)!} \right| |x_k - \bar{x}|^{m+1}$$

Ora, visto che $g \in C^{m+1}$, a meno di restringerci ulteriormente su un intorno $I \subset U \cap W^2$, possiamo supporre (usando il teorema di Weierstrass) l'esistenza del massimo $m := \max_{x \in I} \left\{ \left| \frac{g^{(m+1)}(c_x)}{(m+1)!} \right| \right\}$. Otteniamo quindi la disuguaglianza

$$|x_{k+1} - \bar{x}| \leq m |x_k - \bar{x}|^{m+1}$$

e quindi la convergenza del metodo ha ordine $m + 1$ □

6.2 Il metodo di bisezione

Consideriamo una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ continua tale che esistono $a, b \in \mathbb{R}$ per cui vale $f(a)f(b) < 0$. Definiamo allora il seguente algoritmo:

INIZIO: $a_0 := a$ e $b_0 = b$;

FOR: $i = 0, \dots, n$: Sia $c_i := \frac{a_i + b_i}{2}$;

IF $f(c_i) == 0$: STOP;

IF $f(a_i)f(c_i) < 0$: $a_{i+1} := a_i$ e $b_{i+1} := c_i$;

IF $f(b_i)f(c_i) < 0$: $a_{i+1} := c_i$ e $b_{i+1} := b_i$;

Teorema 6.2. Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua tale che esistono $a, b \in \mathbb{R}$ per cui vale $f(a)f(b) < 0$. Allora le successioni $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ e $(c_n)_{n \in \mathbb{N}}$ convergono ad una soluzione di $f(x) = 0$.

Dimostrazione. Possiamo osservare che, per ogni $n \in \mathbb{N}$, vale

$$a_0 \leq a_n \leq b_0, \quad a_0 \leq a_n \leq b_0$$

quindi entrambe le successioni sono limitate. Inoltre abbiamo che

$$a_0 \leq a_{n+1}, \quad b_{n+1} \leq b_0$$

quindi le successioni sono anche monotone, di conseguenza ammettono limite. D'altro canto:

$$b_n - a_n = \frac{b_0 - a_0}{2^n} \xrightarrow{n \rightarrow \infty} 0$$

quindi vale $\lim_n a_n = \lim_n b_n$, e visto che $a_n \leq c_n \leq b_n$, per il teorema del confronto vale $\lim_n a_n = \lim_n c_n = \lim_n b_n$. Supponiamo ora che $f(a) > 0$ e che $f(b) < 0$ (la dimostrazione del caso opposto é analoga), di conseguenza abbiamo che $f(a_n) \geq 0$ e che $(b_n) \leq 0$ per ogni $n \in \mathbb{N}$. Usando quanto appena detto e il teorema ponte all'uguaglianza sopra (possiamo applicarlo perché f é continua) vale:

$$0 \leq \lim_n f(a_n) = \lim_n f(c_n) = \lim_n f(b_n) \leq 0$$

quindi $\lim_n f(c_n) = 0$. Infine, usando di nuovo il teorema ponte, vale

$$f(\lim_n c_n) = 0$$

Il che dimostra la tesi. □

²Questo lo possiamo fare poiché abbiamo mostrato che il metodo converge per ogni $x \in U \cap W$.

6.3 Metodo di Newton

Nel metodo di Newton si pone

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

ponendo quindi $x_{k+1} = g(x_k)$. Questa scelta equivale a scegliere x_{k+1} come lo zero della retta tangente a f nel punto $(x_k, f(x_k))$. Infatti la retta tangente a f nel punto $(x_k, f(x_k))$ ha equazione

$$y - f(x_k) = f'(x_k)(x - x_k)$$

imponendo quindi il passaggio della retta per $(x_{k+1}, 0)$ otteniamo

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Ovviamente dobbiamo richiedere che $f'(\bar{x}) \neq 0$, in modo tale che almeno in un intorno la disequazione si mantenga e così che la successione sia ben definita. In generale la derivata potrebbe essere difficile da calcolare, quindi in certi casi, fissando un parametro h , si può sostituire $f'(x_k)$ con il rapporto incrementale ottenendo dunque:

$$x_{k+1} = x_k - \frac{h}{f(x_k + h) - f(x_k)} f(x_k)$$

Teorema 6.3. Sia $f \in C^2(A)$ per qualche $A \in N(\bar{x})$ dove \bar{x} è tale che $f(\bar{x}) = 0$. Se $f'(\bar{x}) \neq 0$ allora esiste un $x_0 \in A$ tale che la successione $x_{k+1} = g(x_k)$ del metodo di Newton converge a \bar{x} con ordine quadratico.

Dimostrazione. Daremo la dimostrazione supponendo che $f \in C^2(A)$. Usiamo il teorema generale (??). Abbiamo che

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

D'altra parte vale che

$$g'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}$$

e quindi $g'(\bar{x}) = 0$. Visto che $f \in C^3(A)$ allora $g \in C^2(A)$ e usando il teorema (??) abbiamo tesi. \square

Vediamo ora un teorema alternativo del metodo di Newton.

Teorema 6.4. Sia $f \in C^2(A)$ per qualche $A \in N(\bar{x})$ dove \bar{x} è tale che $f(\bar{x}) = 0$ ma $f'(\bar{x}) \neq 0$. Sia ora $x_0 \in A$, tale che se valgono una delle seguenti:

- $x_0 > \bar{x}$, f crescente e convessa in $[\bar{x}, x_0]$;
- $x_0 > \bar{x}$, f decrescente e concava in $[\bar{x}, x_0]$;

- $x_0 < \bar{x}$, f decrescente e convessa in $[x_0, \bar{x}]$;
- $x_0 < \bar{x}$, f crescente e concava in $[x_0, \bar{x}]$;

allora la successione del metodo di Newton converge a \bar{x} .

Dimostrazione. Mostriamo solo il primo caso, gli altri sono analoghi. A meno di prendere x_0 La successione é

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

dalle ipotesi abbiamo che $f'(x_k) > 0$ (quindi la successione é ben definita in $[\bar{x}, x_0]$) e che $f(x_k) \geq 0$, quindi vale che

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \leq x_k$$

e quindi la successione $(x_k)_{k \in \mathbb{N}}$ é decrescente. Mostriamo che la successione é anche limitata. Visto x_{k+1} é lo zero della retta tangente a f nel punto $(x_k, f(x_k))$, e che la funzione é convessa in $[\bar{x}, x_0]$, la retta stará sotto il grafico di f in $[\bar{x}, x_0]$. Di conseguenza deve valere $\bar{x} \leq x_k$ per ogni $k \in \mathbb{N}$. Esiste quindi $y \in \mathbb{R}$ limite della successione, in simboli $\lim_k x_k = y$. Applicando il limite alla formula di ricorsione abbiamo

$$y = y - \lim_k \frac{f(x_k)}{f'(x_k)}$$

e quindi

$$\lim_k \frac{f(x_k)}{f'(x_k)} = 0$$

che porta alla condizione $\lim_k f(x_k) = 0$ e usando il teorema ponte abbiamo $f(y) = 0$. Visto che la funzione in $[\bar{x}, x_0]$ é crescente esiste una sola radice, e cist che y é in questo intervallo abbiamo che $y = \bar{x}$. \square

6.4 Metodo delle secanti

Nel metodo delle secanti si definisce la successione nel seguente modo

$$x_{k+1} = x - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})}$$

6.5 Metodo delle corde

Teorema 6.5. Sia A un aperto tale che $f \in C^1(A)$ e tale che $f(\bar{x}) = 0$, $\bar{x} \in A$ e $f'(\bar{x}) \neq 0$. Consideriamo il metodo iterativo

$$g(x) = x - \frac{b - a}{f(b) - f(a)} f(x)$$

che prende il nome di *metodo delle corde*. Allora esistono $a, b, x_0 \in A$ con $a \leq x_0 \leq b$ tali che la successione $x_{k+1} = g(x_k)$ converge a \bar{x} .

Dimostrazione. Fissiamo $a, b \in A$. Sia $g(x) = x - \frac{b-a}{f(b)-f(a)}f(x)$, quindi otteniamo

$$g'(x) = 1 - \frac{b-a}{f(b)-f(a)}f'(x)$$

Dal teorema di Lagrange, per ogni $x \in [a, b]$ esiste $c \in [a, b]$ tale che:

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

quindi abbiamo che:

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

mandando ora $x, a, b \rightarrow \bar{x}$, e di conseguenza anche c vista la sua dipendenza da x da a e da b , otteniamo:

$$\lim_{x, a, b \rightarrow \bar{x}} g'(x) = \frac{f'(\bar{x}) - f'(\bar{x})}{f'(\bar{x})} = 0$$

dove le ultime uguaglianze sono vere poiché \bar{x} è una radice semplice per f . Quindi esiste un intorno $W \subset A$ di \bar{x} tale che $g'(x) < 1$ per ogni $x \in W$. Quindi, riscegliendo $a, b \in W$ e applicando il teorema di Banach, abbiamo la tesi. \square

Consideriamo ora le stesse ipotesi del teorema precedente ma con $f \in C^2[y, z]$. Allora possiamo riutilizzare il teorema di Lagrange e ottenere per ogni $x \in W$ l'esistenza di un $d \in W$ tale che:

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

di conseguenza vale che:

$$\begin{aligned} \sup_{x \in W} |g'(x)| &= \sup_{x \in W} \left| \frac{f'(c) - f'(x)}{f'(c)} \right| \\ &= \sup_{x \in W} \left| \frac{f''(d)(c - x)}{f'(c)} \right| \\ &= \frac{\sup_{x \in W} |c - x| \sup_{x \in W} |f''(d)|}{\inf_{x \in W} |f'(x)|} \\ &\leq \frac{|W| \sup_{x \in W} |f''(d)|}{\inf_{x \in W} |f'(x)|} \end{aligned}$$

6.6 Metodo di Steffensen

Il metodo di Steffensen si basa sulla funzione

$$g(x) = x - \frac{f(x)}{f(x + f(x)) - f(x)}f(x)$$

Funzione simile a quelle precedenti ma con la differenza di aver usato co e incremento $f(x)$ la funzione stessa.

Teorema 6.6. Se $f \in C^2(A)$, con $\bar{x} \in A$ dove $f(\bar{x}) = 0$, $f'(\bar{x}) \neq 0$ ed x_0 è sufficientemente vicino a \bar{x} , allora esiste un indice k finito tale che $x_k = \bar{x}$, oppure la successione x_k definita dal metodo di Steffensen converge con ordine quadratico alla soluzione \bar{x} .

Dimostrazione. Mostriamo che per la funzione g vale il teorema generale sui metodi iterativi (???). Usando il teorema di Lagrange otteniamo l'esistenza di un $c \in A$ (che dipende da x) tale che

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

Visto che $c \in (x, f(x) + x)$ (oppure $x \in (f(x) + x, x)$ se vale che $x \geq f(x) + x$), e visto che se $x \rightarrow \bar{x}$ allora $f(x) + x \rightarrow \bar{x}$, abbiamo

$$\lim_{x \rightarrow \bar{x}} g(x) = \bar{x} - \lim_{x \rightarrow \bar{x}} \frac{f(x)}{f'(c)} = \bar{x} - \frac{f(\bar{x})}{f'(\bar{x})} = \bar{x}$$

Mostriamo ora che $\lim_{x \rightarrow \bar{x}} g'(x) = 0$. Usando quanto trovato prima con il teorema di Lagrange otteniamo:

$$\begin{aligned} g'(x) &= 1 - \frac{2f(x)f'(x)[f(x + f(x)) - f(x)] - f(x)^2[f'(x + f(x))(1 + f'(x)) - f'(x)]}{(f(x + f(x)) - f(x))^2} \\ &= 1 - \frac{2f(x)f'(x)[f(x + f(x)) - f(x)] - f(x)^2[f'(x + f(x))(1 + f'(x)) - f'(x)]}{f(x)^2 f'(c)^2} \\ &= 1 - \frac{2f(x)f'(x)[f(x + f(x)) - f(x)]}{f(x)^2 f'(c)^2} + \frac{f(x)^2[f'(x + f(x))(1 + f'(x)) - f'(x)]}{f(x)^2 f'(c)^2} \\ &= 1 - \frac{2f'(x)f'(c)}{f'(c)^2} + \frac{[f'(x + f(x))(1 + f'(x)) - f'(x)]}{f'(c)^2} \\ &\xrightarrow{x \rightarrow \bar{x}} 1 - 2 + \frac{[f'(\bar{x})(1 + f'(\bar{x})) - f'(\bar{x})]}{f'(\bar{x})^2} = 0 \end{aligned}$$

Dove abbiamo usato che per $x \rightarrow \bar{x}$, vale $c \rightarrow x$ (come abbiamo osservato sopra) e che $f(\bar{x}) = 0$. \square

7 Residui, ordini e velocità di convergenza, criteri d'arresto e errori

7.1 Residuo

Il *residuo al passo k* , non é nient'altro che

$$\|f(x_k) - f(\bar{x})\| = \|f(x_k)\|$$

quindi si conta, quanto vicino al valore della soluzione si é arrivati.

7.2 Velocità e ordine di convergenza

Considerando sempre che per un sistema iterativo valga

$$\|x_k - \bar{x}\| \leq L\|x_{k-1} - \bar{x}\|^\gamma$$

allora diremo che L *il coefficiente di contrazione* (misura quanto rapidamente l'errore va a 0) e che γ *é l'ordine di convergenza* (quanto velocemente converge il sistema alla soluzione). Osserviamo anche che, a parità di ordine, se due sistemi iterativi A, B hanno rispettivamente $t < u$ come coefficienti di contrazione, allora il sistema A , oltre che a ridurre l'errore prima, convergerà piú velocemente di B .

7.3 Criteri d'arresto

Per tutti i metodi iterativi sopra descritti, ci si può porre la seguente domanda:

”Quando devo fermarmi con le iterazioni?”

In generale si sceglie un razionale $\epsilon > 0$ e si fermano le iterazioni quando vale $\|x_k - x_{k-1}\| < \epsilon$.

7.4 Errori

Supponiamo ora che L sia la costante tale che, dato un metodo iterativo $(x_k)_{k \in \mathbb{N}}$, vale

$$\|x_k - \bar{x}\| \leq L\|x_{k-1} - \bar{x}\|$$

allora abbiamo la disuguaglianza

$$\|x_k - \bar{x}\| \leq L\|x_{k-1} - \bar{x}\| \leq \dots \leq L^k\|x_0 - \bar{x}\|$$

7.4.1 Numero di iterazioni necessarie per errore minore di ϵ

Vista la disuguaglianza sopra, basta imporre

$$\|x_k - \bar{x}\| \leq L^k\|x_0 - \bar{x}\|$$

e risolvere per k .

7.4.2 Numero di iterazioni necessarie per ottenere una riduzione dell'errore iniziale di un fattore e

Quindi, supponendo di voler calcolare il numero di iterazioni necessarie per ottenere una riduzione dell'errore iniziale di un fattore e , allora vogliamo trovare un k per il quale valga:

$$\|x_k - \bar{x}\| \leq e\|x_0 - \bar{x}\|$$

Sappiamo anche però che vale

$$\|x_k - \bar{x}\| \leq L^k \|x_0 - \bar{x}\| \quad (11)$$

Potremmo allora imporre

$$e \leq L^k$$

e risolvere la disequazione trovando k . A questo punto avremmo trovato un k per cui vale la relazione

$$e\|x_0 - \bar{x}\| \leq L^k \|x_0 - \bar{x}\|$$

ma dalla (11) l'errore $\|x_k - \bar{x}\|$ potrebbe trovarsi benissimo nell'intervallo

$$(e\|x_0 - \bar{x}\|, L^k \|x_0 - \bar{x}\|)$$

e quindi avremmo trovato un k per cui non è detto valga la condizione chiesta.

Se invece imponiamo

$$L^k \leq e$$

allora possiamo trovare un k tale che valga

$$L^k \|x_0 - \bar{x}\| \leq e\|x_0 - \bar{x}\|$$

e a questo punto, visto che dalla (11) abbiamo che l'errore $\|x_k - \bar{x}\| \leq L^k \|x_0 - \bar{x}\|$, avremmo trovato un k per il quale si verifica la riduzione richiesta.

Che forma ha questa L ? In generale per i sistemi iterativi di sistemi lineari, L è la norma della matrice iterativa del sistema; più in generale invece, dato un metodo iterativo con funzione g definito su intervallo $[a, b]$, allora $L = \sup_{x \in [a, b]} \{|g'(x)|\}$.

7.4.3 Stima a priori e fattore $\frac{1}{1-L}$

Dal teorema del punto fisso di Banach sappiamo che vale la stima dell'errore

$$\|x_k - \bar{x}\| \leq \frac{L^k}{1-L} \|x_1 - x_0\|.$$

Questa stima è particolarmente utile perché non richiede la conoscenza della soluzione esatta \bar{x} , ma solo la distanza tra le prime due iterate. Il fattore $\frac{1}{1-L}$ amplifica la stima iniziale e diventa tanto più grande quanto L si avvicina a 1: questo riflette il fatto che se la costante di contrazione è vicina a 1, la convergenza è più lenta e l'errore si riduce meno rapidamente. Possiamo inoltre ottenere una stima a posteriori utilizzando la distanza tra iterate consecutive:

$$\|x_k - \bar{x}\| \leq \frac{L}{1-L} \|x_k - x_{k-1}\|.$$

Quest'ultima formula è utile nella pratica per stimare l'errore corrente senza conoscere \bar{x} .

8 Approssimazioni di funzioni di una variabile

8.1 Teoria generale dell'interpolazione

Definizione 8.1. Data una funzione f e x_0, \dots, x_n $n+1$ punti tutti diversi tra loro (detti *nodi dell'interpolazione*) un polinomio $\Pi_n(x)$ di grado al più n (a volte anche indicato con $p_n(x)$) tale che

$$\Pi_n(x_i) = f(x_i), \quad \forall i \in \{0, \dots, n\}$$

é chiamato *polinomio interpolatore*.

Piú avanti vedremo che il polinomio esiste sempre (e ne esibiremo anche due forme) ed è unico.

Teorema 8.2 (Convergenza delle interpolazioni in una base generica.). Si supponga che $f, \phi_0, \dots, \phi_n \in C^0([a, b])$, e che $\phi_i(x_j) = \delta_{ij}$. Denotando con X_n lo spazio generato dalla base $\{\phi_0, \dots, \phi_n\}$, con

$$\varepsilon_n(f) = \inf_{p_n \in X_n} \sup_{x \in [a, b]} |f(x) - p_n(x)| = \min_{p_n \in X_n} \max_{x \in [a, b]} |f(x) - p_n(x)| = \max_{x \in [a, b]} |f(x) - p_n^*(x)|,$$

e con

$$\Lambda_n = \max_{x \in [a, b]} \sum_{i=0}^n |\phi_i(x)|,$$

la si ha:

$$\max_{x \in [a, b]} \left| f(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| \leq (1 + \Lambda_n) \varepsilon_n(f).$$

Dimostrazione. Abbiamo che $p_n^*(x) \in X_n$: visto che $\{\phi_0, \dots, \phi_n\}$ è una base per lo spazio vettoriale X_n , che $\phi_i(x_j) = \delta_{ij}$ e che la scrittura in combinazione lineare di un elemento (fissata una base) è unica allora abbiamo che

$$p_n^*(x) = \sum_{i=0}^n p_n^*(x_i) \phi_i(x)$$

Sia ora $x \in [a, b]$, quindi otteniamo:

$$\begin{aligned}
\left| f(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| &= \left| f(x) - p_n^*(x) + p_n^*(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| \\
&\leq |f(x) - p_n^*(x)| + \left| p_n^*(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| \\
&\leq \epsilon_n(f) + \left| \sum_{i=0}^n p_n^*(x_i) \phi_i(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| \\
&\leq \epsilon_n(f) + \left| \sum_{i=0}^n (p_n^*(x_i) - f(x_i)) \phi_i(x) \right| \\
&\leq \epsilon_n(f) + \left| \sum_{i=0}^n (p_n^*(x_i) - f(x_i)) \phi_i(x) \right| \\
&\leq \epsilon_n(f) + \sum_{i=0}^n |(p_n^*(x_i) - f(x_i)) \phi_i(x)| \\
&\leq \epsilon_n(f) + \epsilon_n(f) \sum_{i=0}^n |\phi_i(x)| = \left(1 + \sum_{i=0}^n |\phi_i(x)| \right) \epsilon_n(f)
\end{aligned}$$

Passando al massimo sull'ultima relazione ottenuta, abbiamo la tesi. \square

Definizione 8.3. La costante $\Lambda_n = \max_{x \in [a, b]} \sum_{i=0}^n |\phi_i(x)|$ del teorema prima, prende il nome di *Costante di Lebesgue*.

Teorema 8.4 (Unicità del polinomio interpolatore). Data una funzione f e x_0, \dots, x_n $n+1$ punti tutti diversi tra loro, allora il polinomio interpolatore è unico. In particolare, se f stessa è un polinomio allora vale $f = \Pi_n$.

Dimostrazione. Siano ora a e b due polinomi interpolatori per i nodi nell'enunciato, quindi sono polinomi di grado n e che, per ogni $i \in \{0, \dots, n\}$, rispettano le condizioni

$$a(x_i) = b(x_i) = f(x_i)$$

Consideriamo ora il polinomio $a - b$; questo è ancora di grado non superiore ad n ma ha $n+1$ zeri. Di conseguenza, dal teorema fondamentale dell'algebra, segue che $a - b = 0$. Quindi i polinomi coincidono. \square

Teorema 8.5 (Errore di approssimazione). Siano $x_0 < \dots < x_n$, $n+1$ punti distinti e definiamo $I := [x_0, x_n]$. Se $f \in C^{n+1}(I)$ allora per ogni $x \in I$ esiste un $\xi \in I$ tale che vale

$$f(x) - \Pi_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x)$$

dove $\omega_n(x) = \prod_{i=0}^n (x - x_i)$ e Π_n è il polinomio interpolatore costruito sui nodi x_0, \dots, x_n .

Dimostrazione. Fissiamo $x \in I$ ³ e possiamo assumere per simmetria che $x \in [x_0, x_1]$; definiamo la funzione

$$\begin{aligned} G: I &\rightarrow \mathbb{R} \\ z &\mapsto f(z) - \Pi_n(z) - Q(x)\omega_n(z) \end{aligned}$$

dove Q é una funzione tale che $G(x) = 0$. Quindi tale che vale:

$$f(x) - \Pi_n(x) = Q(x)\omega_n(x)$$

Ora mostriamo che Q ha la forma che compare nella tesi del teorema, e per farlo usiamo il teorema di Rolle. Iniziamo con osservare che, avendo fissato x , $Q(x)$ é una costante e quindi G é combinazione lineare di funzioni in $C^{n+1}(I)$. Di conseguenza vale che $G \in C^{n+1}(I)$; inoltre abbiamo che $\{x, x_0, \dots, x_n\}$ é un insieme di cardinalitá $n + 2$ che contiene gli zeri della funzione G . Abbiamo quindi

$$G(x) = G(x_0) = \dots = G(x_n) = 0$$

quindi possiamo applicare il teorema di Rolle all'intervallo $[x_0, x]$ e assumere quindi l'esistenza di un $\xi_1 \in [x_0, x]$ tale che $G'(\xi_1) = 0$. Possiamo ora applicare il teorema di Rolle anche sull'intervallo $[x, x_1]$ e ottenere un $\xi_2 \in [x, x_1]$ tale che $G'(\xi_2) = 0$. Cosí facendo otteniamo che la funzione G' ha $n + 1$ zeri in I e inoltre $G' \in C^n(I)$. Ripetendo questo procedimento altre n volte otteniamo che esiste un $\xi \in I$ tale che $G^{(n+1)}(\xi) = 0$. Ma osservando che $\Pi_n^{(n+1)} = 0$ e che $\omega_n^{(n+1)} = (n + 1)!$, allora abbiamo

$$0 = G^{(n+1)}(\xi) = f^{(n+1)}(\xi) - Q(x)(n + 1)!$$

Quindi otteniamo che

$$Q(x) = \frac{f^{(n+1)}(\xi)}{(n + 1)!}$$

□

8.2 Polinomio interpolatore di Lagrange

Teorema 8.6. Data una funzione f e x_0, \dots, x_n $n + 1$ punti tutti diversi tra loro, allora il polinomio

$$\Pi_n(x) = \sum_{i=0}^n f(x_i)L_i(x)$$

dove

$$L_i(x) = \begin{cases} 1 & \text{se } i = n = 0 \\ \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} & \text{se } n > 0 \end{cases}$$

soddisfa l'equazione $\Pi_n(x_i) = f(x_i)$ per ogni $i \in \{0, \dots, n + 1\}$.

³Visto che se $x \in \{x_0, \dots, x_n\}$ l'asserito é banalmente vero, possiamo supporre che $x \notin \{x_0, \dots, x_n\}$.

Dimostrazione. Se $n = 0$ abbiamo $\Pi = f$, quindi l'asserto é banalmente vero. Supponiamo quindi $n > 0$. Mostriamo che in generale vale la formula

$$L_i(x_k) = \delta_{ik}$$

Se $i \neq k$ vale

$$L_i(x_k) = \prod_{j \neq i} \frac{x_k - x_j}{x_i - x_j} = \frac{x_k - x_k}{x_i - x_k} \cdot \prod_{j \notin \{i, k\}} \frac{x_k - x_j}{x_i - x_j} = 0$$

quindi abbiamo $L_i(x_k) = 0$. D'altronde, se $i = k$ allora

$$L_i(x_k) = \prod_{j \neq i} \frac{x_k - x_j}{x_i - x_j} = \prod_{j \neq k} \frac{x_k - x_j}{x_k - x_j} = 1$$

Quindi ora, valutando Π_n in un x_k e usando il risultato appena ottenuto, abbiamo

$$\Pi_n(x_k) = \sum_{i=0}^n f(x_i) L_i(x_k) = \sum_{i=0}^n f(x_i) \delta_{ik} = f(x_k)$$

□

8.3 Polinomio interpolatore di Newton

Lemma 8.7 (Formula di Neville). Sia f una funzione e x_0, \dots, x_n $n + 1$ punti tutti diversi tra loro; allora vale la seguente formula per il polinomio interpolatore p_n :

$$p_n(x) = \frac{(x - x_0)p_{1\dots n}(x) - (x - x_n)p_{0\dots n-1}(x)}{x_n - x_0} \quad 4$$

Dimostrazione. Per cominciare, definiamo

$$q(x) := \frac{(x - x_0)p_{1\dots n}(x) - (x - x_n)p_{0\dots n-1}(x)}{x_n - x_0}$$

Osserviamo poi che, per ogni $i \in \{1, \dots, n - 1\}$, vale che

$$q(x_i) = \frac{(x_i - x_0)p_{1\dots n}(x_i) - (x_i - x_n)p_{0\dots n-1}(x_i)}{x_n - x_0} = \frac{(x_i - x_0)f(x_i) - (x_i - x_n)f(x_i)}{x_n - x_0} = f(x_i)$$

D'altra parte, nel nodo x_0 , vale

$$q(x_0) = -\frac{(x_0 - x_n)p_{0\dots n-1}(x_0)}{x_n - x_0} = -\frac{(x_0 - x_n)f(x_0)}{x_n - x_0} = f(x_0)$$

e analogamente, nel nodo x_n vale

$$q(x_n) = \frac{(x_n - x_0)p_{1\dots n}(x_n)}{x_n - x_0} = \frac{(x_n - x_0)f(x_n)}{x_n - x_0} = f(x_n)$$

Osservando che $\deg(q) \leq n$ e ricordando che il polinomio interpolatore é unico, vale che $p_n = q$. □

⁴Con $p_{k\dots l}$ indichiamo il polinomio interpolatore costruito sui nodi $\{x_k, \dots, x_l\}$.

Teorema 8.8. Data una funzione f e x_0, \dots, x_n $n + 1$ punti tutti diversi tra loro, allora il polinomio

$$\Pi_n(x) = \sum_{i=0}^n \left(f[x_0, \dots, x_i] \prod_{k=0}^{i-1} (x - x_k) \right)$$

dove

$$\begin{aligned} f[x_0] &= f(x_0) \\ f[x_0, \dots, x_k] &= \frac{f[x_1, \dots, x_k] - f[x_0, \dots, x_{k-1}]}{x_k - x_0} \end{aligned}$$

soddisfa l'equazione $\Pi_n(x_i) = f(x_i)$ per ogni $i \in \{0, \dots, n\}$. Le costanti $f[x_0, \dots, x_i]$ prendono il nome di differenze divise.

Dimostrazione. Mostriamo per induzione su n che il polinomio interpolatore p_n coincide con Π_n , una volta fatto questo la tesi é automatica. Dalla definizione del polinomio vale che $\Pi_0(x_0) = f(x_0)$, quindi $p_0 = \Pi_0$. Supponiamo quindi la tesi vera per $n - 1$ e mostriamola per n . Definiamo ora

$$g := p_n - p_{n-1} = p_n - \Pi_{n-1}$$

Necessariamente, visto che $\deg(p_n) \leq n$ e $\deg(\Pi_{n-1}) \leq n - 1$, deve valere $\deg(g) \leq n$. Inoltre, per ogni $i \in \{0, \dots, n - 1\}$ deve valere

$$g(x_i) := p_n(x_i) - \Pi_{n-1}(x_i) = 0$$

quindi ho trovato n radici per il polinomio g che é di grado al piú n , quindi per il teorema fondamentale dell'algebra esiste una costante c tale che

$$g(x) = c \cdot \prod_{i=0}^{n-1} (x - x_i)$$

Visto che g é di grado n e che Π_{n-1} é di grado al piú $n - 1$ deduciamo che p_n é un polinomio di grado n . Di conseguenza, la costante c coincide con il coefficiente di grado n di p_n . Se valesse che $c = f[x_0, \dots, x_n]$ allora avremmo che $p_n = g + \Pi_{n-1} = \Pi_n$, e quindi avremmo anche la tesi. Mostriamo quindi che $c = f[x_0, \dots, x_n]$. Osserviamo che, per il lemma sopra, vale

$$p_n(x) = \frac{(x - x_0)p_{1\dots n}(x) - (x - x_n)p_{0\dots n-1}(x)}{x_n - x_0}$$

e che per ipotesi induttiva vale

$$\begin{aligned} p_{1\dots n} &= \Pi_{1\dots n} \\ p_{0\dots n-1} &= \Pi_{0\dots n-1} \end{aligned}$$

⁶ quindi, visto che la costante c corrisponde al coefficiente di grado massimo di p_n , dovrà valere

$$c = \frac{f[x_1, \dots, x_n] - f[x_0, \dots, x_{n-1}]}{x_n - x_0}$$

Di conseguenza vale che $p_n = \Pi_n$. □

⁵La seconda uguaglianza é vera per ipotesi induttiva.

⁶Dove con $\Pi_{m\dots l}$ indichiamo il polinomio nell'enunciato definito sui punti x_m, \dots, x_l .

8.4 Strategie di infittimento dei nodi

8.4.1 Nodi Chebishev

In questa strategia, i nodi vengono posizionati negli zeri di un polinomio di Chebishev ⁷ di grado $n + 1$, vediamo due metodi:

- *Chebyshev-Gauss*: Si definiscono i nodi nel modo seguente:

$$x_j = a + \frac{b-a}{2} \left(1 - \cos \left(\frac{(2j+1)\pi}{2n+2} \right) \right)$$

e corrispondono proprio agli zeri del polinomio di Chebishev di grado $n + 1$ nell'intervallo $[a, b]$. Si può mostrare che questa precisa scelta dei nodi (tra tutte quelle possibili) è quella che porta al minimo valore del $\sup_{[a,b]} |\omega_n(x)|$ nel caso in cui gli estremi interpolazione non siano vincolati ad essere nodi dell'interpolazione.

- *Chebyshev-Gauss-Lobatto*: In questo caso invece la forma dei nodi è

$$a + \frac{b-a}{2} \left(1 - \cos \left(\frac{j\pi}{n} \right) \right)$$

8.4.2 Interpolazioni composite

Suddividendo l'intervallo $[a, b]$ in sottointervalli $[a_j, b_j]$ tutti di ampiezza H_j in modo tale che ogni intervallo contenga $n + 1$ nodi di interpolazione si ottiene una interpolazione di grado n a tratti. Assumendo anche che due degli $n + 1$ nodi dentro ciascun intervallo siano agli estremi, allora avremmo anche la continuità della funzione approssimante. In caso contrario, la continuità potrebbe fallire proprio nei bordi degli intervalli.

8.5 Convergenza uniforme del polinomio interpolatore

Teorema 8.9 (Convergenza uniforme per nodi generici). Sia $f : I \rightarrow \mathbb{R}$ con $I = [a, b]$. Se è soddisfatta la condizione

$$\sup_{x \in I} |f^{(n+1)}(x)| = o \left(\frac{(n+1)!}{(b-a)^{n+1}} \right)$$

allora per $n \rightarrow \infty$ il polinomio interpolatore $\Pi_n \rightarrow f(x)$ uniformemente in I per ogni scelta arbitraria dei nodi.

Dimostrazione. Ricordiamo che dalla stima sull'errore di approssimazione 8.5 abbiamo che $f(x) - \Pi_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x)$. Per mostrare la tesi ora basta osservare che

$$\begin{aligned} \sup_{x \in I} |f(x) - \Pi_n(x)| &= \sup_{x \in I} \left| \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x) \right| = \sup_{x \in I} |f^{(n+1)}(\xi)| \left| \frac{\omega_n(x)}{(n+1)!} \right| \\ &\leq \sup_{x \in I} |f^{(n+1)}(\xi)| \frac{(b-a)^{n+1}}{(n+1)!} \end{aligned}$$

⁷Con polinomio di Chebishev di grado n si intende un polinomio che nell'intervallo $[0, 1]$ assume la forma $T_n(x) = \cos(n \arccos x)$. Se consideriamo un intervallo $[a, b]$ generico, allora il polinomio di Chebishev assume la forma $\cos(n \arccos(\frac{2t-(a+b)}{b-a}))$ (abbiamo applicato la traslazione $x = \frac{2t-(a+b)}{b-a}$ con $t \in [a, b]$).

Dove l'ultima disuguaglianza e' vera poiché $|x_i - x| \leq (b - a)$ per ogni $i \in \{0, \dots, n\}$ e per ogni $x \in I$. Usando le ipotesi e mandando n a infinito allora otteniamo che la successione Π_n di polinomi interpolatori convergono uniformemente a $f(x)$ in I . \square

Teorema 8.10 (Convergenza uniforme per nodi di Chebishev). Se $f \in C^k(I)$ per un intervallo I e con $k \in \mathbb{N}_{\geq 1}$, allora il polinomio interpolatore $\Pi_n(x)$ costruito sui nodi di Chebishev converge uniformemente a f per $n \rightarrow \infty$. Inoltre esiste un $C \in \mathbb{R}$ tale che vale la stima $|f(x) - \Pi_n(x)| \leq Cn^{-k}$ per ogni $x \in I$.

Teorema 8.11 (Convergenza uniforme per approssimazioni composite). Se $f \in C^{n+1}(I)$, con I intervallo reale, allora l'interpolazione composita $\Pi_{n,H}$ di grado n a tratti sui k sottointervalli $[a_j, b_j]$ ciascuno con ampiezza $H_j \leq H$ converge uniformemente a f ⁸. Per di piu' vale la stima

$$\sup_{x \in I} |f(x) - \Pi_{n,H}(x)| \leq \frac{1}{(n+1)!} \sup_{j \in \{0, \dots, k\}} \left(H_j^{n+1} \sup_{x \in [a_j, b_j]} |f^{(n+1)}(x)| \right)$$

In particolare vale anche

$$\sup_{x \in I} |f(x) - \Pi_{n,H}(x)| \leq \frac{\sup_{x \in I} |f^{(n+1)}(x)|}{(n+1)!} H^{n+1}$$

Dimostrazione. Ricordiamo che dalla stima sull'errore di approssimazione 8.5 abbiamo che per ogni $x \in [a_j, b_j]$ esistono per ogni $j \in \{1, \dots, k\}$ degli $\xi_j \in [a_j, b_j]$ tali che $f(x) - \Pi_{n,j}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_{n,j}(x)$ ⁹. Per mostrare la tesi ora basta osservare che

$$\begin{aligned} \sup_{x \in I} |f(x) - \Pi_{n,H}(x)| &= \sup_{j \in \{0, \dots, k\}} \sup_{x \in [a_j, b_j]} \left| \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_{n,j}(x) \right| \\ &= \frac{1}{(n+1)!} \sup_{j \in \{0, \dots, k\}} \sup_{x \in [a_j, b_j]} |f^{(n+1)}(\xi) \omega_{n,j}(x)| \\ &\leq \frac{1}{(n+1)!} \sup_{j \in \{0, \dots, k\}} \sup_{x \in [a_j, b_j]} |f^{(n+1)}(\xi) H_j^{n+1}| \\ &= \frac{1}{(n+1)!} \sup_{j \in \{0, \dots, k\}} \left(H_j^{n+1} \sup_{x \in [a_j, b_j]} |f^{(n+1)}(\xi)| \right) \end{aligned}$$

Dove l'ultima disuguaglianza e' vera poiché $|x_{i_j} - x| \leq (b_j - a_j)$ per ogni $i \in \{0, \dots, n\}$, $j \in 0, \dots, k$ e per ogni $x \in [a_j, b_j]$. Usando le ipotesi e mandando n a infinito allora otteniamo che la successione Π_n converge uniformemente a $f(x)$ in I . Per la seconda stima basta usare l'ipotesi di $H_j \leq H$. \square

⁸In norma infinito per $H \rightarrow 0$

⁹Con $\Pi_{n,j}$ indichiamo il polinomio interpolatore nel sottointervallo $[a_j, b_j]$ e con $\omega_{n,j}$ indichiamo il polinomio della forma $\prod_{i=0}^n (x - x_{i_j})$ dove gli x_{i_j} sono i nodi per l'interpolazione nell'intervallo $[a_j, b_j]$.

8.6 Interpolazione di Hermite

Nella interpolazione di Hermite, si usa in modo combinato la strategia del polinomio di Taylor e quella vista sopra. Infatti, fissata f una funzione, $m_i \in \mathbb{N}$ e dei nodi x_i , si impone che il polinomio di Hermite $H_m(x)$ passi per i nodi x_i e che le derivate coincidano nel punto, quindi si chiede che:

$$H_m^{(p)}(x_i) = f^{(p)}(x_i) \quad \forall p \in \{0, \dots, m_i - 1\}$$

Piú precisamente, vale il seguente teorema:

Teorema 8.12 (Esistenza del polinomio interpolatore di Hermite). Sia $f : [a, b] \rightarrow \mathbb{R}$, $I := \{0, \dots, n\}$. Consideriamo per ogni $i \in I$ dei nodi $x_i \in [a, b]$ tali che f ammetta la derivata $(m_i - 1)$ -esima. Allora esiste un polinomio $L_{ik}(x)$ tale che

$$L_{ik}^{(p)}(x_j) = \begin{cases} 1, & \text{se } i = j \text{ e } p = k \\ 0 & \text{altrimenti} \end{cases}$$

Si definisce quindi il polinomio di Hermite come il polinomio seguente

$$H_m(x) = \sum_{i=0}^n \sum_{k=0}^{m_i-1} f^{(k)}(x_i) L_{ik}(x)$$

dove $m = \sum_{i \in I} m_i - 1$. Inoltre, esiste una formula per ricorsione per calcolare i polinomi L_{ik} ed é la seguente:

$$\begin{aligned} L_{i, m_i-1}(x) &= l_{i, m_i-1}(x) \\ L_{ik}(x) &= l_{ik}(x) - \sum_{p=k+1}^{m_i-1} l_{ik}^{(p)}(x_i) L_{ip}(x) \quad (k = m_i - 2, \dots, 0), \end{aligned}$$

dove

$$l_{ik}(x) = \frac{(x - x_i)^k}{k!} \prod_{j \neq i} \left(\frac{x - x_j}{x_i - x_j} \right)^{m_j} \quad (k = 0, \dots, m_i - 1).$$

Anche per questo tipo di interpolazione esiste una stima dell'errore:

Teorema 8.13 (Rappresentazione dell'errore). Se $I = [\min(x, x_0, \dots, x_n), \max(x, x_0, \dots, x_n)]$ e se $f \in C^{m+1}(I)$, allora l'errore di approssimazione si può rappresentare come:

$$f(x) - H_m(x) = \frac{f^{(m+1)}(\xi)}{(m+1)!} \Omega_m(x) \quad (5.25)$$

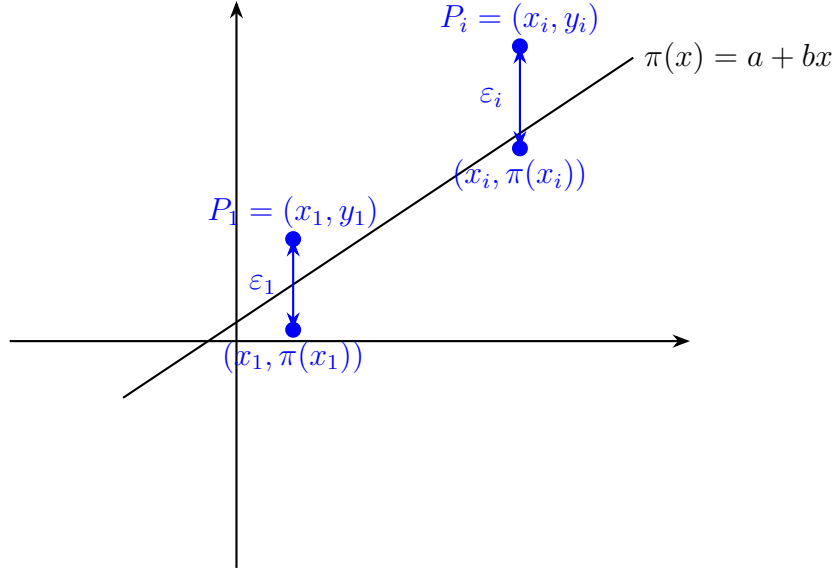
dove

$$\Omega_m(x) = (x - x_0)^{m_0} (x - x_1)^{m_1} \dots (x - x_n)^{m_n}, \quad \xi \in I.$$

Osservazione 8.14. Prendendo nel teorema sopra, $x_0 = \dots = x_n$ otteniamo proprio il resto di Lagrange. La cosa non ci meraviglia poiché, nell'ipotesi data, il polinomio di Hermite coincide esattamente con il polinomio di Taylor.

8.7 Errore quadratico minimo

Esempio 8.15.



Fissiamo dei punti $p_i := (x_i, y_i)$ con $i \in \{0, \dots, m\}$. Vogliamo trovare a, b tali che le distanze ϵ_i assumano il valore minimo tra quelli possibili al variare di a e b . Quindi vogliamo minimizzare la funzione:

$$\begin{aligned} r(a, b) &= \sum_{i=1}^m (\pi(x_i) - y_i)^2 = \sum_{i=1}^m (a + bx_i - y_i)^2 \\ &= \sum_{i=1}^m a^2 + b^2 x_i^2 + y_i^2 + 2abx_i - 2ay_i - 2bx_i y_i \\ &= ma^2 + b^2 \sum_{i=1}^m x_i^2 + \sum_{i=1}^m y_i^2 + 2ab \sum_{i=1}^m x_i - 2a \sum_{i=1}^m y_i - 2b \sum_{i=1}^m x_i y_i \end{aligned}$$

Per trovare i punti critici imponiamo le condizioni

$$\begin{aligned} 0 &= \frac{\partial r}{\partial a} = 2ma + 2b \sum_{i=1}^m x_i - 2 \sum_{i=1}^m y_i \\ 0 &= \frac{\partial r}{\partial b} = 2b \sum_{i=1}^m x_i^2 + 2a \sum_{i=1}^m x_i - 2 \sum_{i=1}^m x_i y_i \end{aligned}$$

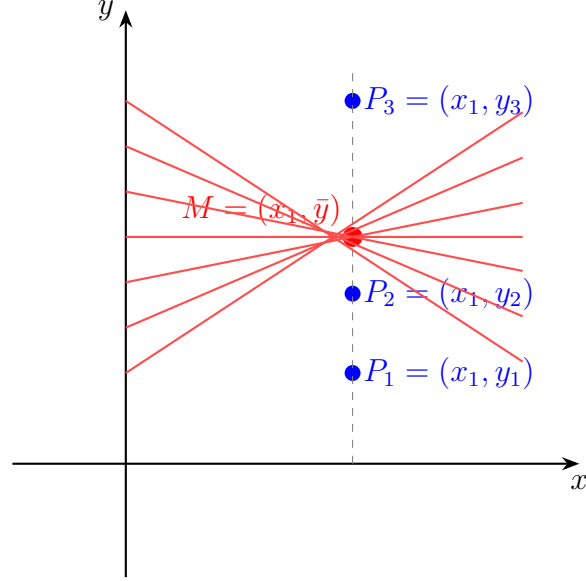
Questo é un sistema, e se dividiamo per 2 otteniamo la matrice associata al sistema

$$A = \begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix}$$

Quindi i punti critici (x, y) sono le soluzioni del sistema:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix}$$

É giusto osservare che ci sono casi in cui A é singolare, ad esempio scegliendo i punti P_i della forma (x_i, y_0) (quindi tutti con la stessa ordinata) l'unica condizione che si impone é che le rette devono passare per il punto medio della retta verticale che passa per i punti. Di conseguenza esisteranno infinite rette che passano per quel punto (cfr. figura)



Ora generalizziamo l'esempio sopra: siano $p_i = (x_i, y_i)$ m punti e consideriamo questa volta $\pi(x) = \sum_{i=1}^n a_i \phi_i(x)$ con $\{\phi_i\}_{i \in \{1, \dots, n\}}$ base dell'interpolazione scelta. Definendo

$$\Phi = \begin{pmatrix} \phi_1(x_1) & \dots & \phi_n(x_1) \\ \vdots & & \vdots \\ \phi_1(x_m) & \dots & \phi_n(x_m) \end{pmatrix}$$

e $a = (a_1, \dots, a_n)^T$ abbiamo

$$\Phi a = \begin{pmatrix} \sum_{i=1}^n a_i \phi_i(x_1) \\ \vdots \\ \sum_{i=1}^n a_i \phi_i(x_m) \end{pmatrix}$$

definiamo la funzione r come

$$\begin{aligned} r(a) &= \|\Phi a - y\|^2 \\ &= (\Phi a - y)^T (\Phi a - y) \\ &= (\Phi a)^T (\Phi a) - (\Phi a)^T y - y^T (\Phi a) + y^T y \\ &= a^T \Phi^T \Phi a - a^T \Phi^T y - y^T \Phi a + y^T y \\ &= a^T \Phi^T \Phi a - 2a^T \Phi^T y + y^T y \end{aligned}$$

dove abbiamo usato che $a^T \Phi^T y = y^T \Phi a$ poiché due scalari ¹⁰. Ora il gradiente sarà

$$\nabla(r)(a) = 2\Phi^T \Phi a - 2\Phi^T y$$

Imponendo uguale a zero il gradiente otteniamo la condizione

$$\Phi^T \Phi a = \Phi^T y$$

¹⁰Infatti vale $a^T \Phi^T y = (y^T \Phi a)^T = y^T \Phi a$ dove abbiamo usato che la trasposta di uno scalare é se stessa.

9 Integrazione numerica

9.1 Teoria generale sull'integrazione numerica

Supponiamo di scrivere una funzione f come combinazione lineare di funzioni più un certo errore:

$$f(x) = \sum_{i=0}^n c_i \phi_i(x) + E_n(x)$$

Il relativo integrale su un intervallo $[a, b]$ allora sarebbe

$$\int_a^b f(x) dx = \sum_{i=0}^n c_i \int_a^b \phi_i(x) dx + \int_a^b E_n(x) dx$$

Supponiamo ora di star approssimando la funzione con il polinomio di Lagrange, otteniamo allora

$$\int_a^b f(x) dx = \sum_{i=0}^n f(x_i) \int_a^b L_i(x) dx + \int_a^b E_n(x) dx$$

Quindi, il valore approssimato dell'integrale é

$$I_n(f; a, b) = \sum_{i=0}^n f(x_i) \int_a^b L_i(x) dx$$

che prende nome di *formula di quadratura* associata ad f sull'intervallo $[a, b]$. I coefficienti $\int_a^b L_i(x)$ si chiamano *pesi*.

Osservazione 9.1. Di solito il calcolo dei pesi viene fatto su un intervallo di riferimento $[c, d]$ che non deve per forza coincidere con quello di partenza. Per poi ottenere i pesi relativi all'intervallo $[a, b]$ basta osservare che

$$\int_a^b L_i(x) dx = \frac{b-a}{d-c} \int_c^d L_i(\gamma(t)) dt$$

dove abbiamo effettuato il cambio di variabile $\gamma(t) = a + \frac{b-a}{d-c}(t-c)$

Osservazione 9.2 (Errore). Possiamo osservare che c'è un'ovvia maggiorazione dell'errore

$$\int_a^b E_n(x) dx \leq (b-a) \|E_n\|_{L^\infty([a,b])}$$

ma per averne delle migliori dovremmo aspettare (cfr. sotto).

Osservazione 9.3 (Grado di precisione). Fissiamo x_0, \dots, x_{n+1} dei nodi; l'accuratezza di una formula di quadratura si caratterizza spesso tramite il parametro seguente

$$\nu = \max \left\{ k : \int_a^b f(x) dx = \sum_{i=0}^n f(x_i) \int_a^b L_i(x) dx \quad \forall f \in \mathbb{P}_k \right\}$$

Supponiamo ora di costruire un polinomio interpolatore su $n + 1$ nodi, in questo caso allora per ogni $f \in \mathbb{P}_n$ vale $\Pi_n^f = f$ ¹¹ e quindi $E_n = 0$. Tutto ciò implica $\int_a^b f(x) dx = 0$. Quindi in generale vale $\nu \geq n$.

Vediamo ora un altro risultato che sarà utile piú tardi.

Proposizione 9.4. Fissiamo $f : [a, b] \rightarrow \mathbb{R}$ una funzione; allora la somma dei pesi uguaglia sempre l'ampiezza dell'intervallo, in formule

$$b - a = \sum_{i=0}^n \int_a^b L_i(t) dt$$

Dimostrazione. Supponiamo di voler interpolare la funzione $f(t) = 1$ in $[a, b]$ con il polinomio di Lagrange. Fissiamo quindi $n + 1$ punti $\{x_i\}_{i \in \{0, \dots, n\}}$ in $[a, b]$ e, costruendo il polinomio interpolatore, otteniamo:

$$\Pi_n(t) = \sum_{i=0}^n L_i(t) f(x_i) = \sum_{i=0}^n L_i(t)$$

Usando il fatto che il polinomio interpolatore é unico e che f é già un polinomio, abbiamo che $f = \Pi_n$. Quindi per ogni t vale

$$\sum_{i=0}^n L_i(t) = 1$$

Integriamo ora nell'intervallo $[a, b]$ e otteniamo

$$b - a = \int_a^b 1 dt = \int_a^b \sum_{i=0}^n L_i(t) dt = \sum_{i=0}^n \int_a^b L_i(t) dt$$

Abbiamo quindi la tesi. □

Osservazione 9.5 (Propagazione delle perturbazioni). La quantità $\sum_{i=0}^n \left| \int_a^b L_i(x) dx \right|$ é legata anche alla propagazione delle perturbazioni. Supponiamo che i valori $f(x_i)$ siano affetti da una perturbazione δ_i con $|\delta_i| \leq \delta$, si ha allora

$$\sum_{i=0}^n \int_a^b L_i(x) dx (f(x_i) + \delta_i) = I_n(f; a, b) + \sum_{i=0}^n \int_a^b L_i(x) dx \delta_i$$

quindi, otteniamo la stima

$$\left| \sum_{i=0}^n \int_a^b L_i(x) dx (f(x_i) + \delta_i) \right| \leq |I_n(f; a, b)| + \delta \sum_{i=0}^n \left| \int_a^b L_i(x) dx \right|$$

¹¹ Π_n^f é il polinomio interpolatore di f associato agli $n + 1$ nodi fissati.

Teorema 9.6 (Teorema di Polya). Sia $I_n(f, a, b)$, con $x_0, \dots, x_n \in [a, b]$. Condizione necessaria e sufficiente perché si abbia, per ogni $f \in C^0([a, b])$ $I_n(f, a, b) \rightarrow \int_a^b f(x)dx$ per $n \rightarrow \infty$, è che, al variare di n , esista una costante $M > 0$ tale che

$$\sum_{i=0}^n \left| \int_a^b L_i(x) dx \right| \leq M$$

e che, per ogni polinomio di grado fissato $p(x) \in \mathbb{P}_k$, si abbia

$$\lim_{n \rightarrow \infty} I_n(p, a, b) = \int_a^b p(x)dx.$$

Dimostrazione. Ci limitiamo a dimostrare la parte relativa alla sufficienza. Per la densità dei polinomi in C^0 (teorema di Stone-Weierstrass), dato un $\varepsilon > 0$, esiste un polinomio $p_\varepsilon \in \mathbb{P}_k$ tale che $|f(x) - p_\varepsilon(x)| < \varepsilon$ per ogni $x \in [a, b]$. D'altra parte,

$$\begin{aligned} \left| \int_a^b f(x) dx - I_n(f, a, b) \right| &\leq \left| \int_a^b f(x) dx - \int_a^b p_\varepsilon(x) dx \right| + \\ &\quad + \left| \int_a^b p_\varepsilon(x) dx - I_n(p_\varepsilon, a, b) \right| + |I_n(p_\varepsilon, a, b) - I_n(f, a, b)|. \end{aligned}$$

Per il primo dei tre termini a secondo membro si ha:

$$\left| \int_a^b f(x) dx - \int_a^b p_\varepsilon(x) dx \right| \leq \int_a^b |f(x) - p_\varepsilon(x)| dx \leq \varepsilon(b-a)$$

mentre per il secondo,

$$\left| \int_a^b p_\varepsilon(x) dx - I_n(p_\varepsilon, a, b) \right| \rightarrow 0.$$

Il terzo termine, si può maggiorare come:

$$|I_n(p_\varepsilon, a, b) - I_n(f, a, b)| \leq \sum_{i=0}^n |(p_\varepsilon(x_i) - f(x_i))| \left| \int_a^b L_i(x) dx \right| \leq M\varepsilon.$$

Se ne deduce quindi che tutti e tre i termini possono essere resi arbitrariamente piccoli. \square

9.2 Formule di quadratura di Newton-Cotes

Queste formule di quadratura prevedono l'utilizzo dell'interpolazione usando nodi equidistanti tra loro. Molte volte si usa anche l'interpolazione composita, quindi dividendo in più intervalli il dominio di partenza e considerando più polinomi interpolatori e nodi equidistanti all'interno di ogni intervallo. Diremo che la formula è *semplice* se viene costruita su un unico polinomio interpolatore, altrimenti diremo che la formula è *composita*.

Definizione 9.7. Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione. Dividiamo l'intervallo $[a, b]$ in m intervalli $[a_i, a_{i+1}]$ con $i \in \{0, \dots, m-1\}$ e scegliamo $n+1$ nodi $\{x_0^{(i)}, \dots, x_n^{(i)}\}$ equidistanti all'interno di ogni intervallo. La *formula di quadratura di Newton - Cotes* è:

$$I_{n,m} = \sum_{j=0}^{m-1} \sum_{k=0}^n \int_{a_j}^{a_{j+1}} L_k^{(j)}(t) dt f(x_k^{(j)})$$

dove $L_k^{(j)}(t)$ è il polinomio presente nella definizione del polinomio interpolatore di Lagrange relativo al nodo k dell'intervallo $[a_j, a_{j+1}]$. Indicheremo con $H_j := a_{j+1} - a_j$ e $H = \max_j H_j$. Inoltre, chiameremo *passo* la distanza tra i nodi dell'interpolazione e la indicheremo con h .

Teorema 9.8. Costruendo la base di Lagrange $\{L_i\}$ su nodi equidistanti si ottiene una simmetria delle funzioni L_i rispetto la posizione dei nodi.

Dimostrazione. Dimostriamo il caso in cui abbiamo un numero pari di nodi (il caso dispari é analogo). Inoltre, fissiamo l'intervallo $[0, 1]$ ¹² e per comodità chiamiamo i punti $\{x_{-n}, \dots, x_{-1}, x_1, \dots, x_n\}$. Visto che i punti sono equidistanti vale la relazione $x_i = -1 + \frac{2i}{n}$. Fissiamo ora $i > 0$ (per $i < 0$ é analogo), sempre per equidistanza vale che $x_i = -x_{-i}$. A questo punto osserviamo che:

$$L_i(x) = \prod_{k < 0} \frac{x - x_k}{x_i - x_k} \prod_{k > 0, k \neq i} \frac{x - x_k}{x_i - x_k}$$

e che

$$L_{-i}(x) = \prod_{k < 0, k \neq -i} \frac{x - x_k}{x_{-i} - x_k} \prod_{k > 0} \frac{x - x_k}{x_{-i} - x_k}$$

Ora mostriamo la simmetria (...)

□

Quindi, dalle proprietà dell'integrazione, avremo che l'integrale di L_i su $[a, b]$ sarà uguale all'integrale di L_{-i} sempre su $[a, b]$.

Teorema 9.9. Dati n punti equidistanti su $[a, b]$, allora i pesi di L_i sui punti x_i su $[a, b]$ sono uguali ai pesi $l_i(x)$ definiti sui nodi i tali che $x_i = a + \frac{b-a}{n}i$ moltiplicati per un fattore $\frac{b-a}{n}$; in formule:

$$\int_a^b L_i(t) dt = \frac{b-a}{n} \int_0^n l_i(t) dt$$

¹²Se l'intervallo é $[a, b]$, applicando una trasformazione affine, la geometria dei nodi (e quindi la loro equidistanza) si mantiene. Quindi ridurci a questo caso non é limitativo.

Dimostrazione. Usando il cambio di variabile $t := a + \frac{b-a}{n}s$ (e ricordando che i nodi sono stati costruiti tali che $x_i = a + \frac{b-a}{n}i$) abbiamo che:

$$\begin{aligned}
\int_a^b L_k(t) dt &= \int_a^b \prod_{i \neq k} \frac{t - x_i}{x_k - x_i} dt \\
&= \int_0^n \prod_{i \neq k} \frac{(a + \frac{b-a}{n}s) - (a + \frac{b-a}{n}i)}{(a + \frac{b-a}{n}k) - (a + \frac{b-a}{n}i)} \cdot \frac{b-a}{n} ds \\
&= \int_0^n \prod_{i \neq k} \frac{\frac{b-a}{n}(s-i)}{\frac{b-a}{n}(k-i)} \cdot \frac{b-a}{n} ds \\
&= \frac{b-a}{n} \int_0^n \prod_{i \neq k} \frac{s-i}{k-i} ds \\
&= \frac{b-a}{n} \int_0^n l_k(s) ds
\end{aligned}$$

□

Questo teorema ci permette di calcolare i pesi solo su un intervallo fissato $[-1, 1]$ e ottenere i pesi su un qualsiasi intervallo semplicemente moltiplicando un fattore. Vedremo nelle prossime sezioni, come questi due teoremi ci aiutino a calcolare delle formule di integrazione note in modo facile.

9.2.1 Formule chiuse

Definizione 9.10. Una formula si dice *chiusa* se i due nodi estremi coincidono con gli estremi dell'intervallo.

Quindi il passo di questa formula sarà $h = \frac{b-a}{n}$ e i nodi (visto che sono equidistanti) saranno della forma

$$x_i = a + ih$$

Osservazione 9.11 (Regola del trapezio). Mettendo $n = 1$ e $m = 1$ abbiamo che $I_{1,1} = \int_a^b \frac{t-a}{b-a} dt f(b) + \int_a^b \frac{t-b}{a-b} dt f(a) = \frac{b-a}{2}(f(a) + f(b))$ ¹³, otteniamo in questo modo la *regola del trapezio*. Più in generale, si può fare lo stesso conto in forma composta, quindi modificando m e facendo lo stesso conto in ogni sottointervallo (con n punti equidistanti all'interno di ogni sottointervallo) otteniamo quella che è la *regola del trapezio generalizzata*.

Osservazione 9.12 (Regola di Simpson). Imponiamo ora $n = 2$ e lasciamo sempre $m = 1$, allora vale:

$$I_{2,1}(f; a, b) = \frac{b-a}{2} \cdot \frac{1}{3}(f(x_0) + 4f(x_1) + f(x_2))$$

¹³Senza fare il conto degli integrali, si poteva usare il fatto che la somma dei pesi uguaglia l'ampiezza dell'intervallo

Per mostrare la formula, visto quanto detto sopra sulla simmetria delle L e sui pesi, basta fare il conto del peso di L_0 nell'intervallo $[-1, 1]$. Siano $x_0 = -1$, $x_1 = 0$ e $x_2 = 1$; abbiamo $L_0(x) = \frac{x}{-1} \frac{x-1}{-2} = \frac{x^2}{2} - \frac{x}{2}$ e quindi

$$\int_{-1}^1 L_0(x) dx = \int_{-1}^1 \left(\frac{x^2}{2} - \frac{x}{2} \right) dx = \left[\frac{x^3}{6} - \frac{x^2}{4} \right]_{-1}^1 = \left(\frac{1}{6} - \frac{1}{4} \right) - \left(-\frac{1}{6} - \frac{1}{4} \right) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Indichiamo ora con α_1 e α_2 i pesi relativi a x_1 e x_2 ; usando la simmetria delle L_I otteniamo che $\alpha_2 = \frac{1}{3}$, mentre usando che la somma dei pesi é la lunghezza dell'intervallo otteniamo $\alpha_1 = 2 - \frac{2}{3} = \frac{4}{3}$. Ricordando che, nel caso di nodi equidistanti, vale

$$\int_a^b L_i(x) dx = \frac{b-a}{2} \int_{-1}^1 L_i(x) dx$$

otteniamo la formula. Supponiamo ora di voler generalizzare la regola di Simpson, quindi mettere m arbitrario e 3 nodi in ogni intervallo. Applicando la formula di simpson a ciascun intervallo otterremo la formula

$$I_{2,m}(f; a, b) = \frac{b-a}{2} \frac{1}{3} (f(x_0^1) + 4f(x_1^1) + f(x_2^1)) + \cdots + \frac{b-a}{2} \frac{1}{3} (f(x_0^m) + 4f(x_1^m) + f(x_2^m))$$

dove x_j^i é il j -esimo nodo nell' i -esimo intervallo. Osservando che i punti agli estremi degli intervalli sono nodi per entrambi degli intervalli, vale che $x_2^i = x_0^{i+1}$ per ogni $i \in \{0, \dots, m-1\}$. Quindi la formula diventa

$$I_{2,m}(f; a, b) = \frac{b-a}{2} \frac{1}{3} ((f(x_0^1) + 4f(x_1^1) + 2f(x_2^1) + \cdots + 2f(x_0^m) + 4f(x_1^m) + f(x_2^m)))$$

9.2.2 Formule aperte

Definizione 9.13. Una formula si dice *aperta* se nessun nodo coincide con uno degli estremi dell'intervallo.

Quindi il passo di questa formula sará $h = \frac{b-a}{n+2}$ e i nodi (visto che sono equidistanti) saranno della forma

$$x_i = a + (i+1)h$$

Osservazione 9.14 (Regola del punto medio). Mettendo $n = 0$ e chiamando $x_0 = a + h$ otteniamo $I_{0,1} = \int_a^b 1 dt f(x_0)$, che si chiama *formula del punto medio*.

Osservazione 9.15. Mettendo invece $n = 1$ e chiamando $x_0 = a + h$ e $x_1 = a + 2h$ i due punti in $[a, b]$ allora abbiamo che

$$I_{1,1} = \int_a^{a+3h} \frac{t-x_1}{x_0-x_1} f(x_0) dt + \int_a^{a+3h} \frac{t-x_0}{x_1-x_0} f(x_1) dt$$

calcolando il primo integrale otteniamo

$$\begin{aligned}
\int_a^{a+3h} \frac{t - (a + 2h)}{-h} dt &= -\frac{1}{h} \int_a^{a+3h} [t - (a + 2h)] dt \\
&= -\frac{1}{h} \left[\frac{t^2}{2} - (a + 2h)t \right]_a^{a+3h} \\
&= -\frac{1}{h} \left[\frac{(a + 3h)^2}{2} - (a + 2h)(a + 3h) - \frac{a^2}{2} + (a + 2h)a \right] \\
&= -\frac{1}{h} \left[\frac{a^2 + 6ah + 9h^2}{2} - (a^2 + 3ah + 2ah + 6h^2) - \frac{a^2}{2} + a^2 + 2ah \right] \\
&= -\frac{1}{h} \left[\frac{a^2 + 6ah + 9h^2 - a^2}{2} - a^2 - 5ah - 6h^2 + a^2 + 2ah \right] \\
&= -\frac{1}{h} \left[\frac{6ah + 9h^2}{2} - 5ah - 6h^2 + 2ah \right] \\
&= -\frac{1}{h} \left[3ah + \frac{9h^2}{2} - 3ah - 6h^2 \right] \\
&= -\frac{1}{h} \left[\frac{9h^2}{2} - 6h^2 \right] \\
&= -\frac{1}{h} \cdot \left(-\frac{3h^2}{2} \right) \\
&= \frac{3h}{2}
\end{aligned}$$

con un conto analogo otteniamo lo stesso risultato per il secondo integrale ottenendo la formula

$$I_{1,1} = \int_a^{a+3h} \frac{t - x_1}{x_0 - x_1} f(x_0) dt + \int_a^{a+3h} \frac{t - x_0}{x_1 - x_0} f(x_1) dt = \frac{3h}{2} (f(x_0) + f(x_1))$$

Osservazione 9.16. Le seguenti sono alcune formule di quadratura aperte con $m = 1$:

$$I_2(f, a, b) = \frac{2(b - a)}{3} [2f(x_0) - f(x_1) + 2f(x_2)],$$

$$I_3(f, a, b) = \frac{5(b - a)}{72} [11f(x_0) + f(x_1) + f(x_2) + 11f(x_3)],$$

$$I_4(f, a, b) = \frac{3(b - a)}{40} [11f(x_0) - 14f(x_1) + 26f(x_2) - 14f(x_3) + 11f(x_4)],$$

$$I_5(f, a, b) = \frac{7(b - a)}{7200} [611f(x_0) - 453f(x_1) + 562f(x_2) + 562f(x_3) - 453f(x_4) + 611f(x_5)].$$

9.2.3 Grado di precisione, errore e convergenza

Teorema 9.17 (Grado di precisione delle formule di Newton-Cotes). Il grado di precisione è n se n è dispari e $n + 1$ se n è pari.

Dimostrazione. Ci limitiamo a dimostrare che il grado di precisione ν soddisfa $\nu \geq n$ nel primo caso, e $\nu \geq n + 1$ nel secondo. Come già osservato sopra, visto che il polinomio interpolatore ha grado n , vale che $\nu \geq n$. Abbiamo quindi mostrato la prima disuguaglianza, ora supponiamo f un polinomio di grado $n + 1$ e n pari. Osservando che $f^{(n+1)}(x) = c$ per qualche $c \in \mathbb{R}$ e usando la formula di rappresentazione dell'errore, abbiamo che

$$\int_a^b E_n(x) dx = \int_a^b \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x) dx = \frac{c}{(n+1)!} \int_a^b \omega_n(x) dx$$

Osservando che il numero di nodi è dispari e che stiamo considerando tutti nodi equidistanti, abbiamo che la funzione $\omega_n(x)$ è dispari rispetto al nodo centrale di $[a, b]$. Di conseguenza il suo integrale è nullo. Otteniamo così anche la seconda disuguaglianza. \square

Teorema 9.18 (Espressione esplicita dell'errore di quadratura). Sia $p = n + 1$ se n è dispari, $p = n + 2$ se n è pari. Allora, se $f \in C^p([a, b])$, l'errore di quadratura nelle formule di Newton–Cotes semplici si può scrivere come

$$\int_a^b f(x) dx - I_n(f, a, b) = C_n h^{p+1} f^{(p)}(\xi)$$

dove $\xi \in (a, b)$, e la costante C_n vale:

$$C_n = \begin{cases} \frac{1}{(n+1)!} \int_0^n s(s-1) \cdots (s-n) ds < 0 & \text{chiusa, } n \text{ dispari} \\ \frac{1}{(n+1)!} \int_{-1}^{n+1} s(s-1) \cdots (s-n) ds > 0 & \text{aperte, } n \text{ dispari} \\ \frac{1}{(n+2)!} \int_0^n s^2(s-1) \cdots (s-n) ds < 0 & \text{chiusa, } n \text{ pari} \\ \frac{1}{(n+2)!} \int_{-1}^{n+1} s^2(s-1) \cdots (s-n) ds > 0 & \text{aperte, } n \text{ pari.} \end{cases}$$

Teorema 9.19 (Convergenza delle formule di Newton-Cotes composite). Se f è Riemann-integrabile su $[a, b]$ e se la formula di quadratura $I_{n,m}$ di Newton-Cotes è esatta sulle costanti¹⁴, allora

$$\lim_{H \rightarrow 0} I_{n,m}(f; a, b) = \int_a^b f(x) dx$$

Dimostrazione. Se $I_{n,m}$ è esatta sulle costanti, allora abbiamo che

$$b - a = \int_a^b 1 dx = \sum_{j=0}^{m-1} \sum_{k=0}^n \int_{a_j}^{a_{j+1}} L_k^{(j)}(t) dt$$

Se la quadratura è chiusa poniamo $l := n$, altrimenti $l = n + 2$. Definiamo ora sull'intervallo di lunghezza l la formula di quadratura ottenuta usando nodi a distanza 1 l'uno dall'altro e

¹⁴In realtà, come visto all'inizio del capitolo, la somma dei pesi è sempre uguale alla lunghezza dell'intervallo. Quindi questa ipotesi, nel caso di formule definite sulla base di Lagrange, è sempre vera.

chiamiamo ω_k i pesi di questa quadratura relativo al nodo k . Vale quindi che $\int_{a_j}^{a_{j+1}} L_k^{(j)}(t) dt = h_j \omega_k$ con $h_j = H_j/l$.¹⁵ Dall'ipotesi di esattezza di $I_{n,m}$ vale che $l = \sum_k \omega_k$. Vale quindi

$$\begin{aligned} I_{n,m}(f, a, b) &= \sum_{j=0}^{m-1} \frac{H_j}{l} \sum_{k=0}^n w_k f(x_k^{(j)}) = \\ &= \frac{1}{l} \sum_{k=0}^n w_k \sum_{j=0}^{m-1} H_j f(x_k^{(j)}) \end{aligned}$$

Ora, visto che f è integrabile, e che nella somma più interna è presente una somma integrale, abbiamo che

$$I_{n,m}(f, a, b) \rightarrow \frac{1}{l} \sum_{k=0}^n w_k \int_a^b f(x) dx = \int_a^b f(x) dx$$

□

Teorema 9.20 (Ordine di convergenza delle formule di Newton-Cotes composite). Data la formula di quadratura composita, se valgono le ipotesi del Teorema sull'espressione esplicita dell'errore di quadratura, allora, per qualche costante $C > 0$ si ha

$$\left| \int_a^b f(x) dx - I_{n,m}(f, a, b) \right| \leq C \sup_j \left(H_j^p \sup_{[a_j, b_j]} |f^{(p)}(x)| \right)$$

ed in particolare:

$$\left| \int_a^b f(x) dx - I_{n,m}(f, a, b) \right| \leq C \sup_{[a,b]} |f^{(p)}(x)| H^p.$$

Dimostrazione. Una volta mostrata la prima disuguaglianza, la seconda è immediata.

Usando la stessa costruzione della dimostrazione precedente, usiamo l . Sommando gli errori di quadratura ottenuti dagli intervalli elementari si ha

$$\begin{aligned} \left| \int_a^b f(x) dx - I_{n,m}(f, a, b) \right| &\leq \sum_{j=0}^{m-1} \left| \int_{a_j}^{b_j} f(x) dx - I_n(f, a_j, b_j) \right| = \\ &= \sum_{j=0}^{m-1} |C_n| \left(\frac{H_j}{l} \right)^{p+1} |f^{(p)}(\xi_j)| = \\ &= \frac{|C_n|}{l^{p+1}} \sum_{j=0}^{m-1} H_j^{p+1} |f^{(p)}(\xi_j)| \end{aligned}$$

¹⁵Abbiamo visto il teorema all'inizio del paragrafo che mostra questo fatto. Lo rifacciamo qui con gli indici su ogni sottointervallo anche se non necessario solamente a scopo didattico. Usando il cambio di variabile $t := a_j + \frac{H_j}{l}s$ (e ricordando che i nodi sono stati costruiti tali che $x_k^{(j)} = a_j + \frac{H_j}{l}k$), abbiamo che $\int_{a_j}^{a_{j+1}} L_k^{(j)}(t) dt = \int_{a_j}^{a_{j+1}} \prod_{i \neq k} \frac{t - x_i^{(j)}}{x_k^{(j)} - x_i^{(j)}} dt = \int_0^l \prod_{i \neq k} \frac{(a_j + \frac{H_j}{l}s) - (a_j + \frac{H_j}{l}i)}{(a_j + \frac{H_j}{l}k) - (a_j + \frac{H_j}{l}i)} \cdot \frac{H_j}{l} ds = \int_0^l \prod_{i \neq k} \frac{\frac{H_j}{l}(s-i)}{\frac{H_j}{l}(k-i)} \cdot \frac{H_j}{l} ds = \frac{H_j}{l} \int_0^l \prod_{i \neq k} \frac{s-i}{k-i} ds = \frac{H_j}{l} \int_0^l L_k(s) ds = \frac{H_j}{l} \omega_k = h_j \omega_k$.

dove si è applicata la formula dell'errore su ogni intervallo della quadratura composita. Ora, scrivendo $H_j^{p+1} = H_j^p H_j$ e maggiorando il prodotto $H_j^p |f^{(p)}(\xi_j)|$, si ottiene finalmente

$$\left| \int_a^b f(x) dx - I_{n,m}(f, a, b) \right| \leq \frac{|C_n|}{l^{p+1}} \sup_j \left(H_j^p \sup_{[a_j, b_j]} |f^{(p)}(x)| \right) \sum_{j=0}^{m-1} H_j$$

□

9.3 Quadrature Gaussianhe

Queste formule di quadrature si basano su nodi che non sono equidistanti, e sono messe in corrispondenza a degli zeri di determinati polinomi ortogonali. Consideriamo quindi lo spazio di Banach $L^2([a, b])$ delle funzioni di quadrato integrabili su $[a, b]$ con il prodotto scalare tra due funzioni f, g come

$$f \cdot g = \int_a^b f(t)g(t) dt$$

Definizione 9.21. La famiglia dei *polinomi di Legendre in* $[-1, 1]$ é la famiglia di polinomi $\{P_k\}_{k \geq 0}$, costruita tramite il processo di ortogonalizzazione della base $\{1, x, x^2, \dots\}$ dello spazio dei polinomi di grado arbitrariamente alto \mathbb{P}_∞ .

Definizione 9.22. Data una funzione $f : [-1, 1] \rightarrow \mathbb{R}$, e un $n \in \mathbb{N}_{\geq 0}$ la *formula di quadratura di Gauss-Legendre per l'intervallo* $[-1, 1]$ é la formula di quadratura data da

$$\sum_{i=0}^n \int_{-1}^1 L_i(x) f(x_i)$$

dove x_i sono le radici del polinomio di *Legendre* P_{n+1} .

Di seguito mostriamo una tabella con all'interno il numero di nodi $n + 1$, i t_i (le radici del polinomio P_i) e w_i (pesi della formula di quadratura ottenuta costruendo i polinomi della

base di Lagrange L_i sui nodi t_i), tutti calcolati sull'intervallo $(-1,1)$:

n	t_i	w_i
0	0.0	2.0
1	± 0.57735027	1.0
2	± 0.77459667	0.55555556
	0.0	0.88888889
3	± 0.86113631	0.34785485
	± 0.33998104	0.65214515
4	± 0.90617985	0.23692689
	± 0.53846931	0.47862867
	0.0	0.56888889
5	± 0.93246951	0.17132449
	± 0.66120939	0.36076157
	± 0.23861919	0.46791393
6	± 0.94910791	0.12948497
	± 0.74153119	0.27970539
	± 0.40584515	0.38183005
	0.0	0.41795918

Supponiamo ora di voler applicare la stessa formula di quadratura su un intervallo $[a, b]$ generico. In che modo si scelgono i nodi? In questo caso basta costruire una trasformazione lineare $\delta : [-1, 1] \rightarrow [a, b]$ tale che $\delta(-1) = a$ e $\delta(1) = b$, ovvero:

$$\delta(x) = \frac{a+b}{2} + \frac{b-a}{2}x$$

Definizione 9.23. Definiamo la *formula di quadratura di Gauss-Legendre per l'intervallo* $[a, b]$ la formula di quadratura

$$\sum_{i=0}^n \int_a^b l_i(t) dt f(t_i)$$

dove i nodi t_i sono l'immagine tramite δ degli zeri x_i dei polinomi di Legendre (ovvero $t_i := \delta(x_i)$) e i polinomi l_i sono i polinomi nella base di Lagrange associati ai nodi t_i . Inoltre se P_i sono i polinomi di Legendre relativi all'intervallo $[-1, 1]$, allora con $P_i(\delta(x))$ indicheremo i *polinomi di Legendre relativi all'intervallo* $[a, b]$.

Anche in questo caso, per ottenere l'integrale di un L_i su un intervallo generico basta fare il cambio di variabile. Infatti, con il cambio di variabile $t = \delta(x)$ e ricordando che $t_i := \delta(x_i)$, abbiamo:

$$\sum_{i=0}^n \int_a^b l_i(t) dt f(t_i) = \sum_{i=0}^n \frac{b-a}{2} \int_{-1}^1 l_i(\delta(x)) dx f(\delta(x_i)) = \sum_{i=0}^n \frac{b-a}{2} \int_{-1}^1 L_i(x) dx f(\delta(x_i))$$

A questo punto, usando la tabella dei valori degli integrali su $[-1, 1]$ e dopo aver trovato la trasformazione corretta, per calcolare l'integrale basta sostituire i valori correttamente. Vediamo ora le proprietà dei polinomi di Legendre.

Teorema 9.24 (Zeri dei polinomi di Legendre). Consideriamo l'insieme dei polinomi di Legendre nello spazio $L_2([a, b])$. Allora per ogni $k \geq 1$ P_k ha esattamente k radici dentro $[a, b]$.

Dimostrazione. Supponiamo per assurdo che, nell'intervallo (a, b) , $P_n(x)$ abbia $r < n$ radici reali di molteplicità dispari x_0, \dots, x_{r-1} . Definiamo il polinomio

$$Q(x) = \begin{cases} 1 & \text{se } r = 0 \\ (x - x_0)(x - x_1) \cdots (x - x_{r-1}) & \text{altrimenti.} \end{cases}$$

Supponendo $r \neq 0$ sia Q che $P_n(x)$ cambiano segno in corrispondenza dei punti x_0, \dots, x_{r-1} , quindi il prodotto $P_n(x)Q(x)$ ha segno costante e non è identicamente nullo, invece se $r = 0$ $P_n(x)$ è prodotto di monomi elevati ad una potenza pari quindi anche lei è una funzione positiva. Quindi vale che

$$\int_a^b P_n(x)Q(x) dx \neq 0$$

ma ciò è assurdo in quanto il grado di Q è $r < n$, e P_n è ortogonale (rispetto al prodotto scalare) a tutti i primi n polinomi P_0, \dots, P_{n-1} , i quali d'altra parte generano tutto lo spazio dei polinomi di grado non superiore ad $n - 1$. Ne segue che $r = n$ e che le radici sono tutte semplici. □

Teorema 9.25 (Grado di precisione per le formule Gaussiane). Supponendo di usare la formula di quadratura di Gauss sugli $n + 1$ nodi menzionati sopra, si ha grado di precisione $\nu = 2n + 1$.

Dimostrazione. Data una $f(x)$ nello spazio dei polinomi di grado non superiore a $2n + 1$, in base a note proprietà dei polinomi essa può essere scritta come

$$f(x) = p_n(x)\omega(x) + q_n(x)$$

dove come al solito $\omega(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$ e p_n, q_n sono polinomi di grado non superiore ad n . Poiché la formula ha $n + 1$ nodi, il suo grado di precisione è almeno n . Si ha inoltre:

$$\begin{aligned} \int_a^b f(x) dx &= \int_a^b p_n(x)\omega(x) dx + \int_a^b q_n(x) dx = \\ &= \int_a^b p_n(x)\omega(x) dx + I_n(f, a, b) \end{aligned}$$

dove l'ultimo passaggio è giustificato dal fatto che q_n viene integrato esattamente e che $I_n(f, a, b) = I_n(q_n, a, b)$ (infatti $f(x_i) = q_n(x_i)$ per l'annullarsi di ω nei nodi e i pesi sono costruiti sugli stessi nodi x_i). Ora, $\omega(x)$ coincide, a meno di una costante, con l' $(n + 1)$ -esimo polinomio della famiglia di Legendre ed è quindi ortogonale ai polinomi P_0, \dots, P_n . Di

conseguenza (come si è già notato) sarà ortogonale a tutto lo spazio dei polinomi di grado non superiore ad n . Ne consegue che è nullo il prodotto scalare (p_n, ω) , ovvero

$$\int_a^b p_n(x) \omega(x) dx = 0$$

ed in conclusione

$$\int_a^b f(x) dx = I_n(f, a, b).$$

Resta da verificare che il grado di precisione non può essere maggiore di $2n + 1$. Infatti, se per assurdo i polinomi di grado $2n + 2$ fossero integrati esattamente, lo sarebbe anche il polinomio $\omega(x)^2$, il quale però è nullo in tutti i nodi (quindi $I_n(f; a, b) = 0$) ed è una funzione positiva (quindi ha integrale positivo). Otteniamo quindi un assurdo. \square

Teorema 9.26 (Convergenza delle formule Gaussiane). Per ogni $f \in C^0([a, b])$, la formula di quadratura di Gauss-Legendre definita sopra converge al valore esatto dell'integrale; in formule:

$$\lim_{n \rightarrow \infty} I_n(f; a, b) = \int_a^b f(x) dx$$

Dimostrazione. Per farlo applichiamo il teorema di Polya¹⁶. Sia k un intero positivo e $p \in \mathbb{P}_k$. Visto che, dato un insieme di punti $\{x_0, \dots, x_n\}$ il grado di precisione è almeno n , per ogni $m \in \mathbb{N}$ tale che $m \geq k$ vale che $I_m(f; a, b) = \int_a^b f(x) dx$. E quindi la seconda condizione sufficiente per il teorema di Polya è soddisfatta. Ora mostriamo la prima condizione. Fissiamo ora $n \in \mathbb{N}$, e consideriamo per ogni i l'integrale $\int_a^b L_i^2(x) dx$. Abbiamo che $\deg(L_i) \leq n$ e quindi che $\deg(L_i) \leq 2n$, mentre dal teorema sopra sappiamo che il grado di precisione per questa formula di quadratura è $2n + 1$. Quindi la formula è esatta per i polinomi L_i^2 . Ricordando che l'integrale di una funzione positiva è ancora positivo, abbiamo:

$$0 < \int_a^b L_i^2(x) dx = I_n(L_i^2; a, b) = \sum_{j=0}^n \int_a^b L_j(x) dx L_i(x_j) = \sum_{j=0}^n \int_a^b L_j(x) dx \delta_{ij} = \int_a^b L_i(x) dx$$

di conseguenza abbiamo che

$$\sum_{i=0}^n \left| \int_a^b L_i(x) dx \right| = \sum_{i=0}^n \int_a^b L_i(x) dx = b - a$$

e quindi anche la seconda condizione è soddisfatta. \square

¹⁶Ricordiamo l'enunciato del teorema: Sia $I_n(f, a, b)$, con $x_0, \dots, x_n \in [a, b]$. Condizione necessaria e sufficiente perché si abbia, per ogni $f \in C^0([a, b])$ $I_n(f, a, b) \rightarrow \int_a^b f(x) dx$ per $n \rightarrow \infty$, è che, al variare di n , esista una costante $M > 0$ tale che $\sum_{i=0}^n \left| \int_a^b L_i(x) dx \right| \leq M$ e che, per ogni polinomio di grado fissato $p(x) \in \mathbb{P}_k$, si abbia $\lim_{n \rightarrow \infty} I_n(p, a, b) = \int_a^b p(x) dx$.