

UNIVERSITÁ DEGLI STUDI DI PADOVA

Facoltá di Scienze MM. FF. NN.

Laurea Triennale in Matematica

Metodi di
Interpolazione Razionale
e Applicazioni

Laureanda: Irene Federici

Relatore: Dott. Fabio Marcuzzi

Anno Accademico 2003 - 2004

a Lisa, Augusto, mamma e papà

Prefazione

Perché l'interpolazione?

L'interpolazione di funzioni o di dati sperimentali, é un argomento molto vasto (che mi é sempre piaciuto molto). Certo richiederebbe una trattazione di alcune migliaia di pagine. Non potendomi, ovviamente, soffermare cosí a lungo, ho preferito introdurre il discorso in modo generale elencando solo alcune casistiche interessanti. Quindi ho approfondito l'aspetto dell'interpolazione razionale per poi introdurre Spline e B-spline in modo da arrivare alle NURBS.

Queste ultime sono molto interessanti dal punto di vista applicativo, ad esempio nella 'computer graphics'. Quindi ho provato a sperimentarne l'utilizzo in un programma (Maya) che mi ha dato l'occasione di vedere, a livello pratico, come si concretizzano formule matematiche abbastanza complesse.

Indice

Prefazione	3
Perché l'interpolazione?	3
Introduzione: il Problema dell'Interpolazione	7
1 Interpolazione Polinomiale	9
1.1 Il Polinomio di Interpolazione	9
1.1.1 La Matrice di Vandermonde	10
1.2 Interpolazione in Forma di Lagrange	11
1.3 L'Errore di Interpolazione	15
1.4 Formula di Newton	17
1.4.1 L'Errore Usando le Differenze Divise	23
1.4.2 Il Fenomeno di Runge	23
2 Interpolazione Razionale	27
2.1 Proprietá Generali dell'Interpolazione Razionale	27
2.2 Frazioni Continue	32
2.2.1 Differenze Inverse	37
2.2.2 Differenze Reciproche	41
3 Funzioni Spline e B-Spline	47
3.1 Polinomi di Interpolazione a Tratti di Primo Grado	47
3.2 L'Errore nell'Interpolazione a Tratti	49
3.3 Splines	50
3.4 B-Splines	51
3.4.1 Formula di Ricorrenza per B-splines	54

4	Curve e Superfici B-splines	57
4.1	Curve e Superfici	57
4.2	Curve di Bézier	59
4.3	Curve di Bézier Razionali	63
4.4	Superfici Prodotto di Tensori	66
4.5	Curve e Superfici Polinomiali o Razionali a Tratti	69
4.6	Derivate delle Funzioni B-Spline	72
4.7	Ulteriori Proprietá delle Funzioni di Base	74
4.8	Definizione e Proprietá delle Curve B-splines	76
4.9	Le Derivate di una Curva B-spline	77
4.10	Definizione e Proprietá di una Superficie B-spline	80
4.11	Derivate di una Superficie B-spline	82
5	NURBS	85
5.1	Definizione e Proprietá delle Curve NURBS	85
5.2	Derivate di una Curva NURBS	88
5.3	Definizione e Proprietá delle Superfici NURBS	89
5.4	Derivate di una Superficie NURBS	92
5.5	Alcuni Esempi di Utilizzo delle NURBS	93
5.5.1	Come Costruire una Rappresentazione NURBS di un Cilindro	94
5.5.2	Riposizionare un Punto di Controllo	95
5.5.3	Modificare un Peso	96
5.5.4	Aumentare il Grado di una Curva	96
5.6	Creazione di una Figura con Maya	97

Introduzione: il Problema dell'Interpolazione

Date $m+1$ coppie di valori (x_i, y_i) , assegnati, si parla di **interpolazione** quando si cerca una funzione $\Phi = \Phi(x)$ dipendente da x_0, x_1, \dots, x_m tale che $\Phi(x_i) = y_i$ per $i=0, \dots, m$. Le suddette coppie possono essere misure di un fenomeno naturale oppure gli y_i potranno essere i valori assunti nei **nodi** x_i da una funzione f nota analiticamente. Nel primo caso si vorrá determinare una ‘relazione quantitativa’ che descriva la legge del fenomeno in esame e che sia facilmente calcolabile. Nel secondo si cerca una funzione piú semplice da trattare in sostituzione di f . Si parlerá di **approssimazione polinomiale** se Φ é un polinomio algebrico o un polinomio trigonometrico; **approssimazione composita o mediante splines** se Φ é solo localmente un polinomio. In generale si parla di **interpolazione lineare** quando Φ puo’ essere scritta nella forma:

$$\Phi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_n\varphi_n(x) \quad (0.0.1)$$

ove $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ sono ‘funzioni elementari’ definite sull’intervallo $[a, b]$, che contiene tutti i punti di osservazione x_0, x_1, \dots, x_m , e ivi linearmente indipendenti, cioé tali che una loro combinazione lineare a coefficienti reali

$$\sum_{j=1}^n c_j \varphi_j(x)$$

sia identicamente nulla su $[a, b]$ se e solo se $c_j = 0$, $j = 0, \dots, n$. Inoltre, a_0, a_1, \dots, a_n sono numeri reali. La scelta delle ‘funzioni elementari’ $\varphi_j(x)$ deve tener conto di specifiche proprietá della funzione $f(x)$, inoltre si richiede che le funzioni $\varphi_j(x)$ siano facilmente calcolabili e dotate di buone proprietá di regolaritá.

Ai problemi di interpolazione lineare appartiene il problema di **interpolazione mediante polinomi**

$$\Phi(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

dove gli x_i sono definiti nell'asse reale $(-\infty, +\infty)$, e l'**interpolazione trigonometrica**

$$\Phi(x) = a_0 + a_1 e^{xi} + a_2 e^{2xi} + \dots + a_n e^{nxi} \quad (i^2 = -1)$$

Inoltre anche l'**interpolazione mediante funzioni spline** fa parte dei problemi di interpolazione lineare. Mentre fanno parte dei problemi di **interpolazione non lineare** l'**interpolazione mediante funzioni razionali**

$$\Phi(x) = \frac{a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n}{b_0 + b_1 x + b_2 x^2 + \dots + b_m x^m}$$

e mediante **somme esponenziali**

$$\Phi(x) = a_0 e^{\lambda_0 x} + a_1 e^{\lambda_1 x} + \dots + a_n e^{\lambda_n x}$$

Capitolo 1

Interpolazione Polinomiale

1.1 Il Polinomio di Interpolazione

In generale, una volta assegnati i **punti di osservazione** x_i e le **osservazioni** y_i , per $i=0,1,\dots,m$, il problema dell'approssimazione di una funzione mediante funzioni elementari, consiste nel determinare i coefficienti a_0, a_1, \dots, a_n che appaiono nella (0.0.1) imponendo la condizione che

$$\Phi(x_i) = a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \dots + a_n\varphi_n(x_i) \quad (1.1.1)$$

sia uguale o sia il piú possibile vicina a y_i per $i=0,1,\dots,m$.

Nel caso particolare in cui $\mathbf{m}=\mathbf{n}$, la condizione

$$y_i = a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \dots + a_n\varphi_n(x_i) \quad (1.1.2)$$

per $i=0,1,\dots,n$, fornisce il seguente sistema di $n+1$ equazioni lineari nelle $n+1$ incognite a_0, a_1, \dots, a_n

$$a_0\varphi_0(x_0) + a_1\varphi_1(x_0) + \dots + a_n\varphi_n(x_0) = y_0$$

$$a_0\varphi_0(x_1) + a_1\varphi_1(x_1) + \dots + a_n\varphi_n(x_1) = y_1$$

$$\dots \quad \dots \quad \dots \quad (1.1.3)$$

$$a_0\varphi_0(x_n) + a_1\varphi_1(x_n) + \dots + a_n\varphi_n(x_n) = y_n$$

che ammette un'unica soluzione se gli $n+1$ vettori a $n+1$ componenti

$$\varphi_0 = \begin{bmatrix} \varphi_0(x_0) \\ \varphi_0(x_1) \\ \dots \\ \varphi_0(x_n) \end{bmatrix} \quad \varphi_1 = \begin{bmatrix} \varphi_1(x_0) \\ \varphi_1(x_1) \\ \dots \\ \varphi_1(x_n) \end{bmatrix} \quad \dots \quad \varphi_n = \begin{bmatrix} \varphi_n(x_0) \\ \varphi_n(x_1) \\ \dots \\ \varphi_n(x_n) \end{bmatrix}$$

sono linearmente indipendenti. Considerando in particolare, $\varphi_k(x) = x^k$ per $k=0,1,\dots,n$, il sistema (1.1.3) fornisce i coefficienti di un polinomio di grado n (ordine $n+1$) che soddisfa la condizione (1.1.2)

$$a_0 + a_1x_i + a_2x_i^2 + \dots + a_nx_i^n = y_i \quad (1.1.4)$$

per $i=0,1,\dots,n$. Un tale polinomio si chiama **polinomio di interpolazione**.

1.1.1 La Matrice di Vandermonde

Teorema 1.1: se (x_i, y_i) , $i=0,\dots,n$, sono $n+1$ punti tali che $x_i \neq x_j$, per $i \neq j$, esiste ed è unico il polinomio Φ di grado al più n , tale che

$$\Phi(x_i) = a_0 + a_1x_i + \dots + a_nx_i^n = y_i, \quad i = 0, \dots, n.$$

Tale polinomio è detto **polinomio interpolatore** sui nodi x_i , $i=0,\dots,n$.

Dim: supponiamo, infatti, che oltre a $\Phi(x)$ esista un altro polinomio di interpolazione $\theta(x)$ di grado n . Allora, $\theta(x_i) = y_i$ per $i=0,1,\dots,n$. Ma, essendo anche $\Phi(x_i) = y_i$, si ha $\theta(x_i) = \Phi(x_i)$, ovvero $\theta(x_i) - \Phi(x_i) = 0$ per $i=0,1,\dots,n$. Quindi, il polinomio $\theta(x) - \Phi(x)$, che ha grado non superiore a n , si annulla negli $n+1$ punti distinti x_0, x_1, \dots, x_n dell'asse reale. Per il *principio di identità dei polinomi*, $\theta(x) - \Phi(x)$ è identicamente nullo. Pertanto, $\theta(x)$ è uguale a $\Phi(x)$ per qualsiasi valore della variabile x . Per dimostrare l'esistenza di $\Phi(x)$ si considera il vettore $a = [a_0, a_1, \dots, a_n]^T$ dei coefficienti del polinomio

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$

il vettore $y = [y_0, y_1, \dots, y_n]^T$ e la matrice V , detta **matrice di Vandermonde**, così definita

$$v_{ij} = x_{i-1}^{j-1}, \quad i=1, \dots, n+1, j=1, \dots, n+1.$$

Imponendo che $p_n(x)$ verifichi le $n+1$ condizioni $p_n(x_i) = y_i$, si ottiene il sistema lineare di $n+1$ equazioni in $n+1$ incognite

$$Va = y. \tag{1.1.5}$$

La matrice di Vandermonde é non singolare, in quanto

$$\det V = \prod_{i,j=0, i>j}^n (x_i - x_j)$$

e i punti x_i sono per ipotesi a due a due distinti. Ne segue che il sistema (1.1.5) ha una e una sola soluzione e quindi il polinomio $p_n(x)$ esiste ed è unico. Solo raramente del polinomio di interpolazione sono richiesti i coefficienti, in generale ciò che si vuole é il valore di $p_n(x)$ in uno o piú punti. Per calcolare $p_n(x)$ non é però conveniente risolvere il sistema (1.1.5) perché ciò richiederebbe un numero di operazioni moltiplicative dell'ordine di $n^3/3$. Inoltre la matrice di Vandermonde può essere mal condizionata (una variazione relativamente piccola negli elementi della matrice causa una variazione relativamente grande nella soluzione): in tal caso il calcolo di $p_n(x)$ attraverso la risoluzione del sistema (1.1.5) sarebbe numericamente instabile.

1.2 Interpolazione in Forma di Lagrange

Un altro modo per dimostrare l'esistenza del polinomio di interpolazione $\Phi(x)$ si ha procedendo in maniera costruttiva fornendo un'espressione dello stesso. Si utilizzano come funzioni $\varphi_j(x)$ i polinomi di grado n

$$l_j(x) = \prod_{i=0, i \neq j}^n \frac{x - x_i}{x_j - x_i}, \quad j = 0, 1, \dots, n. \quad (1.2.6)$$

che sono linearmente indipendenti in quanto

$$l_j(x_i) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Allora il polinomio

$$p_n(x) = \sum_{j=0}^n y_j l_j(x) \quad (1.2.7)$$

che ha grado al più n , è tale che

$$p_n(x_i) = \sum_{j=0}^n y_j l_j(x_i) = y_i \quad i=0,1,\dots,n$$

e quindi $p_n(x)$ soddisfa alle (1.1.2). Un algoritmo per calcolare il polinomio p_n nel punto $x = \xi$, usando le formule (1.2.6) e (1.2.7) (vedi [2]) è costituito dalle seguenti istruzioni

Dati : $\xi, n, \{x_0, x_1, \dots, x_n\}, \{y_0, y_1, \dots, y_n\}$

$p \leftarrow 0$

per $j = 0, 1, \dots, n$

$q \leftarrow 1$

per $i = 0, 1, \dots, n$

se $j \neq i$, $q \leftarrow q \cdot (\xi - x_i) / (x_j - x_i)$

$p \leftarrow p + y_j \cdot q$

Risultato : p

Definizione 1.1: il polinomio $p_n(x)$, rappresentato nella forma (1.2.7), è il polinomio di interpolazione nella **forma di Lagrange**.

Utilizzando la (1.2.7) il polinomio di interpolazione potrebbe essere trasformato nella forma $p_n(x) = a_n x^n + a_{n-1} x^{n-1} + a_1 x + a_0$, per calcolarne il valore in un punto x con il metodo di Ruffini-Horner, ma in questo modo sono richieste, per il solo calcolo dei coefficienti del polinomio, un numero di operazioni additive e moltiplicative superiore a $6n^2$. Un valore $p_n(x)$ per $x \neq x_j, j=0,1,\dots,n$, si ottiene invece con un minor numero di operazioni procedendo nel modo seguente: posto

$$\pi_n(x) = (x - x_0) \dots (x - x_n)$$

si rappresenta il polinomio nella forma

$$p_n(x) = \pi_n(x) \sum_{j=0}^n \frac{z_j}{x - x_j} \quad (1.2.8)$$

dove

$$z_j = \frac{y_j}{\prod_{i=0, i \neq j}^n (x_j - x_i)}$$

Le operazioni richieste (crf. [1]) sono, a meno di termini di ordine inferiore:

n sottrazioni per calcolare $x - x_j$, $j=0, \dots, n$;

$n^2/2$ sottrazioni per calcolare $x_j - x_i$, $j=0, \dots, n$, $i=j+1, \dots, n$;

n^2 moltiplicazioni per calcolare z_j , $j=0, \dots, n$;

n addizioni e $2n$ moltiplicazioni per calcolare $p_n(x)$.

In totale il calcolo di $p_n(x)$ in un punto x con il polinomio di Lagrange nella forma (1.2.8) richiede $n^2/2$ addizioni e n^2 moltiplicazioni. Una volta calcolato in un punto il polinomio di interpolazione di Lagrange sui nodi (x_i, y_i) , $i=0, 1, \dots, n$, é possibile calcolare nello stesso punto il polinomio di interpolazione di Lagrange sui nodi (x_i, y_i) , $i=0, 1, \dots, n+1$, utilizzando le quantità

$$\frac{z_j}{(x - x_j)}$$

per $j=0, 1, \dots, n$ e il valore di $\pi_n(x)$ già calcolato, con solo altre $2n$ addizioni e $2n$ moltiplicazioni. Infine cambiando i valori y_j ma lasciando inalterati i nodi x_j , con la (1.1.4) si calcola $p_n(x)$ con solo altre n addizioni e $2n$ moltiplicazioni. Per mezzo di $\pi_n(x)$ si può dare di $l_j(x)$ un'altra espressione più compatta che non viene in generale usata per il calcolo effettivo del polinomio di interpolazione. Poiché

$$\pi_n'(x_j) = \prod_{i=0, i \neq j}^n (x_j - x_i)$$

sostituendo nella (1.2.6) si ha

$$l_j(x) = \frac{\pi_n(x)}{(x-x_j)\pi_n'(x_j)}$$

essendo π_n il *polinomio nodale* di grado $n+1$. Il polinomio di Lagrange assume un'espressione particolarmente semplice quando i punti x_i sono *equidistanti di passo h* , cioè

$$x_i = x_0 + ih, \quad i = 0, 1, \dots, n, h > 0.$$

In questo caso, eseguendo il cambiamento di variabile

$$x = x_0 + th$$

e introducendo il polinomio di grado $n+1$

$$\tau_n(t) = t(t-1)\dots(t-n),$$

si ha dalla (1.2.8)

$$p_n(x_0 + th) = \tau_n(t) \sum_{j=0}^n \frac{z_j}{t-j}$$

dove

$$z_j = \frac{y_j}{\prod_{i=0, i \neq j}^n (j-i)}$$

Poiché

$$\prod_{i=0, i \neq j}^n (j-i) = (-1)^{n-j} \prod_{i=0}^{j-1} (j-i) \prod_{i=j+1}^n (i-j) = (-1)^{n-j} j!(n-j)!$$

risulta

$$p_n(x_0 + th) = \tau_n(t) \sum_{j=0}^n \frac{(-1)^{n-j} y_j}{j!(n-j)!(t-j)} = \frac{\tau_n(t)}{n!} \sum_{j=0}^n \binom{n}{j} (-1)^{n-j} \frac{y_j}{t-j}$$

Il costo computazionale per il calcolo di $p_n(x_0 + th)$ é allora di $n^2/4$ addizioni e $4n$ moltiplicazioni. Gran parte delle operazioni additive sono quelle richieste dal calcolo dei coefficienti binomiali, che conviene costruire, sfruttando le simmetrie, con il triangolo di Tartaglia. Se si considerano già noti i valori di $n!$ e dei coefficienti binomiali, il costo computazionale scende a $2n$ addizioni e $3n$ moltiplicazioni.

1.3 L'Errore di Interpolazione

Sostituendo ad una data funzione f il polinomio interpolatore $p_n(x)$ nei nodi x_0, x_1, \dots, x_n si hanno due tipi di errore: analitico, dato da $f(x) - p_n(x)$, e algoritmico. Quest'ultimo viene generato nel calcolo di $p_n(x)$, può diventare preponderante rispetto a quello analitico e dipende dalla particolare espressione del polinomio di interpolazione che viene usata. Si vuole stimare l'errore analitico con il seguente ([3]):

Teorema 1.2: siano x_0, x_1, \dots, x_n , $n+1$ nodi distinti e x un punto appartenente al dominio di una data funzione f . Supponiamo che $f \in C^{n+1}(I_x)$ essendo I_x il piú piccolo intervallo contenente i nodi x_0, x_1, \dots, x_n ed il punto x . Allora l'errore di interpolazione nel generico punto x é dato da

$$E_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \pi_n(x) \quad (1.3.9)$$

con $\xi \in I_x$ e dove π_n é il polinomio nodale di grado $n+1$.

Dim: il risultato é ovvio se x coincide con uno dei nodi di interpolazione. In caso contrario, definiamo per ogni $t \in I_x$, la funzione $G(t) = E_n(x) - \pi_n(t)E_n(x)/\pi_n(x)$. Essendo $f \in C^{(n+1)}(I_x)$ e π_n un polinomio, allora $G \in C^{(n+1)}(I_x)$ ed ha $n+2$ zeri distinti in I_x con

$$\begin{aligned} G(x_i) &= E_n(x_i) - \pi_n(x_i)E_n(x)/\pi_n(x) = 0, i = 0, 1, \dots, n \\ G(x) &= E_n(x) - \pi_n(x)E_n(x)/\pi_n(x) = 0 \end{aligned}$$

Per il teorema del valor medio allora G' avrà $n+1$ zeri distinti e, procedendo in modo ricorsivo, G^j presenterá $n+2-j$ zeri distinti. Di conseguenza G^{n+1} ammetterá un unico zero che denoteremo con ξ . D'altra parte, siccome $E_n^{(n+1)}(t) = f^{(n+1)}(t)$ e $\pi_n^{(n+1)}(x) = (n+1)!$, si ottiene

$$G^{(n+1)}(t) = f^{(n+1)}(t) - \frac{(n+1)!}{\pi_n(x)} E_n(x)$$

da cui, isolando $E_n(x)$, per $t = \xi$ si ricava la tesi.

Per analizzare il comportamento dell'errore di interpolazione dato dalla (1.3.9), quando n tende all'infinito, si definisce per ogni funzione $f \in C^0([a, b])$ la **norma infinito (o norma del massimo)** come

$$\|f\|_\infty = \max_{x \in [a, b]} |f(x)|.$$

Si introduce una matrice triangolare inferiore X di dimensione infinita, **matrice di interpolazione** su $[a, b]$, i cui elementi $x_{i,j}$, $i, j=0, 1, \dots$ rappresentano punti di $[a, b]$, e si assume che in ogni riga gli elementi siano tutti distinti. Allora, per ogni $n \geq 0$, la $n+1$ -esima riga di X contiene $n+1$ valori che si possono identificare come nodi di interpolazione e, per una data funzione f , si puó definire in modo univoco un polinomio interpolatore p_n di grado n rispetto a tali nodi. p_n dipenderá da X e da f . Quindi, una volta fissata f e la matrice X , l'errore di interpolazione sará

$$E_{n,\infty}(X) = \|f - p_n\|_\infty, \quad n = 0, 1, \dots$$

Se indichiamo con $p_n^* \in P_n$ il **polinomio di miglior approssimazione uniforme**, si ha

$$E_n^* = \|f - p_n^*\|_\infty \leq \|f - q_n\|_\infty \quad \forall q_n \in P_n$$

e la seguente

Proprietá: sia $f \in C^0([a, b])$ e X una matrice di interpolazione su $[a, b]$. Allora

$$E_{n,\infty}(X) \leq (1 + \Lambda_n(X)) E_n^*, \quad n = 0, 1, \dots$$

dove $\Lambda_n(X)$ denota la **costante di Lebesgue** di X definita come

$$\Lambda_n(X) = \left\| \sum_{j=0}^n |l_j^{(n)}| \right\|_\infty$$

ed essendo $l_j^{(n)} \in P_n$ il j -esimo polinomio caratteristico associato alla $n+1$ -esima riga di X , cioè tale che $l_j^{(n)}(x_{n,k}) = \delta_{j,k}$, $j,k=0,1,\dots$. Essendo E_n^* indipendente da X , tutte le informazioni relative agli effetti di X su $E_{n,\infty}(X)$ andranno cercate in $\Lambda_n(X)$. Pur esistendo una matrice di interpolazione X^* in corrispondenza della quale $\Lambda_n(X)$ è minimo, solo in rari casi se ne possono determinare esplicitamente gli elementi. D'altro canto, per ogni scelta possibile di X , esiste una costante $C > 0$ tale che

$$\Lambda_n(X) > \frac{2}{\pi} \ln(n+1) - C, \quad n = 0, 1, \dots$$

Questa proprietà mostra che $\Lambda_n(X) \rightarrow \infty$ per $n \rightarrow \infty$. In particolare si deduce che, per ogni matrice di interpolazione X su un intervallo $[a,b]$, esiste sempre una funzione f , continua su $[a,b]$, tale che p_n non converge uniformemente (cioè nella norma infinito) ad f . Quindi non è possibile approssimare tramite l'interpolazione polinomiale *tutte* le funzioni continue.

1.4 Formula di Newton

La forma di Lagrange del polinomio interpolatore non è la più conveniente da un punto di vista pratico. Vediamo una forma alternativa dal costo computazionale inferiore. Consideriamo una funzione $F(x)$ i cui valori $F(x_0), F(x_1), \dots, F(x_m)$ sono assegnati nei punti x_0, x_1, \dots, x_m dell'asse reale. Si suppone che questi punti siano distinti tra loro.

Definizione 1.2: si dice **differenza divisa** di $F(x)$ rispetto agli argomenti x_0, x_1 , si indica con $F[x_0, x_1]$ ed è definita dalla relazione

$$F[x_0, x_1] = \frac{F(x_0) - F(x_1)}{x_0 - x_1} = \frac{F(x_1) - F(x_0)}{x_1 - x_0} = F[x_1, x_0]$$

Analogamente, si definisce la differenza divisa di $F(x)$ rispetto agli argomenti x_1, x_2 e così via. Due differenze divise di due argomenti aventi un argomento in comune possono essere usate per definire una differenza divisa, di ordine 2, rispetto ai tre argomenti. Allora, la differenza divisa di $F(x)$ rispetto ai tre argomenti x_0, x_1, x_2 , indicata con $F[x_0, x_1, x_2]$, è definita dalla relazione

$$F[x_0, x_1, x_2] = \frac{F[x_0, x_1] - F[x_1, x_2]}{x_0 - x_2} = \frac{F[x_0, x_1] - F[x_2, x_1]}{x_0 - x_2}$$

Procedendo in questo modo si possono definire le differenze divise di $F(x)$ rispetto ad $m+1$ argomenti quando sono state definite le differenze divise di $F(x)$ rispetto ad m argomenti. Allora

$$F[x_0, x_1, \dots, x_m] = \frac{F[x_0, x_1, \dots, x_{m-1}] - F[x_1, x_2, \dots, x_m]}{x_0 - x_m}$$

Si dice che $F[x_0, x_1, \dots, x_m]$ é una differenza divisa di $F(x)$ di ordine m . Dalle suddette relazioni si ricava anche il seguente:

Teorema 1.3: le differenze divise $F[x_0, x_1, \dots, x_m]$ sono funzioni simmetriche dei loro argomenti, cioè sono invarianti per permutazioni di x_0, x_1, \dots, x_m .

Le relazioni delle differenze divise possono essere riassunte nella seguente tabella

$$\begin{array}{ccccccc}
 x_0 & F(x_0) = F[x_0] & & & & & \\
 x_1 & F(x_1) = F[x_1] & F[x_0, x_1] & & & & \\
 x_2 & F(x_2) = F[x_2] & F[x_1, x_2] & F[x_0, x_1, x_2] & & & \\
 x_3 & F(x_3) = F[x_3] & F[x_2, x_3] & F[x_1, x_2, x_3] & & F[x_0, x_1, x_2, x_3] & \\
 & \dots & & \dots & & & \\
 x_m & F(x_m) = F[x_m] & F[x_{m-1}, x_m] & F[x_{m-2}, x_{m-1}, x_m] & & &
 \end{array}$$

Esiste quindi un algoritmo (vedi [2]) per determinare le differenze divise di $F(x)$ di ordine k , $F[x_0, x_1, \dots, x_k]$, per ogni $k=0, 1, \dots, m$, assegnati gli argomenti x_0, x_1, \dots, x_m e i valori $F(x_0), F(x_1), \dots, F(x_m)$ della funzione $F(x)$. Per semplicitá di scrittura si pone $d_k = F[x_0, x_1, \dots, x_k]$, $k=0,1,\dots,m$.

Dati : $m, x_0, x_1, \dots, x_m, F(x_0), F(x_1), \dots, F(x_m)$

per $k = 0, 1, \dots, m$

$d_k \leftarrow F(x_k)$

per $i = 1, 2, \dots, m$

per $k = m, m-1, \dots, i$

$d_k \leftarrow (d_k - d_{k-1}) / (x_k - x_{k-i})$

Risultato : $\{d_0, d_1, \dots, d_m\}$

La complessitá computazionale di questo algoritmo si determina osservando che per il calcolo dell'espressione nel ciclo piú interno occorre effettuare due sottrazioni e una divisione. Questa espressione deve essere calcolata

$$m + (m-1) + \dots + 1 = \frac{m(m+1)}{2}$$

volte. Pertanto, il numero totale di sottrazioni é $m(m+1)$ e il numero totale di divisioni é $\frac{m(m+1)}{2}$. Qualora fosse disponibile una nuova valutazione della funzione in un nuovo nodo, il polinomio di grado $m+1$ richiederá solo il calcolo di una nuova riga della tabella vista sopra, per avere il polinomio di grado $m+1$ si avrá quindi un costo computazionale aggiuntivo pari a $m+1$ divisioni e $2(m+1)$ addizioni.

Teorema 1.4: la differenza divisa di $F(x)$ di ordine m rispetto agli argomenti x_0, x_1, \dots, x_m , distinti tra loro, si può esprimere nella forma

$$F[x_0, x_1, \dots, x_m] = \sum_{j=0}^m \frac{F(x_j)}{(x_j - x_0)(x_j - x_1) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_m)} \quad (1.4.10)$$

Le differenze divise godono di alcune proprietà interessanti :

- come già visto sono invarianti per permutazioni degli indici dei nodi. Questo fatto si può sfruttare opportunamente quando problemi di stabilità suggeriscono un'opportuna permutazione degli indici;
- se $f = \alpha g + \beta h$ per $\alpha, \beta \in \mathfrak{R}$, allora

$$f[x_0, \dots, x_m] = \alpha g[x_0, \dots, x_m] + \beta h[x_0, \dots, x_m];$$

- se $f=gh$ vale la seguente formula (detta **formula di Leibnitz**)

$$f[x_0, \dots, x_m] = \sum_{j=0}^m g[x_0, \dots, x_j] h[x_j, \dots, x_m];$$

Usando le differenze divise é possibile fornire una rappresentazione del polinomio di interpolazione che é molto conveniente nelle applicazioni. Siano assegnati i punti di osservazione x_0, x_1, \dots, x_n distinti tra loro e siano assegnate le corrispondenti osservazioni y_0, y_1, \dots, y_n . Sia $y=F(x)$ una funzione, definita nell'intervallo $[a,b]$, che nei punti x_i assume i valori $y_i = F(x_i)$, $i=0,1,\dots,n$; l'intervallo $[a,b]$ contiene tutti i punti x_0, x_1, \dots, x_n . Come già dimostrato, esiste un solo polinomio di interpolazione $f(x)$ di grado n che soddisfa la condizione $f(x_i) = y_i$ per $i=0,1,\dots,n$. Tale polinomio può essere espresso nella forma

$$f(x) = F(x_0) + (x - x_0)F[x_0, x_1] + (x - x_0)(x - x_1)F[x_0, x_1, x_2] + \dots + (x - x_0)(x - x_1) \dots (x - x_{n-1})F[x_0, x_1, \dots, x_n], \quad a \leq x \leq b \quad (1.4.11)$$

Questa formula, detta **formula di Newton**, si ottiene applicando ricorsivamente il seguente:

Teorema 1.5: sia $f_k(x)$ il polinomio di interpolazione di grado k che soddisfa la condizione $f_k(x_i) = y_i$ per $i=0, 1, \dots, k$. Allora la funzione

$$f_{k+1}(x) = f_k(x) + (x - x_0)(x - x_1) \dots (x - x_k)F[x_0, x_1, \dots, x_{k+1}], \quad a \leq x \leq b \quad (1.4.12)$$

é il polinomio di grado $k+1$ che soddisfa la condizione $f_{k+1}(x_i) = y_i$ per $i=0,1,\dots,k+1$.

Dim: sia $g(x)$ il polinomio di grado $k+1$ che soddisfa la condizione $g(x_i) = y_i$ per $i=0,1,\dots,k+1$. Si ha, per $i=0,1,\dots,k$:

$$\begin{aligned} g(x_i) - f_{k+1}(x_i) &= g(x_i) - f_k(x_i) - (x_i - x_0)(x_i - x_1) \dots (x_i - x_i) \dots (x_i - x_k) \cdot \\ &\quad \cdot F[x_0, x_1, \dots, x_{k+1}] = y_i - y_i - 0 = 0 \end{aligned} \quad (1.4.13)$$

Inoltre essendo $F[x_0, x_1, \dots, x_{k+1}]$ il coefficiente principale del polinomio di grado $k+1$

$$(x - x_0)(x - x_1) \dots (x - x_k)F[x_0, x_1, \dots, x_{k+1}]$$

per la formula (1.4.10), tale coefficiente coincide con il coefficiente principale del polinomio di interpolazione $g(x)$ di grado $k+1$. Allora $g(x)$ e $f_{k+1}(x)$ hanno il medesimo coefficiente principale e quindi $g(x) - f_{k+1}(x)$ é un polinomio di grado non superiore a k . Poiché questo polinomio si annulla nei $k+1$ punti x_0, x_1, \dots, x_k , esso é identicamente nullo; cioè $g(x) = f_{k+1}(x)$ per ogni $x \in [a, b]$.

Il polinomio di Newton si può anche scrivere nella forma:

$$\begin{aligned} p_n(x) &= \{ \dots \{ f[x_0, x_1, \dots, x_n] (x - x_{n-1}) + f[x_0, x_1, \dots, x_{n-1}] \} (x - x_{n-2}) + \dots + \\ &\quad f[x_0, x_1] \} (x - x_0) + f[x_0] \end{aligned}$$

Le operazioni richieste per il calcolo del polinomio di interpolazione in un punto x , una volta costruita la tabella delle differenze divise, sono quindi $2n$ addizioni e n moltiplicazioni.

Indicando con A le ‘addizioni’ e con M le ‘moltiplicazioni’ richieste per questo calcolo, si mettono a confronto i costi computazionali relativi ai polinomi di Newton e di Lagrange:

<u>Tipo polinomio</u>	<i>Polinomio di Lagrange</i>	<i>Polinomio di Newton</i>
<i>calcolo in un punto</i>	$\frac{n^2}{2}A + n^2M$	$n^2A + \frac{n^2}{2}M$
<i>calcolo in m punti</i>	$(\frac{n^2}{2} + 2mn)A + (n^2 + 2mn)M$	$(n^2 + 2mn)A + (\frac{n^2}{2} + 2mn)M$
<i>aggiunta di un nodo</i>	$2nA + 2nM$	$2nA + 2nM$

Dati i punti base $(x_i, f(x_i))$, con $i=0,1,\dots,n$, il calcolo delle differenze divise $f[x_0, x_1, \dots, x_i]$ si può fare con un algoritmo (cfr. [4]), applicabile poi in **Scilab** ([9]), del tipo:

```

for i=0 to n do
begin t[i]=f[i];
  for k=i-1 to 0 do
    t[k]=(t[k+1]-t[k])/(x[i]-x[k]);
  a[i]=t[0]
end;

```

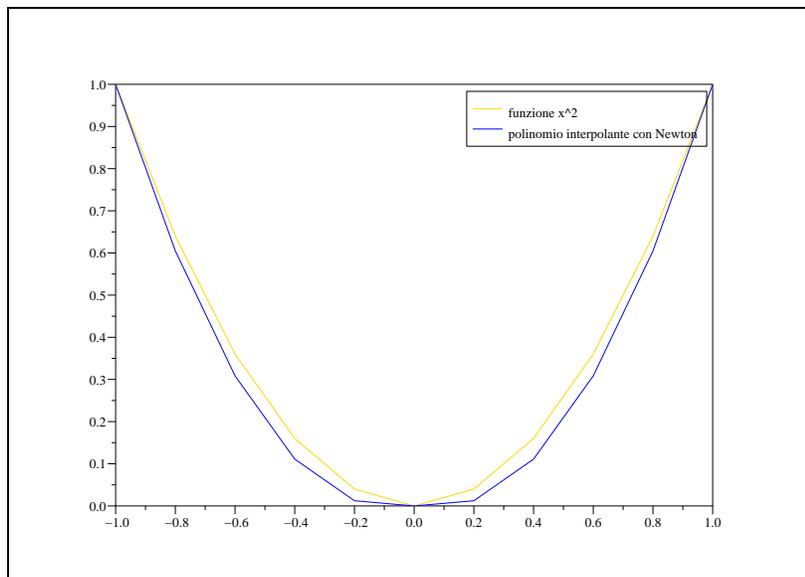
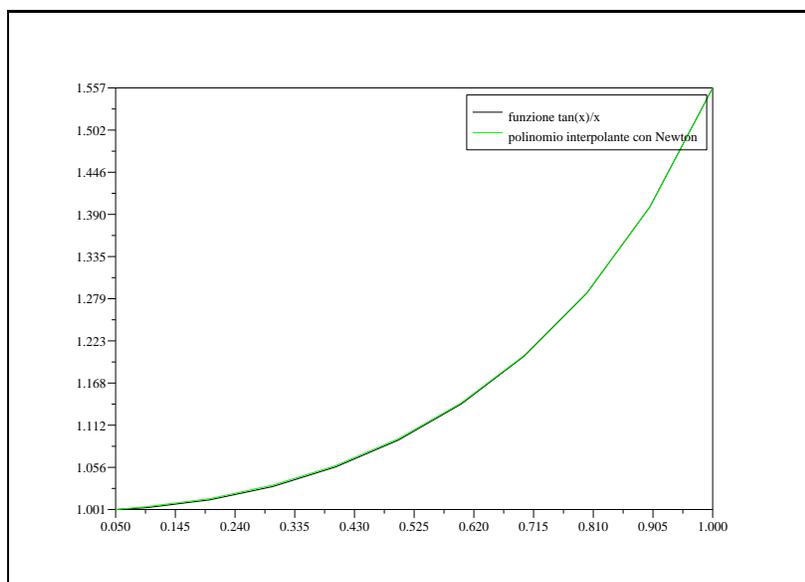
mentre il polinomio interpolatore in qualsiasi punto si ottiene con un algoritmo:

```

p=a[n];
for i=n-1 to 0 do
  p=p*(z-x[i])+a[i];

```

Nelle figure 1.1 e 1.2 vengono applicati questi algoritmi alle funzioni x^2 e $\frac{\tan(x)}{x}$.

Figura 1.1: x^2 Figura 1.2: $\tan(x)/x$

1.4.1 L'Errore Usando le Differenze Divise

Si considerino i nodi x_0, x_1, \dots, x_n e $p_n(x)$ il polinomio interpolatore di f su tali nodi. Sia ora x un nodo distinto dai precedenti; posto $x_{n+1} = x$, si denoti con $p_{n+1}(t)$ il polinomio interpolatore sui nodi $x_k, k=0, \dots, n+1$. Usando la formula delle differenze divise di Newton si trova

$$p_{n+1}(t) = p_n(t) + (t - x_0) \dots (t - x_n) F[x_0, \dots, x_n, t]$$

Essendo $p_{n+1}(x) = f(x)$, si ricava la seguente formula dell'errore di interpolazione in $t=x$

$$E_n(x) = f(x) - p_n(x) = p_{n+1}(x) - p_n(x) = (x - x_0) \dots (x - x_n) F[x_0, \dots, x_n, x] = \pi_n(x) F[x_0, \dots, x_n, x].$$

Supponendo $f \in C^{(n+1)}(I_x)$, dal confronto con la (1.3.9) si trova che

$$F[x_0, \dots, x_n, x] = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

con $\xi \in I_x$. Per la somiglianza di quest'ultima con il resto dello sviluppo in serie di Taylor, la formula di Newton del polinomio interpolatore viene paragonata ad uno sviluppo troncato intorno ad x_0 a patto che $|x_n - x_0|$ non sia troppo grande.

1.4.2 Il Fenomeno di Runge

Ritornando a considerare l'errore di interpolazione definito nel Teorema 1.2, cioè $E_n(x) = f(x) - p_n(x)$ é plausibile congetturare che, al crescere di n e quindi del numero delle osservazioni y_i relative alla funzione $f(x)$, questo si riduca sempre piú. Si vede con un esempio, dovuto a Runge, che questa congettura, in generale, é falsa. Si considera la **funzione di Runge** nell'intervallo $[a, b] = [-5, 5]$

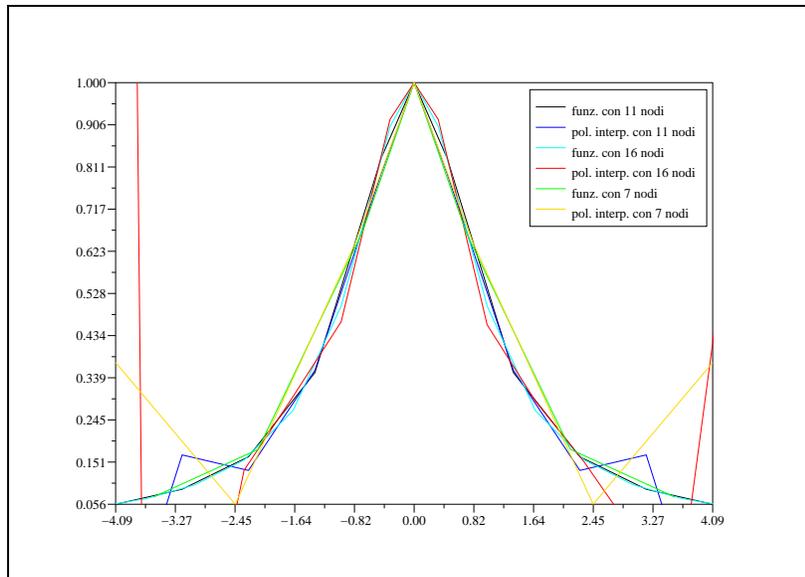
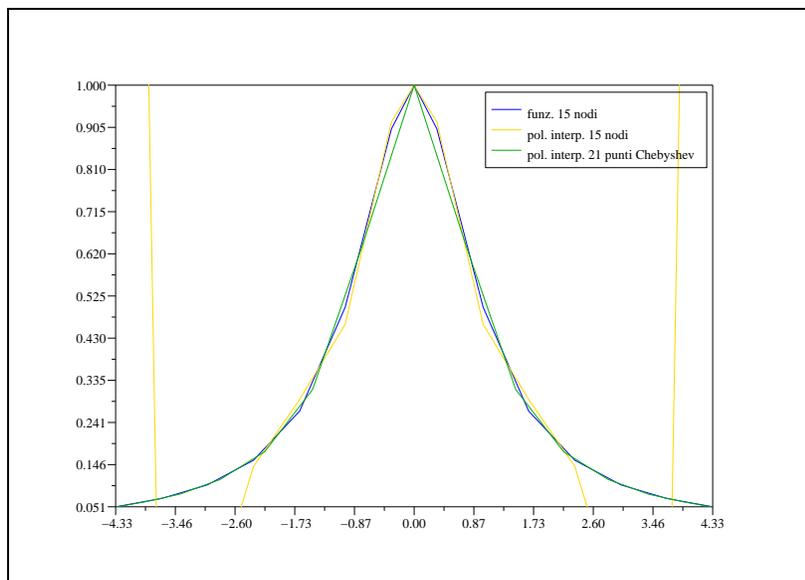
$$f(x) = \frac{1}{1+x^2}, \quad -5 \leq x \leq 5$$

e si considerano i punti $x_i = a + \frac{i(b-a)}{n}$ per $i=0,1,\dots,n$ ove n é un intero positivo. Tali punti sono equidistribuiti in $[a,b]$ e la distanza tra due punti consecutivi é $\frac{(b-a)}{n}$. Poiché $f(-x)=f(x)$, la funzione di Runge é pari e quindi é simmetrica rispetto all'asse delle ordinate. Sia $p_n(x)$ il polinomio di interpolazione di grado n che soddisfa le $n+1$ condizioni $p_n(x_i) = y_i = f(x_i)$ nei punti $x_i = a + \frac{i(b-a)}{n}$. Si vuol studiare l'errore di interpolazione $E_n(x) = f(x) - p_n(x)$ per $x \in [-5, 5]$. $|f(x) - p_n(x)|$ cresce per x prossimo agli estremi dell'intervallo al crescere del grado n del polinomio di interpolazione e quindi all'infittirsi dei punti x_i , quando questi sono equidistanti nell'intervallo $[a,b]$. Si può altresí osservare che, se i punti x_i sono scelti secondo la legge

$$x_i = \cos \left(\frac{(2(n-i)+1)\pi}{2(n+1)} \right), \quad i = 0, 1, \dots, n$$

(i punti x_i sono gli $(n+1)$ zeri reali del polinomio di Chebyshev di grado $n+1$ definito nell'intervallo $[-1,1]$) l'errore di interpolazione, in valore assoluto, decresce al crescere del grado del polinomio di interpolazione; e per $n \rightarrow \infty$ la matrice di interpolazione, X , descritta a pagg. 16 e 17 ha una costante di Lebesgue che cresce lentamente con n .

La figura 1.3 fornisce il grafico della funzione di Runge $y = \frac{1}{1+x^2}$, il polinomio di interpolazione con 11, 16 e 7 nodi equidistribuiti, mentre nella figura 1.4 c'è il polinomio di interpolazione della stessa funzione con 15 punti di osservazione equidistribuiti e con 21 zeri di Chebyshev. Questo risultato mostra che é possibile ottenere in alcuni casi una 'approssimazione' di una funzione $f(x)$ sempre meno accurata quanto piú elevato é il grado n del polinomio di interpolazione, ovvero quanto piú fitti sono i punti x_i in cui sono assegnati i valori della funzione. Mentre se i punti x_i sono scelti in modo opportuno, l'"approssimazione" diventa sempre piú accurata quanto piú elevato é il numero n . Per evitare tali fenomeni di crescita dell'errore di interpolazione all'aumentare delle osservazioni y_i , se queste non sono fornite in punti di osservazione distribuiti sull'asse reale in modo opportuno, si studiano i **polinomi di interpolazione a tratti**. Tra questi i piú semplici sono i polinomi di interpolazione a tratti di **primo grado**.

Figura 1.3: *Funzione di Runge*Figura 1.4: *Funzione di Runge negli zeri di Chebyshev*

Capitolo 2

Interpolazione Razionale

2.1 Proprietá Generali dell'Interpolazione Razionale

La classe delle funzioni razionali é piú ampia di quella dei polinomi e consente di approssimare meglio funzioni che alternano, ad un comportamento quasi rettilineo, brusche variazioni. Siano $p(x)$ un polinomio di grado massimo m e $q(x)$ un polinomio di grado massimo n e si consideri la funzione razionale di grado massimo $m+n$:

$$\omega(x) = \frac{p(x)}{q(x)} = \frac{a_m x^m + a_{m-1} x^{m-1} + \dots + a_0}{b_n x^n + b_{n-1} x^{n-1} + \dots + b_0} \quad (2.1.1)$$

La coppia di interi (m,n) puó essere chiamata il tipo di grado del problema di interpolazione razionale. Dati $m+n+1$ punti distinti $x_i, i=0,1,\dots,m+n$, e i corrispondenti valori $f(x_i)$ della funzione $f(x)$, si dice che la funzione razionale $\omega(x)$ é di interpolazione della funzione $f(x)$ se

$$\omega(x_i) = f(x_i), i = 0, \dots, m + n \quad (2.1.2)$$

Si impongono quindi $m+n+1$ condizioni, cioé quante se ne potrebbero imporre con un polinomio di grado $m+n$. La funzione razionale $\omega(x)$ é determinata dai suoi $m+n+2$ coefficienti $a_m, a_{m-1}, \dots, a_0, b_n, b_{n-1}, \dots, b_0$. D'altra parte, la (2.1.2) determina questi coefficienti

solo a meno di un fattore costante $\rho \neq 0$. Ciò suggerisce che $\omega(x)$ é totalmente determinato dalle $m+n+1$ condizioni di interpolazione $\omega(x_i) = f(x_i)$. Moltiplicando entrambi i membri della (2.1.2) per $q(x_i)$ si ottiene il sistema lineare omogeneo:

$$a_m x_i^m + a_{m-1} x_i^{m-1} + \dots + a_0 - f(x_i)(b_n x_i^n + b_{n-1} x_i^{n-1} + \dots + b_0) = 0 \text{ per } i = 0, \dots, m+n \quad (2.1.3)$$

di $m+n+1$ equazioni nelle $m+n+2$ incognite a_m, a_{m-1}, \dots, a_0 e b_n, b_{n-1}, \dots, b_0 . Le soluzioni del sistema (2.1.3) per cui $b_n x_i^n + b_{n-1} x_i^{n-1} + \dots + b_0 \neq 0$, per $i=0,1,\dots,m+n$, sono anche soluzioni del sistema (2.1.2), però in generale non é detto che i due sistemi siano equivalenti. Per chiarire questo fatto si veda il seguente

ESEMPIO:

dati i punti di osservazione $x_0 = 0$, $x_1 = 1$ e $x_2 = 2$, con corrispondenti osservazioni $f(x_0) = 1$, $f(x_1) = 2$ e $f(x_2) = 2$ e $m=n=1$ si ha il sistema:

$$\begin{aligned} a_0 - 1 \cdot b_0 &= 0 \\ a_0 + a_1 - 2(b_0 + b_1) &= 0 \\ a_0 + 2a_1 - 2(b_0 + 2b_1) &= 0 \end{aligned}$$

che mi da come soluzione del problema di interpolazione razionale

$$\omega(x) = \frac{2x}{x}$$

che, tuttavia, per $x=0$ porta all'espressione $\frac{0}{0}$. Invece semplificando x si ottiene la soluzione

$$\bar{\omega}(x) = 2$$

la quale 'perde' il punto $(0,1)$. Quindi il problema non ha soluzione.

Si noti che non puó esistere una soluzione non nulla di (2.1.3) in cui i coefficienti b_i , $i=0,1,\dots,n$, siano tutti nulli, perché se per assurdo ciò fosse vero, ne seguirebbe che il polinomio $p(x)$ di grado al piú m avrebbe $m+n+1$ zeri distinti. Perciò il polinomio $q(x)$ al denominatore non puó essere mai identicamente nullo. Se la matrice del sistema (2.1.3) ha rango massimo, la sua soluzione é unica a meno di una costante moltiplicativa e la funzione $\omega(x)$ risulta determinata solo da $m+n+1$ parametri, in quanto il numeratore e il denominatore possono essere divisi per la suddetta costante non nulla, senza che

cambi il valore di $\omega(x)$. Per questo si considerano spesso funzioni razionali della forma (2.1.1), in cui uno dei due polinomi $p(x)$ o $q(x)$ é **monico** (cioé con coefficiente del termine di grado massimo uguale a 1). Se invece la matrice non ha rango massimo, vi sono infinite soluzioni. Inoltre dato che i sistemi (2.1.2) e (2.1.3) possono non essere equivalenti, contrariamente a quanto accade per l'interpolazione polinomiale, non é detto che il problema dell'interpolazione razionale abbia sempre una soluzione. Per evitare che la funzione $\omega(x)$ abbia piú di una rappresentazione si impone che $p(x)$ e $q(x)$ non abbiano quali fattori comuni polinomi di grado superiore o uguale a 1, in altre parole si richiede che $\omega(x)$ sia **irriducibile**. Per questo si enuncia il seguente:

Teorema 2.1: fissati i due interi m e n , siano $p_1(x)$ e $p_2(x)$ due polinomi di grado m , $q_1(x)$ e $q_2(x)$ due polinomi di grado n e siano

$$\omega_1(x) = \frac{p_1(x)}{q_1(x)} \quad \omega_2(x) = \frac{p_2(x)}{q_2(x)}$$

due funzioni razionali ottenute risolvendo il sistema (2.1.3), tali che $q_1(x_i), q_2(x_i) \neq 0$ per $i=0,1,\dots,m+n$. Allora $\omega_1(x)$ e $\omega_2(x)$ soddisfano alla (2.1.2), e se $\omega_1(x)$ e $\omega_2(x)$ sono irriducibili, allora $p_1(x) = \alpha p_2(x)$, $q_1(x) = \alpha q_2(x)$ per una opportuna costante $\alpha \neq 0$, cioé la soluzione del problema di interpolazione razionale é unica a meno di costanti comuni a numeratore e denominatore.

Dim: poiché

$$\frac{p_1(x_i)}{q_1(x_i)} = \frac{p_2(x_i)}{q_2(x_i)} = f(x_i), \quad i = 0, \dots, m+n$$

il polinomio, di grado non superiore a $m+n$,

$$t(x) = p_1(x)q_2(x) - p_2(x)q_1(x)$$

é tale che $t(x_i) = 0$, per $i=0,1,\dots,m+n$, e quindi é identicamente nullo. Se $\omega_1(x)$ e $\omega_2(x)$ sono irriducibili, segue che esiste $\alpha \neq 0$ per cui

$$p_1(x) = \alpha p_2(x) \quad q_1(x) = \alpha q_2(x).c.v.d.$$

Il problema di interpolazione razionale puó avere soluzione per certe scelte di m e n , e non avere soluzione per altre scelte di m e n . Infine diciamo che l'espressione razionale $\omega(x)$ é una soluzione della (2.1.3) se i suoi coefficienti la risolvono. Come si é visto $\omega(x)$ risolve la (2.1.3) se risolve la (2.1.2). L'interpolazione razionale é complicata dal fatto che l'inverso non necessariamente esiste.

Teorema 2.2: il sistema lineare omogeneo di equazioni (2.1.3) ha sempre soluzioni non banali. Per ogni soluzione

$$\omega(x) = \frac{p(x)}{q(x)}$$

con $q(x) \neq 0$, cioè tutte le soluzioni non banali definiscono espressioni razionali.

Dim: il sistema lineare omogeneo (2.1.3) ha $m+n+1$ equazioni per $m+n+2$ incognite. Come un sistema lineare omogeneo con più incognite che equazioni, il (2.1.3) ha soluzioni non banali

$$(a_m, a_{m-1}, \dots, a_0, b_n, b_{n-1}, \dots, b_0) \neq (0, \dots, 0).$$

Per ogni soluzione, $q(x) \neq 0$, da

$$q(x) = b_n x^n + b_{n-1} x^{n-1} + \dots + b_0 = 0$$

si avrebbe che il polinomio

$$p(x) = a_m x^m + a_{m-1} x^{m-1} + \dots + a_0$$

ha gli zeri

$$p(x_i) = 0, \quad i = 0, 1, \dots, m+n.$$

Seguirebbe che $p(x)=0$, poiché il polinomio p ha al massimo grado m , e sparisce per $m+n+1 \geq m+1$ differenti posizioni, contraddicendo

$$(a_m, a_{m-1}, \dots, a_0, b_n, b_{n-1}, \dots, b_0) \neq (0, \dots, 0).$$

Dagli ultimi due teoremi ricaviamo che esiste per ogni problema di interpolazione razionale (lo chiamiamo A) un'unica funzione razionale, che è rappresentata da ogni espressione razionale ω , che risolve il corrispondente sistema lineare (2.1.3). Quindi questa funzione razionale soddisfa la (2.1.2), risolvendo il problema A per mezzo di essa, o tale problema non è risolvibile del tutto. Nell'ultimo caso, ci devono essere alcuni *punti di supporto* $(x_i, f(x_i))$ che sono stati 'persi' dalla funzione razionale. Tali punti di supporto sono detti **inaccessibili**. In questo modo A è risolvibile se non ci sono punti inaccessibili. Supponiamo $\omega(x) = \frac{p(x)}{q(x)}$ sia una soluzione alla (2.1.3). Per ogni $i \in \{0, 1, \dots, m+n\}$ distinguiamo due casi:

1. $q(x_i) \neq 0$;
2. $q(x_i) = 0$.

Nel primo caso, chiaramente, $\omega(x_i) = f(x_i)$. Nel secondo caso, tuttavia, il punto di supporto $(x_i, f(x_i))$ può essere inaccessibile. Qui $p(x_i) = 0$ può mantenere $p(x_i) - f(x_i)q(x_i) = 0$. Quindi, sia p che q contengono, per $i=0,1,\dots,m+n$, il fattore $x-x_i$ e sono conseguentemente non primi tra loro. Così:

(a) *Se (2.1.3) ha una soluzione ω che é relativamente prima, allora non ci sono punti inaccessibili: A é risolubile.*

(**Definizione 2.1:** Un'espressione razionale si dice **relativamente prima** se numeratore e denominatore sono primi fra loro, cioè non entrambi divisibili per lo stesso polinomio di grado positivo). Dato ω , sia ϖ un'espressione razionale equivalente che é relativamente prima. Allora abbiamo il risultato generale:

Teorema 2.3: supponiamo ω risolva la (2.1.3). Allora A é risolubile e ω é la soluzione se e solo se ϖ risolve la (2.1.3).

Dim: se ω risolve la (2.1.3), allora A é risolubile tramite la (a). Se ϖ non risolve la (2.1.3), la sua funzione razionale corrispondente non risolve A.

Anche se il sistema lineare (2.1.3) ha rango pieno $m+n+1$, il problema di interpolazione razionale A non può essere risolubile. Tuttavia, finché le soluzioni di (2.1.3) sono, in questo caso, univocamente determinate a meno di un fattore costante comune $\rho \neq 0$ abbiamo:

Corollario: se (2.1.3) ha rango pieno, allora A é risolubile se e solo se la soluzione ω di (2.1.3) é relativamente prima.

Definizione 2.2: diciamo che i punti di supporto $(x_i, f(x_i))$ $i=0,1,\dots,s$ sono in **posizione speciale** se sono interpolati da un'espressione razionale di tipo di grado (κ, λ) con $(\kappa + \lambda < s)$. In altre parole, il problema di interpolazione é risolubile per un grado combinato piú piccolo del numeratore e del denominatore di quello suggerito dal numero di punti di supporto. Osserviamo che:

(b) *I punti di supporto accessibili di un problema di interpolazione non risolubile A sono in posizione speciale.*

Dim: siano i_1, \dots, i_α gli indici dei punti inaccessibili, e sia ω una soluzione di (2.1.3). Il numeratore e il denominatore di ω erano visti sopra come aventi fattori comuni $x - x_{i_1}, \dots, x - x_{i_\alpha}$, la cui cancellazione porta a un'espressione razionale equivalente ω (gradi κ e λ) con $\kappa = m - \alpha, \lambda = n - \alpha$. ω risolve il problema di interpolazione A (con gradi κ e λ) che consiste proprio di $m + n + 1 - \alpha$ punti accessibili. Poiché

$$\kappa + \lambda + 1 = m + n + 1 - 2\alpha < m + n + 1 - \alpha$$

i punti accessibili di A (con gradi m e n) sono chiaramente in posizione speciale.

L'osservazione (b) chiarisce che la non risolvibilità del problema di interpolazione razionale é un fenomeno degenerante: la risolvibilità puó essere ripristinata con arbitrarie piccole perturbazioni dei punti di supporto. Quindi restringeremo la nostra attenzione a problemi completamente non degenerati, cioè problemi per i quali nessun sottoinsieme dei punti di supporto é in posizione speciale. Non solo A (con gradi μ e ν) é risolubile in questo caso, ma lo sono tutti i problemi A (con gradi κ e λ) di $\kappa + \lambda + 1$ dei punti di supporto originali dove $\kappa + \lambda \leq \mu + \nu$. Di seguito si riportano le procedure ricorsive per risolvere problemi di interpolazione razionale A (con gradi m e n).

2.2 Frazioni Continue

Una **frazione continua** finita di ordine k é un'espressione della forma

$$\omega(x) = d_0 + \frac{c_1}{d_1 + \frac{c_2}{d_2 + \frac{c_3}{\dots + \frac{c_k}{d_k}}}} \quad (2.2.4)$$

in cui i c_i e d_i sono polinomi a coefficienti reali, $c_i \neq 0$ per $i=1, \dots, k$ e $d_k \neq 0$. Per semplificare la formula si puó scrivere in questo modo:

$$\omega(x) = d_0 + \frac{c_1}{d_1} + \frac{c_2}{d_2 + \dots} + \frac{c_k}{d_k} \quad (2.2.5)$$

Indicata con ω_i , $i \leq k$, la i-esima **frazione parziale**, cioè

$$\omega_i = d_0 + \frac{c_1}{d_1} + \frac{c_2}{d_2 + \dots} + \frac{c_i}{d_i}$$

é facile verificare che ω_i é una funzione razionale della forma

$$\omega_i = \frac{p_i}{q_i}$$

dove p_i e q_i soddisfano le relazioni ricorrenti:

$$\begin{aligned} p_{-1} &= 1 & q_{-1} &= 0 & p_0 &= d_0 & q_0 &= 1 \\ p_i &= d_i p_{i-1} + c_i p_{i-2} & \text{per } i &= 1, \dots, k \\ q_i &= d_i q_{i-1} + c_i q_{i-2} \end{aligned} \tag{2.2.6}$$

Quindi i p_i e i q_i sono soluzioni della medesima equazione alle differenze, con condizioni iniziali diverse. Poiché

$$\omega(x) = \omega_k = \frac{p_k}{q_k}$$

$\omega(x)$ risulta una funzione della forma (2.1.1). Indicato con z_i l' i -esimo **residuo**

$$z_i = d_i + \frac{c_{i+1}}{d_{i+1} + \dots + \frac{c_k}{d_k}}, \text{ per } i = 0, \dots, k-1 \text{ e } z_k = d_k, \tag{2.2.7}$$

vale la relazione

$$z_i = d_i + \frac{c_{i+1}}{z_{i+1}}, \quad i = 0, \dots, k-1 \tag{2.2.8}$$

che può essere utilizzata come algoritmo di calcolo della funzione $\omega(x) = z_0$, a partire da $z_k = d_k$, facendo variare l'indice i da $k-1$ fino a 0 . La (2.2.8) presenta però lo svantaggio che non si può calcolare ω_{i+1} a partire da ω_i , cosa che è invece possibile fare utilizzando la (2.2.6), però con un costo superiore. Si hanno altri algoritmi sfruttando il fatto che il valore di una frazione continua può essere ricavato con un opportuno sistema lineare. Una frazione razionale $\omega(x)$ può essere rappresentata in diversi modi sotto forma di frazione continua. Il seguente teorema mostra come $\omega(x)$ sia rappresentabile nella forma (2.2.4) in cui i c_i , $i=1, \dots, k$, sono delle costanti. Con questa rappresentazione il calcolo di $\omega(x)$ risulta in generale più efficiente. Successivamente si vedrà come $\omega(x)$ sia rappresentabile nella forma (2.2.4) in cui i d_i , $i=1, \dots, k$, sono delle costanti. Questa forma è anche quella che si ottiene quando $\omega(x)$ è costruita con una tecnica di *interpolazione razionale*.

Teorema 2.4: sia

$$\omega(x) = \frac{p(x)}{q(x)} \quad (2.2.9)$$

una funzione razionale irriducibile, in cui $q(x)$ è un polinomio **monico** di grado $n \geq 1$. Allora $\omega(x)$ può essere espressa sotto forma di frazione continua finita di ordine $k \leq n$.

$$\omega(x) = s_0(x) + \frac{\alpha_1}{s_1(x)} + \frac{\alpha_2}{s_2(x)} + \dots + \frac{\alpha_k}{s_k(x)}$$

dove gli α_i , per $i=1, \dots, k$, sono delle costanti e gli $s_i(x)$, per $i=0, \dots, k$, sono dei polinomi. In particolare, se m è il grado di $p(x)$, allora $s_0(x)$ ha grado $m-n$ se $m \geq n$ ed è identicamente nullo se $m < n$, e gli $s_i(x)$, per $i=1, \dots, k$, sono polinomi monici di grado almeno 1.

Dim: se $m \geq n$, si ponga $t_0(x) = p(x)$ e $t_1(x) = q(x)$. Si applichi il seguente *algoritmo di Euclide* modificato

$$t_i(x) = t_{i+1}(x)s_i(x) + \alpha_{i+1}t_{i+2}(x), \quad \text{per } i = 0, \dots, k-1 \quad (2.2.10)$$

dove $s_i(x)$ è il quoziente e $\alpha_{i+1}t_{i+2}(x)$ è il resto della divisione di $t_i(x)$ per $t_{i+1}(x)$, e α_{i+1} è scelto in modo che $t_{i+2}(x)$ sia monico. L'algoritmo che calcola il massimo comun divisore fra $t_0(x)$ e $t_1(x)$, termina dopo $k \leq n$ passi, dove k è il minimo intero per cui $t_{k+1}(x)$ è un polinomio di grado 0. L'ipotesi di irriducibilità fatta su $\omega(x)$ assicura che $t_{k+1}(x)$ è una costante non nulla e quindi uguale a 1. Poiché il grado del polinomio $t_{i+2}(x)$ è minore del grado di $t_{i+1}(x)$, il quoziente $s_{i+1}(x)$ della divisione di $t_{i+1}(x)$ per $t_{i+2}(x)$ ha grado almeno 1, e poiché i primi coefficienti di $t_{i+1}(x)$ e $t_{i+2}(x)$ sono uguali a 1, il primo coefficiente di $s_{i+1}(x)$ è uguale a 1. Si pone infine $s_k(x) = t_k(x)$. Dividendo $t_i(x)$ per $t_{i+1}(x)$ e ponendo

$$z_i(x) = \frac{t_i(x)}{t_{i+1}(x)},$$

dalla (2.2.10) si ha

$$z_i(x) = s_i(x) + \frac{\alpha_{i+1}}{z_{i+1}(x)}$$

Confrontando con la (2.2.8), ne segue la (2.2.9), poiché

$$\omega(x) = \frac{t_0(x)}{t_1(x)}$$

Se $m < n$, si ponga $t_1(x) = q(x)$ e $\alpha_1 t_2(x) = p(x)$, dove α_1 è tale che $t_2(x)$ sia monico, e si proceda in modo analogo applicando la (2.2.10) a partire dall'indice $i=1$. L'unica differenza con il caso precedente è quindi l'assenza del primo termine, cioè $s_0(x) = 0$. Vanno considerati i seguenti casi molto importanti:

- 1) se $m=n$, allora $s_0(x)$ è una costante;
- 2) se $m=n + 1$, allora $s_0(x)$ è un polinomio di grado 1.

In entrambi i casi se il procedimento (2.2.10) determina come quozienti dei polinomi $s_i(x)$ di grado 1, allora $k=n$. Perciò la frazione continua è della forma

$$\omega(x) = \alpha_0 x + \beta_0 + \frac{\alpha_1(x)}{x + \beta_1} + \frac{\alpha_2(x)}{x + \beta_2} + \dots + \frac{\alpha_n(x)}{x + \beta_n}$$

dove $\alpha_0 = 0$ se $m=n$. Il valore di $\omega(x)$ si calcola allora con l'algoritmo

$$\begin{aligned} z_n &= x + b_n \\ z_i &= x + b_i + \frac{\alpha_{i+1}}{z_{i+1}} \quad \text{per } i = n - 1, \dots, 1, \\ \omega(x) &= \alpha_0 x + \beta_0 + \frac{\alpha_1}{z_1}. \end{aligned}$$

Ovviamente é possibile che il procedimento richieda un numero minore di passi e che risulti $k < n$. Il calcolo di $\omega(x)$, dati i coefficienti dei polinomi $p(x)$ e $q(x)$, é piú costoso del calcolo della funzione stessa, dati i coefficienti della frazione continua, perché il numero di operazioni moltiplicative richieste nel primo caso può essere anche doppio. Se $\omega(x)$ é una funzione pari, nella corrispondente frazione continua i polinomi $s_i(x)$ sono tutti pari e se $\omega(x)$ é funzione dispari, tutti gli $s_i(x)$ sono dispari. Per quanto riguarda l'accuratezza del risultato, però, il calcolo di una funzione razionale mediante una frazione continua può essere numericamente instabile.

Un'altra tecnica per trasformare una funzione razionale in frazione continua é il **metodo di Viskovatov** che dá una frazione continua della forma

$$\omega(x) = \frac{\alpha_1 x^{i_1}}{1} + \frac{\alpha_2 x^{i_2}}{1} + \dots + \frac{\alpha_k x^{i_k}}{1} \quad (2.2.11)$$

se nella (2.1.1) é $b_0 \neq 0$ (se fosse $b_0 = 0$, poiché $\omega(x)$ é irriducibile, dovrebbe essere $a_0 \neq 0$ e quindi il metodo potrebbe essere applicato alla funzione $1/\omega(x)$). Si supponga quindi che $b_0 = 1$; posto $t_1(x) = q(x)$, si indichi con $t_2(x)$ il polinomio tale che

$$p(x) = \alpha_1 x^{i_1} t_2(x),$$

dove $\alpha_1 x^{i_1}$ é il termine non nullo di grado minimo di $p(x)$ (se $a_0 \neq 0$ é $a_1 = a_0$ e $i_1 = 0$). Quindi il termine di grado minimo di $t_2(x)$ é uguale a 1 e si ha

$$\frac{p(x)}{q(x)} = \frac{\alpha_1 x^{i_1}}{\frac{t_1(x)}{t_2(x)}} = \frac{\alpha_1 x^{i_1}}{1 + \frac{t_1(x) - t_2(x)}{t_2(x)}}$$

il polinomio $t_1(x) - t_2(x)$ é divisibile per x e si può scrivere

$$t_1(x) - t_2(x) = \alpha_2 x^{i_2} t_3(x),$$

in cui $\alpha_2 x^{i_2}$ é il termine non nullo di grado minore di $t_1(x) - t_2(x)$. Procedendo in questo modo fino ad un indice k per cui $t_k(x) = t_{k+1}(x)$, si ottiene la (2.2.11). Se $i_1 = 0$ e $i_2 = \dots = i_k = 1$, si può verificare che $k = m + n + 1$, per cui la (2.2.11) risulta

$$\omega(x) = \frac{\alpha_1}{1} + \frac{\alpha_2 x}{1} + \dots + \frac{\alpha_k x}{1} \quad (2.2.12)$$

Frazioni continue della forma (2.2.12) vengono dette **S-frazioni**. Infine con piccole varianti del **metodo di Viskovatov** é possibile scrivere $\omega(x)$ in frazione continua in modo ancora diverso, ad esempio applicando il metodo alla funzione $\omega(x) - \omega(0)$, ottenendo

$$\omega(x) = \frac{\alpha_1 x^{i_1}}{1} + \frac{\alpha_2 x^{i_2}}{1} + \dots + \frac{\alpha_k x^{i_k}}{1}$$

in cui $i_1 \geq 1$, oppure nel caso $m \geq n$ si può prima calcolare il quoziente $s(x)$ e il resto $r(x)$ della divisione di $p(x)$ per $q(x)$, ottenendo

$$\omega(x) = s(x) + \frac{r(x)}{q(x)} = s(x) + \frac{\alpha_1 x^{i_1}}{1} + \frac{\alpha_2 x^{i_2}}{1} + \dots + \frac{\alpha_k x^{i_k}}{1}$$

In generale le diverse frazioni continue che corrispondono ad una stessa funzione razionale non sono equivalenti, dove per equivalenza si intende che le successioni delle frazioni parziali coincidono.

2.2.1 Differenze Inverse

I coefficienti della frazione continua di una funzione razionale $\omega(x)$ che interpola la funzione $f(x)$ sui nodi $x_i, i=0, \dots, m+n$, possono essere ottenuti con un procedimento analogo a quello usato per costruire il polinomio di interpolazione di Newton, usando al posto delle differenze divise le differenze inverse.

Definizione 2.3: si chiama differenza inversa di ordine k della funzione $f(x)$ relativa ai punti $x_i, i=0, \dots, k-1$, a due a due distinti, la funzione $\phi[x_0, x_1, \dots, x_{k-1}, x]$ definita ricorsivamente nel modo seguente

$$\begin{aligned} \text{per } k = 0 & \quad \phi[x] = f(x), \\ \text{per } k = 1 & \quad \phi[x_0, x] = \frac{x - x_0}{\phi[x] - \phi[x_0]}, \\ \text{per } k \geq 2 & \end{aligned}$$

$$\phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x] = \frac{x - x_{k-1}}{\phi[x_0, x_1, \dots, x_{k-2}, x] - \phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}]} \quad (2.2.13)$$

Oltre che nel punto x_{k-1} la funzione $\phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x]$ risulta non definita anche in ogni altro punto x tale che $\phi[x_0, x_1, \dots, x_{k-2}, x] = \phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}]$. Per il calcolo delle differenze inverse si usa una tabella simile a quella delle differenze divise:

$$\begin{array}{ccccccc}
 x_0 & \phi[x_0] & & & & & \\
 x_1 & \phi[x_1] & \phi[x_0, x_1] & & & & \\
 x_2 & \phi[x_2] & \phi[x_0, x_2] & \phi[x_0, x_1, x_2] & & & \\
 \vdots & \vdots & \vdots & \vdots & \dots\dots\dots & \phi[x_0, x_1, \dots, x_{m+n}] & \\
 \vdots & \vdots & \vdots & \vdots & & & \\
 x_{m+n} & \phi[x_{m+n}] & \phi[x_0, x_{m+n}] & \phi[x_0, x_1, x_{m+n}] & & &
 \end{array}$$

Gli elementi della tabella, considerati come elementi di una matrice A , triangolare inferiore di ordine $m+n+1$, vengono costruiti mediante le relazioni

$$\begin{aligned}
 a_{i1} &= f(x_{i-1}), \quad i = 1, \dots, m+n+1 \\
 a_{ij} &= \frac{x_{i-1} - x_{j-2}}{a_{i,j-1} - a_{j-1,j-1}}, \quad j = 2, \dots, m+n+1, \quad i = j, \dots, m+n+1.
 \end{aligned}$$

Poiché per ogni elemento sono richieste due sottrazioni e una divisione, la costruzione della tabella richiede, a meno di termini di ordine inferiore, $(m+n)^2$ addizioni a $(m+n)^2/2$ moltiplicazioni. Il seguente teorema mostra che la funzione $f(x)$ é esprimibile come frazione continua mediante le sue differenze inverse.

Teorema 2.5: se le differenze inverse $\phi[x_0, x_1, \dots, x_k]$ sono definite per $k=0, \dots, m+n$, allora vale la seguente relazione

$$f(x) = \phi[x_0] + \frac{x - x_0}{\phi[x_0, x_1]_+} \frac{x - x_1}{\phi[x_0, x_1, x_2]_+ \dots} \frac{x - x_{m+n-1}}{\phi[x_0, x_1, \dots, x_{m+n}]_+} \frac{x - x_{m+n}}{\phi[x_0, \dots, x_{m+n}, x]} \quad (2.2.14)$$

in ogni punto x per cui é definita $\phi[x_0, \dots, x_{m+n}, x]$.

Dim: dal fatto che $\phi[x_0, \dots, x_{m+n}, x]$ é definita, segue che in x sono definite anche tutte le differenze inverse di ordine inferiore. Si procede quindi per induzione. Per $m=n=0$, poiché $\phi[x_0, x] \neq 0$ per $x \neq x_0$, si ha

$$f(x) = \phi[x_0] + \frac{x - x_0}{\phi[x_0, x]} \quad (2.2.15)$$

come segue direttamente dalla definizione di differenze inversa per $k=1$. Per $m+n > 0$ si supponga che la (2.2.14) valga fino all'indice $m+n-1$, cioè

$$f(x) = \phi[x_0] + \frac{x-x_0}{\phi[x_0, x_1]_+} \frac{x-x_1}{\phi[x_0, x_1, x_2]_+ \dots} \frac{x-x_{m+n-1}}{\phi[x_0, \dots, x_{m+n-1}, x]} \quad (2.2.16)$$

in cui $\phi[x_0, \dots, x_{m+n-1}, x] \neq 0$ per $x \neq x_{m+n-1}$. Ma per la (2.2.13) é per $k=m+n+1$

$$\phi[x_0, \dots, x_{m+n-1}, x] = \phi[x_0, \dots, x_{m+n-1}, x_{m+n}] + \frac{x-x_{m+n}}{\phi[x_0, \dots, x_{m+n}, x]},$$

e sostituendo nella (2.2.16), ne segue subito la (2.2.14).

Teorema 2.6: nell'ipotesi del teorema 2.5, si considerino la frazione continua di ordine $m+n$

$$\omega(x) = \phi[x_0] + \frac{x-x_0}{\phi[x_0, x_1]_+} \frac{x-x_1}{\phi[x_0, x_1, x_2]_+ \dots} \frac{x-x_{m+n-1}}{\phi[x_0, \dots, x_{m+n}]}, \quad (2.2.17)$$

e i suoi residui per $i=0, \dots, m+n+1$

$$z_i(x) = \phi[x_0, x_1, \dots, x_i] + \frac{x-x_i}{\phi[x_0, x_1, \dots, x_{i+1}]_+ \dots} \frac{x-x_{m+n-1}}{\phi[x_0, x_1, \dots, x_{m+n}]}$$

Allora la (2.2.17) verifica la condizione $\omega(x_k) = f(x_k)$ per $k=m+n-1$, per $k=m+n$ e per quei $k=0, \dots, m+n-2$, per cui $z_{k+1}(x_k) \neq 0$. La (2.2.17), detta **frazione continua di Thiele**, é funzione razionale di interpolazione della $f(x)$ sui nodi x_i , $i=0, \dots, m+n$.

Dim: dalla (2.2.14) segue la (2.2.16) e quindi $f(x_{m+n}) = \omega(x_{m+n})$. Posto $z_{m+n}(x) = \phi[x_0, x_1, \dots, x_{m+n}]$, per $k=0, \dots, m+n-1$ risulta

$$\omega(x) = \phi[x_0] + \frac{x-x_0}{\phi[x_0, x_1]_+} \frac{x-x_1}{\phi[x_0, x_1, x_2]_+ \dots} \frac{x-x_{k-1}}{\phi[x_0, \dots, x_k]_+} \frac{x-x_k}{\phi[x_0, \dots, x_{k+1}, x]},$$

Per $k=m+n-1$ é $z_{m+n}(x) \neq 0$ perché $x_{m+n-1} \neq x_{m+n}$, e quindi $f(x_{m+n-1}) = \omega(x_{m+n-1})$.
 Per $k=0, \dots, m+n-2$, se $z_{k+1}(x_k) \neq 0$, risulta per la (2.2.14) che

$$\omega(x_k) = \phi[x_0] + \frac{x_k - x_0}{\phi[x_0, x_1]_{+\dots+}} \frac{x_k - x_{k-1}}{\phi[x_0, x_1, \dots, x_k]} = f(x_k).$$

Se per un indice k accade che

$$\phi[x_0, x_1, \dots, x_{k-2}, x_k] = \phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}] \quad (2.2.18)$$

allora la differenza inversa $\phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x_k]$ non può essere calcolata e quindi la funzione razionale di interpolazione non può essere rappresentata nella forma (2.2.18).
 Se $k < n$ é talvolta possibile, con un opportuno riordinamento dei punti, evitare che ciò accada. Dal teorema 2.6 segue che la i -esima frazione parziale della (2.2.17)

$$\omega(x) = \phi[x_0] + \frac{x - x_0}{\phi[x_0, x_1]_{+}} \frac{x - x_1}{\phi[x_0, x_1, x_2]_{+\dots+}} \frac{x - x_{i-1}}{\phi[x_0, \dots, x_i]},$$

é funzione razionale di interpolazione della $f(x)$ sui nodi x_0, \dots, x_{i-1} . Dalla (2.2.6) si ha che, indicata con

$$\omega_i(x) = \frac{p_i(x)}{q_i(x)} \quad (2.2.19)$$

la i -esima frazione parziale, i polinomi $p_i(x)$ e $q_i(x)$ soddisfano alle relazioni ricorrenti

$$\begin{aligned} p_{-1}(x) &= 1 & q_{-1}(x) &= 0 & p_0(x) &= f(x_0) & q_0(x) &= 1 \\ p_i(x) &= \phi[x_0, x_1, \dots, x_i]p_{i-1}(x) + (x - x_{i-1})p_{i-2}(x) \\ q_i(x) &= \phi[x_0, x_1, \dots, x_i]q_{i-1}(x) + (x - x_{i-1})q_{i-2}(x) \end{aligned} \quad (2.2.20)$$

Se non intervengono elisioni dei termini di grado massimo, dalle (2.2.20) segue che

$$\begin{aligned} \text{se } i = 2s \quad & \text{grado di } p_i(x) = \text{grado di } q_i(x) = s, \\ \text{se } i = 2s + 1 \quad & \text{grado di } p_i(x) = s + 1 \text{ e grado di } q_i(x) = s, \end{aligned} \quad (2.2.21)$$

per cui la funzione razionale di interpolazione che si ottiene per mezzo della frazione continua di Thiele é quella che ha lo stesso grado al numeratore ed al denominatore se n é pari, ed ha grado del numeratore di 1 superiore a quello del denominatore se n é dispari. Nel calcolo di $p_i(x)$ con i pari e di $q_i(x)$ con i dispari puó accadere che i termini di grado massimo si elidano, per cui il grado di $p_i(x)$ o di $q_i(x)$ sia minore di s . In generale però i gradi dei polinomi $p_j(x)$ e $q_j(x)$ con $j > i$ non risultano alterati.

2.2.2 Differenze Reciproche

L'analogia fra le differenze divise e quelle inverse si perde quando si passa a considerare la proprietá di simmetria espressa dal teorema 1.3: nel caso delle differenze inverse la simmetria riguarda solo gli ultimi due argomenti, ma in generale non gli altri. La proprietá di simmetria vale invece per le differenze reciproche della $f(x)$, cosí definite.

Definizione 2.4: si chiama **differenza reciproca** di ordine k della funzione $f(x)$ relativa ai punti x_i , $i=0,1,\dots,k-1$, a due a due distinti, la funzione $\rho[x_0, x_1, \dots, x_{k-1}, x]$ definita ricorsivamente nel modo seguente

$$\begin{aligned} \text{per } k = 0 \quad & \rho[x] = f(x), \\ \text{per } k = 1 \quad & \rho[x_0, x] = \frac{x-x_0}{\rho[x]-\rho[x_0]}, \\ \text{per } k \geq 2 \quad & \rho[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x] = \rho[x_0, x_1, \dots, x_{k-2}] + \\ & \frac{x-x_{k-1}}{\rho[x_0, x_1, \dots, x_{k-2}, x] - \rho[x_0, x_1, \dots, x_{k-2}, x_{k-1}]} \end{aligned} \quad (2.2.22)$$

Dal confronto con la definizione 2.3 risulta che $\rho[x] = \phi(x)$ e $\rho[x_0, x] = \phi[x_0, x]$. Il seguente teorema dá la relazione che lega fra loro le differenze inverse e le differenze reciproche di ordine piú elevato.

Teorema 2.7: per $k \geq 2$ vale la relazione

$$\phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x] = \rho[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x] - \rho[x_0, x_1, \dots, x_{k-2}]. \quad (2.2.23)$$

Dim: si procede per induzione su k: per k=2 si ha

$$\rho[x_0, x_1, x] = \rho[x_0] + \frac{x - x_1}{\rho[x_0, x] - \rho[x_0, x_1]} = \rho[x_0] + \frac{x - x_1}{\phi[x_0, x] - \phi[x_0, x_1]} = \rho[x_0] + \phi[x_0, x_1, x]. \quad (2.2.24)$$

Per $k > 2$ per l'ipotesi induttiva si ha

$$\rho[x_0, \dots, x_{k-3}, x_{k-2}, x] = \rho[x_0, \dots, x_{k-3}] + \phi[x_0, x_1, \dots, x_{k-2}, x], \quad (2.2.25)$$

per cui

$$\rho[x_0, \dots, x_{k-3}, x_{k-2}, x_{k-1}] = \rho[x_0, \dots, x_{k-3}] + \phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}]. \quad (2.2.26)$$

Sottraendo la (2.2.25) dalla (2.2.26) e utilizzando la definizione di *differenza reciproca* si ha

$$\rho[x_0, x_1, \dots, x_{k-2}, x_{k-1}, x] = \rho[x_0, x_1, \dots, x_{k-2}] + \frac{x - x_{k-1}}{\phi[x_0, x_1, \dots, x_{k-2}, x] - \phi[x_0, x_1, \dots, x_{k-2}, x_{k-1}]}, \quad (2.2.27)$$

da cui segue la (2.2.23) tenendo conto della definizione di *differenza inversa*.

Teorema 2.8: la differenza reciproca di ordine k

$$\rho[x_0, x_1, \dots, x_k]$$

é funzione simmetrica dei suoi argomenti x_0, x_1, \dots, x_k , cioè é invariante comunque essi vengano permutati.

Dim: per semplificare la dimostrazione si fa l'ipotesi che nei nodi x_0, x_1, \dots, x_k esista la funzione razionale di interpolazione della $f(x)$ e che nella costruzione dei polinomi $p_i(x)$ e $q_i(x)$ della (2.2.19) non vi siano cancellazioni dei termini di grado massimo. Indicati con a_i e b_i i coefficienti dei termini di grado massimo dei polinomi $p_i(x)$ e $q_i(x)$, si dimostra per induzione che, se non si moltiplicano numeratore e denominatore per una costante, allora

$$\begin{aligned} \text{per } i = 2s \quad & e' \quad a_i = \rho[x_0, x_1, \dots, x_i] \text{ e } b_i = 1, \\ \text{per } i = 2s + 1 \quad & e' \quad a_i = 1 \text{ e } b_i = \rho[x_0, x_1, \dots, x_i], \end{aligned} \quad (2.2.28)$$

Infatti per $i=0$ e $i=1$ queste relazioni sono immediate. Per $i = 2s > 1$ si ha dalla (2.2.21) che i polinomi $p_{i-1}(x)$ e $(x - x_{i-1})p_{i-2}(x)$ hanno lo stesso grado minore o uguale a s , e quindi per la (2.2.20) e per l'ipotesi induttiva il primo coefficiente di $p_i(x)$ é dato da

$$a_i = \phi[x_0, x_1, \dots, x_i]a_{i-1} + a_{i-2} = \phi[x_0, x_1, \dots, x_i] + \rho[x_0, x_1, \dots, x_{i-2}]$$

da cui per la (2.2.23) si ottiene la prima delle (2.2.27). Le altre relazioni si verificano nello stesso modo. Permutando in qualsiasi modo i punti x_0, x_1, \dots, x_i , la i -esima frazione parziale $\omega_i(x)$ in (2.2.19), che é la funzione razionale di interpolazione di $f(x)$ nei punti x_0, x_1, \dots, x_i , non varia. Segue che anche i primi coefficienti dei due polinomi $p_i(x)$ e $q_i(x)$ non variano permutando i punti x_0, x_1, \dots, x_i e quindi neanche le differenze reciproche $\rho[x_0, x_1, \dots, x_i]$. Le differenze reciproche si possono sistemare nella tabella seguente:

$$\begin{array}{ccccccc} x_0 & \rho[x_0] & & & & & \\ x_1 & \rho[x_1] & \rho[x_0, x_1] & & & & \\ x_2 & \rho[x_2] & \rho[x_1, x_2] & \rho[x_0, x_1, x_2] & & & \\ x_3 & \rho[x_3] & \rho[x_2, x_3] & \rho[x_1, x_2, x_3] & \rho[x_0, x_1, x_2, x_3] & & \\ \vdots & \vdots & \vdots & \vdots & & & \end{array}$$

La simmetria delle differenze reciproche consentirebbe di rappresentare la funzione razionale di interpolazione non tenendo conto dell'ordinamento dei punti nella costruzione della tabella delle differenze, purché nella (2.2.17) si esprimano i coefficienti mediante le differenze reciproche con la (2.2.23). In pratica, però, le differenze reciproche, dato che sono

piú complesse da costruire, vengono usate solo se alcuni o tutti i punti x_i , $i=0, \dots, m+n$, coincidono. In quel caso la proprietá di simmetria delle differenze reciproche consente di rappresentare ugualmente la funzione razionale di interpolazione, in modo analogo a quanto fatto per le differenze divise. Per semplicitá consideriamo il caso in cui tutti i punti x_i sono coincidenti col punto x , ma si otterrebbero sviluppi analoghi nel caso in cui solo alcuni dei punti x_i coincidessero. A differenza, però, del caso polinomiale, anche se $f(x) \in C^k[a, b]$, non é detto in generale che la funzione

$$\rho \underbrace{[x, x, \dots, x]}_{k+1 \text{ volte}} = \lim_{\substack{x_i \rightarrow x \\ i=0, \dots, k}} \rho[x_0, x_1, \dots, x_k]$$

sia definita per tutti i punti dell'intervallo $[a, b]$: infatti la funzione puó non esistere nei punti in cui si annulla almeno un denominatore. Negli sviluppi che seguono per ogni k si indicherá con Ω_k il massimo sottoinsieme aperto, che si suppone esista non vuoto, di $[a, b]^{k+1}$ in cui la funzione $\rho[x_0, x_1, \dots, x_k]$ é estendibile per continuitá ed ha derivate prime continue rispetto a tutti gli argomenti. Se $\underbrace{(x, x, \dots, x)}_{k+1 \text{ volte}} \in \Omega_k$, per semplicitá si userá la notazione

$$\rho^{(0)}[x] = \rho[x] \text{ e } \rho^{(k)}[x] = \rho \underbrace{[x, x, \dots, x]}_{k+1 \text{ volte}}$$

Teorema 2.9: sia $f(x) \in C^k[a, b]$, per un intero $k \geq 1$.

a) Se per $k=1$ é $f'(x) \neq 0$, allora $[x, x] \in \Omega_1$ e vale $\rho^{(1)}[x] = 1/f'(x)$.

b) Se per $k > 1$ é $\underbrace{(x, \dots, x)}_{k-1 \text{ volte}} \in \Omega_{k-2}$, $\underbrace{(x, \dots, x)}_k \in \Omega_{k-1}$, e $\frac{d}{dx}(\rho^{(k-1)}[x]) \neq 0$, allora

$$\underbrace{(x, \dots, x)}_{k+1 \text{ volte}} \in C^k$$

e vale

$$\rho^{(k)}[x] = \rho^{(k-2)}[x] + \frac{k}{\frac{d}{dx}(\rho^{(k-1)}[x])}. \quad (2.2.29)$$

Dim: a) per la (2.2.22) per $|\epsilon| \neq 0$ e sufficientemente piccolo, é $[x, x + \epsilon] \in \Omega_1$ e

$$\rho[x, x + \epsilon] = \frac{\epsilon}{f(x + \epsilon) - f(x)}.$$

Per ipotesi il limite per $\epsilon \rightarrow 0$ del secondo membro esiste ed é

$$\rho^{(1)}[x] = \lim_{\epsilon \rightarrow 0} \rho[x, x + \epsilon] = \lim_{\epsilon \rightarrow 0} \frac{\epsilon}{f(x + \epsilon) - f(x)} = \frac{1}{f'(x)}$$

b) per $|\epsilon| \neq 0$ e sufficientemente piccolo é

$$\underbrace{(x, \dots, x, x + \epsilon)}_{k-1 \text{ volte}} \in \Omega_{k-1}$$

e

$$\lim_{\epsilon \rightarrow 0} \frac{\rho[\underbrace{x, \dots, x, x + \epsilon}_{k-1 \text{ volte}}] - \rho[\underbrace{x, \dots, x}_{k \text{ volte}}]}{\epsilon} = \left. \frac{\partial \rho[x_0, \dots, x_{k-1}]}{\partial x_{k-1}} \right|_{x_0, \dots, x_{k-1} = x} \quad (2.2.30)$$

Poiché $\rho[x_0, \dots, x_{k-1}]$ é funzione simmetrica dei suoi argomenti, vale

$$\rho[\underbrace{x, \dots, x}_{k \text{ volte}}] = \sum_{j=0}^{k-1} \left. \frac{\partial \rho[x_0, \dots, x_{k-1}]}{\partial x_j} \frac{dx_j}{dx} \right|_{x_0, \dots, x_{k-1} = x} = k \left. \frac{\partial \rho[x_0, \dots, x_{k-1}]}{\partial x_{k-1}} \right|_{x_0, \dots, x_{k-1} = x}$$

in quanto $\frac{dx_j}{dx} = 1$ perché $x_j(x) = x$. Sostituendo nella (2.2.29) si ha per la (2.2.22)

$$\rho[\underbrace{x, \dots, x}_{k+1 \text{ volte}}] = \rho[\underbrace{x, \dots, x}_{k-1 \text{ volte}}] + \lim_{\epsilon \rightarrow 0} \frac{\epsilon}{\rho[\underbrace{x, \dots, x, x + \epsilon}_{k-1 \text{ volte}}] - \rho[\underbrace{x, \dots, x}_{k \text{ volte}}]} = \rho[\underbrace{x, \dots, x}_{k-1 \text{ volte}}] + \frac{k}{\frac{d}{dx} \rho[\underbrace{x, \dots, x}_{k \text{ volte}}]}$$

da cui si ottiene la (2.2.29). La (2.2.29) consente quindi, a partire da

$$\rho^{(0)}[x] = f(x) \text{ e } \rho^{(1)}[x] = 1/f'(x),$$

di costruire le differenze reciproche di ordine piú elevato. Inoltre, per la (2.2.23), se $\rho^{(k-2)}[x]$ e $\rho^{(k)}[x]$ esistono, allora esiste anche

$$\phi^{(k)}[x] = \phi[\underbrace{x, \dots, x}_{k+1 \text{ volte}}] = \lim_{\substack{x_i \rightarrow x \\ i=0, \dots, k}} \phi[x_0, \dots, x_k]$$

e vale

$$\phi^{(k)}[x] = \rho^{(k)}[x] - \rho^{(k-2)}[x] = \frac{k}{\frac{d}{dx} \rho^{(k-1)}[x]}. \quad (2.2.31)$$

É cosí possibile scrivere la frazione continua (2.2.17) nel caso in cui tutti i punti x_i coincidono con il punto x_0 .

Teorema 2.10: sia $x_0 \in [a, b]$. Se le differenze inverse $\phi^{(k)}[x_0]$ esistono finite per $k=0, \dots, m+n$, allora la frazione continua di Thiele di ordine $m+n$

$$\omega[x] = f(x_0) + \frac{x - x_0}{\phi^{(1)}[x_0]} + \frac{x - x_0}{\phi^{(m+n)}[x_0]} \quad (2.2.32)$$

é la funzione razionale che nel punto x_0 assume il valore $f(x_0)$ e le cui derivate fino all'ordine $m+n$ assumono i valori delle corrispondenti derivate della $f(x)$ in x_0 . Si osservi che l'ipotesi di non annullamento dei residui fatta nel teorema 2.6 é qui soddisfatta in quanto $\phi^{(k)}[x_0] \neq 0$ per ogni k .

Capitolo 3

Funzioni Spline e B-Spline

3.1 Polinomi di Interpolazione a Tratti di Primo Grado

Indicando con

$$P_n = \left\{ p(x) : p(x) = \sum_{i=1}^n c_i x^{i-1}; \quad c_i, x \in \mathfrak{R} \right\},$$

lo spazio dei polinomi di ordine n (grado $n-1$). Esso ha dimensione finita e base $\{1, x, x^2, \dots, x^{n-1}\}$. I polinomi sono funzioni regolari (C^∞), facili da memorizzare e da valutare. Le derivate sono ancora polinomi e i coefficienti sono ottenibili algebricamente, gli zeri sono meno di $n-1$, le matrici per i problemi di interpolazione o di approssimazione sono sempre *non singolari*, ma per il *fenomeno di Runge* sono *inflessibili*. Per questa inflessibilit  si suggerisce di usare polinomi di grado *limitatamente basso* e di *suddividere* l'intervallo in pezzi piccoli.

Definizione 3.1: siano assegnati i punti di osservazione in ordine crescente:

$$a = x_0 < x_1 < \dots < x_k < x_{k+1} < \dots < x_n = b$$

e siano assegnate le osservazioni $y_0, y_1, \dots, y_k, y_{k+1}, \dots, y_n$. In particolare, si pu  supporre che l'osservazione y_i sia il valore in x_i di una funzione $f(x)$ definita nell'intervallo $[a, b]$,

$i=0,1,\dots,n$. Possiamo ricostruire una funzione empirica (e, in particolare, costruire una ‘approssimazione’ della funzione $f(x)$ in $[a,b]$) usando il seguente criterio. Anziché considerare un polinomio $p_n(x)$ di grado n che soddisfa la condizione $p_n(x_i) = y_i$ per $i=0,1,\dots,n$, si considerano n polinomi di grado prefissato $\mu \ll n$ (ad esempio, $\mu = 1$ oppure $\mu = 3$) $p_1(x), p_2(x), \dots, p_k(x), \dots, p_n(x)$, ove $p_k(x)$ é definito nell’intervallo $[x_{k-1}, x_k]$ e soddisfa le condizioni ($k=1,2,\dots,n$):

$$p_k(x_{k-1}) = y_{k-1}, \quad p_k(x_k) = y_k$$

Se si considera $\mu = 1$, dalla (1.2.5) e (1.2.6) si ottiene

$$p_k(x) = y_{k-1} \frac{x-x_k}{x_{k-1}-x_k} + y_k \frac{x-x_{k-1}}{x_k-x_{k-1}} \quad x_{k-1} \leq x \leq x_k$$

Tale funzione é continua in $[a,b]$ e si chiama **polinomio di interpolazione a tratti di primo grado**. Ponendo

$$B_0(x) = \begin{cases} \frac{x-x_1}{x_0-x_1} & \text{per } x_0 \leq x \leq x_1 \\ 0 & \text{per } x_1 \leq x \leq x_n \end{cases}$$

$$B_k(x) = \begin{cases} \frac{x-x_{k-1}}{x_k-x_{k-1}} & \text{per } x_{k-1} \leq x \leq x_k \\ \frac{x-x_{k+1}}{x_k-x_{k+1}} & \text{per } x_k \leq x \leq x_{k+1} \\ 0 & \text{per } x_0 \leq x \leq x_{k-1} \text{ e } x_{k+1} \leq x \leq x_n \end{cases}$$

$$k=1,2,\dots,n-1$$

$$B_n(x) = \begin{cases} \frac{x-x_{n-1}}{x_n-x_{n-1}} & \text{per } x_{n-1} \leq x \leq x_n \\ 0 & \text{per } x_0 \leq x \leq x_{n-1} \end{cases}$$

il polinomio a tratti di primo grado che si indica con $p(x)$, si può scrivere nella forma:

$$p(x) = y_0 B_0(x) + y_1 B_1(x) + \dots + y_k B_k(x) + \dots + y_n B_n(x) \quad a \leq x \leq b$$

Sono stati studiati anche polinomi di interpolazione a tratti di terzo grado, caratterizzati dalla proprietà che il polinomio di terzo grado $p_k(x)$ definito nell'intervallo $[x_{k-1}, x_k]$ e il polinomio di terzo grado $p_{k+1}(x)$ definito nell'intervallo $[x_k, x_{k+1}]$ assumono in x_k il medesimo valore y_k e posseggono in x_k un 'contatto del secondo ordine', $k=1,2,\dots,n-1$. Allora l'utilizzo dei **polinomi di interpolazione a tratti di grado basso su piccoli intervalli** può risolvere il problema dell'aumento dell'errore all'aumentare di n . Su questo si basa l'interpolazione polinomiale a tratti di cui le **funzioni spline** sono un'applicazione e un'estensione.

3.2 L'Errore nell'Interpolazione a Tratti

Nella sezione 1.3 pagg. 16 e 17 si è visto come, in generale, per distribuzioni equispaziate dei nodi di interpolazione, non si abbia convergenza uniforme di p_n a f per $n \rightarrow \infty$. D'altra parte, da un lato l'equispaziatura dei nodi presenta considerevoli vantaggi computazionali, dall'altro l'interpolazione di Lagrange risulta ragionevolmente accurata per *gradi bassi*, a patto di interpolare su intervalli sufficientemente *piccoli*. Risulta, quindi, naturale introdurre una partizione \mathcal{T}_h di $[a,b]$ in K sottointervalli $I_j = [x_j, x_{j+1}]$ di lunghezza h_j , con $h = \max_{0 \leq j \leq K-1} h_j$, tali che $[a, b] = \cup_{j=0}^{K-1} I_j$ ed usare poi l'interpolazione di Lagrange su ciascuna intervallo I_j con n piccolo e su $n+1$ nodi equispaziati $\{x_j^{(i)}, 0 \leq i \leq n\}$. Per $k \geq 1$, introduciamo su \mathcal{T}_h lo spazio dei *polinomi a tratti*

$$X_h^k = \{v \in C^0([a, b]) : v|_{I_j} \in P_k(I_j) \forall I_j \in \mathcal{T}_h\}$$

definito come lo spazio delle funzioni continue su $[a,b]$ le cui restrizioni a ciascun I_j sono polinomi di grado $\leq k$. Allora, per ogni funzione f continua su $[a,b]$, il *polinomio interpolatore a tratti*, p_h^k , coincide su I_j con il polinomio interpolatore di $f|_{I_j}$ sui $k+1$ nodi $\{x_j^{(i)}, 0 \leq i \leq k\}$. Di conseguenza, se $f \in C^{k+1}([a, b])$, per la (1.3.9) applicata su ogni intervallo, si ricava la seguente stima dell'errore

$$\|f - p_h^k\|_\infty \leq Ch^{k+1} \|f^{(k+1)}\|_\infty.$$

Si noti che si può ottenere un piccolo errore di interpolazione anche per k piccolo purché h sia sufficientemente 'piccolo'.

3.3 Splines

Indichiamo con

$$\mathcal{P}P_n(\{x_0, x_1, \dots, x_k, x_{k+1}\}) = \{p : \exists p_0, \dots, p_k \in P_n \mid p_i(x) = p(x), x \in I_i = [x_i, x_{i+1}], i = 0, 1, \dots, k\},$$

per $n > 0$, lo **spazio dei polinomi continui a tratti** di ordine n con nodi $x_0, x_1, \dots, x_k, x_{k+1}$. Allora un elemento di $\mathcal{P}P_n(\{x_0, x_1, \dots, x_k, x_{k+1}\})$ consiste di $k+1$ pezzi di grado $\leq n-1$ e i suoi elementi possono essere discontinui.

Definizione 3.2: sia $\Delta = \{x_0, x_1, \dots, x_k, x_{k+1}\}$ una partizione di $[a, b]$ ed $n \in \mathcal{N}$. Lo spazio delle funzioni **spline polinomiali** é

$$S_n(\Delta) = \mathcal{P}P_n(\Delta) \cap C^{n-2}[a, b]$$

Ovvero polinomi a tratti di ordine n con nodi *semplici* nei punti interni x_1, \dots, x_k . Generalizziamo la definizione 3.2 al caso di *nodii interni multipli*. Allora indichiamo con $\mathcal{M} = (m_1, \dots, m_k)$, tale che $1 \leq m_i \leq n$, $i=1, \dots, k$, il *vettore delle molteplicitá* corrispondente ai nodii interni della partizione Δ .

Definizione 3.3:

$$S(P_n; \mathcal{M}; \Delta) = \{s : \exists s_0, \dots, s_k \in P_n \mid s(x) = s_i(x), x \in I_i, i = 0, 1, \dots, k$$

$$D^j s_{l-1}(x_l) = D^j s_l(x_l), j = 0, 1, \dots, n-1-m_l, l = 1, \dots, k\}$$

é lo spazio delle **spline polinomiali di ordine n con nodii (interni) x_1, \dots, x_k di molteplicitá m_1, \dots, m_k rispettivamente**.

Il vettore \mathcal{M} controlla la natura dello spazio $S(P_n; \mathcal{M}; \Delta)$, controllando la regolaritá della spline nei nodii. Se $m_i = n$, $\forall i$, allora s_{i-1} e s_i non sono collegate in x_i , invece se $m_i < n$, allora s_{i-1} e s_i sono obbligate a passare per il punto x_i . Ma non solo, la generica spline e le sue derivate $n-1-m_i$ sono continue sui nodii interni.

Teorema 3.1: sia $K = \sum_{i=1}^k m_i$, allora $S(P_n; \mathcal{M}; \Delta)$ é uno spazio lineare di dimensione $n+K$.

Nel caso di nodii semplici (cioé $\mathcal{M} = (1, 1, \dots, 1)$) allora $\dim S_n(\Delta) = n + K$. Osservando che $P_n \subset S(P_n; \mathcal{M}; \Delta)$, una base per S deve includere almeno una base per P_n . A tal fine consideriamo le funzioni $1, x-a, \dots, (x-a)^{n-1}$. Dobbiamo trovare altri K elementi linearmente indipendenti. Inoltre consideriamo la funzione ‘*potenza troncata*’ (vista come funzione di x):

$$t(x; y) = (x - y)_+^j = (x - y)^j (x - y)_+^0,$$

ove

$$(x - y)_+^0 = \begin{cases} 0 & x < y \\ 1 & x \geq y \end{cases}$$

Teorema 3.2: le funzioni

$$\sigma_{ij}(x) = (x - x_i)_+^{n-j}, \quad j = 1, \dots, m_i, \quad i = 0, 1, \dots, k \quad (3.3.1)$$

con $x_0 = a$ e $m_0 = n$, formano una base per $S(P_n; \mathcal{M}; \Delta)$ che include la base di P_n .

Per le dimostrazioni dei teoremi appena enunciati si veda [6].

Osservazione: la base (3.3.1) non é opportuna per le applicazioni numeriche. Infatti per valutare $s(x)$ per x vicina a b é necessario valutare tutte le funzioni di base e farne la somma. Per questa ragione sono state definite le **B-Spline** che invece sono a base simmetrica e supporto locale.

3.4 B-Splines

Si definiscono le **B-splines** di ordine k come k -esime differenze divise di potenze troncate e si prova che ogni spazio P_k , di funzioni polinomiali a tratti continui e derivabili, ha una base che consiste di certe splines di base o **B-splines**.

Definizione 3.4: sia $x = \{x_i\}$ una sequenza non decrescente di punti (finita o infinita). La **i -esima B-spline normalizzata di ordine k per la sequenza di nodi x** si indica con $B(x; x_i, \dots, x_{i+k})$ ed é definita come la k -esima differenza divisa di ordine k della funzione

$$p(t, x) = (t - x)_+^{k-1} = \begin{cases} (t - x)^{k-1} & x < t \\ 0 & \text{altrimenti} \end{cases}$$

ovvero

$$B(x; x_i, \dots, x_{i+k}) = (x_{i+k} - x_i)p[x_i, \dots, x_{i+k}](x) \quad (3.4.2)$$

dove $p[\cdot](x)$ é la k -esima differenza divisa sui punti x_i, \dots, x_{i+k} della funzione $p(\cdot, x)$ vista come funzione di x (a t viene sostituito di volta in volta uno dei nodi x_i).

Se la partizione $\{x_i\}$ é fissata, per indicare la i -esima B-spline di ordine k la notazione piú usata é $B_{i,k}(x)$ invece di $B(x; x_i, \dots, x_{i+k})$.

Osservazioni: la definizione 3.4 é buona, infatti se $x = x_j$ per qualche j , allora si definisce

$$B_{i,k}(x_j) = \lim_{x \rightarrow x_j^+} B_{i,k}(x),$$

di modo che le B-splines sono *continue a destra*. Sapendo che

$$p[x_i, \dots, x_{i+k}](x) = \sum_{j=i}^{i+k} a_j (x_j - x)_+^{k-1},$$

con $a_j = \prod_{h=i, h \neq j}^{i+k} \frac{1}{x_h - x_j}$, la B-spline é quindi un polinomio continuo a tratti di grado $k-1$ e appartiene a $C^{k-2}(\mathfrak{R})$.

Proposizione 3.1:

1. $B_{i,k}(x) = 0$ se $x \notin (x_i, x_{i+k})$.
2. $B_{i,k} > 0$ nel suo supporto (x_i, x_{i+k}) .
3. $\forall x \in \mathfrak{R}, \sum_{i=-\infty}^{\infty} B_{i,k}(x) = 1$ o equivalentemente $\int_{\mathfrak{R}} B_{i,k}(x) dx = \int_{x_i}^{x_{i+k}} B_{i,k}(x) dx = 1$.

Osservazioni: una sequenza di funzioni non negative che soddisfa la proprietá (3) si dice una **partizione dell'unitá**; come si é giá detto $\{x^i, i = 0, \dots, k-1\}$ e $\{(x - x_j)_+^{k-1}, j = 1, \dots, n\}$ sono una base per lo spazio delle splines di ordine k , ma *mal condizionata* dal punto di vista numerico per la rappresentazione di splines arbitrarie. Al contrario le B-splines sono *ben condizionate* per questo si utilizzano per rappresentare le splines come una loro combinazione.

Sia $[a, b]$ tale che $x_0 < a < x_1$ e $x_n < b < x_{n+1}$. Definiamo

$$B_i(x) = B(x; x_{i-k}, \dots, x_i), \quad i = 1, \dots, n+k$$

Teorema 3.3: le B-splines $B_i(x)$, $i=1, \dots, n+k$ formano una base per lo spazio delle spline di grado $k-1$, $S(k-1; x_1, \dots, x_n)$ su $[a, b]$ (dove x_1, \dots, x_n sono i nodi interni ad $[a, b]$).

Lemma 3.1: una funzione $s \in S(k-1; x_1, \dots, x_n)$ é una **spline** se e solo se

$$s(x) = p_{k-1}(x) + \sum_{j=1}^n a_j (x - x_j)_+^{k-1}, \quad (3.4.3)$$

dove $p_{k-1}(x)$ é un polinomio di grado $\leq k-1$.

Una conseguenza del Lemma 3.1 é la definizione di **spline naturale**, ovvero di una spline di grado $2k-1$ che é data da un polinomio di grado $\leq k-1$ in $[a, x_i] \cup [x_n, b]$ e tale che

$$s^j(a) = s^j(b) = 0, \quad j = k, \dots, 2k-2$$

Corollario 3.1: una funzione $s : [a, b] \rightarrow \mathfrak{R}$ é una **spline naturale** di grado $2k-1$ con sequenza dei nodi x_1, \dots, x_n se e solo se

$$s(x) = p_{k-1}(x) + \sum_{j=1}^n a_j (x - x_j)_+^{2k-1},$$

dove $p_{k-1} \in P_{k-1}$ e

$$\sum_{j=1}^n a_j x_j^i = 0, \quad i = 0, \dots, k-1.$$

Lemma 3.2: sia s una spline in $S(k-1; x_1, \dots, x_n)$. Se $s(x) = 0$ per ogni $x \notin [x_1, x_k]$ allora $s(x) \equiv 0, \forall x \in \mathfrak{R}$.

Lemma 3.3: le B-splines $B_i(x)$, $i=1, \dots, n+k$ definite come nel teorema 3.3 sono linearmente indipendenti su \mathfrak{R} .

Per le dimostrazioni si veda [6] o anche [5] al cap. IX.

3.4.1 Formula di Ricorrenza per B-splines

Proposizione 3.2: siano f e g due funzioni sufficientemente differenziabili e i punti $x_1 \leq \dots \leq x_{n+1}$ siano dati. Allora

$$(f.g)[x_1, \dots, x_{n+1}] = \sum_{j=1}^{n+1} f[x_1, \dots, x_j] g[x_j, \dots, x_{n+1}] \quad (3.4.4)$$

Per la dimostrazione si veda [6].

Consideriamo ora le **B-splines non normalizzate** di ordine k sui nodi x_i, \dots, x_{i+k}

$$N_{i,k}(x) = \frac{B_{i,k}(x)}{x_{i+k} - x_i}$$

Poniamo

$$(t - x)_+^{k-1} = (f.g)(t; x)$$

dove $f(t; x) = t - x$ e $g(t; x) = (t - x)_+^{k-2}$. Ora

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{x_{i+1} - x - x_i + x}{x_{i+1} - x_i} = 1$$

e le differenze divise di ordine superiore al primo di f sono nulle. Perciò

$$\begin{aligned} N_{i,k} &= (f.g)[x_i, \dots, x_{i+k}] \\ &= \sum_{j=0}^k f[x_i, \dots, x_{i+j}] g[x_{i+j}, \dots, x_{i+k}] \\ &= f(x_i)g[x_i, \dots, x_{i+k}] + f[x_i, x_{i+1}]g[x_{i+1}, \dots, x_{i+k}] \\ &= f(x_i) \frac{g[x_{i+1}, \dots, x_{i+k}] - g[x_i, \dots, x_{i+k-1}]}{x_{i+k} - x_i} + g[x_{i+1}, \dots, x_{i+k}] \\ &= \left(\frac{x_i - x}{x_{i+k} - x_i} + 1 \right) g[x_{i+1}, \dots, x_{i+k}] - \left(\frac{x_i - x}{x_{i+k} - x_i} \right) g[x_i, \dots, x_{i+k-1}]. \end{aligned}$$

Concludendo la formula richiesta é

$$N_{i,k}(x) = \left(\frac{x_i - x}{x_{i+k} - x_i} + 1 \right) N_{i+1,k-1}(x) + \left(\frac{x - x_i}{x_{i+k} - x_i} \right) N_{i,k-1}(x)$$

che é una *combinazione convessa* di $N_{i,k-1}$ e $N_{i+1,k-1}$. Moltiplicandola per $x_{i+k} - x_i$ si ha la formula di ricorrenza per le B-splines normalizzate

$$B_{i,k}(x) = \left(\frac{x_{i+k} - x}{x_{i+k} - x_{i+1}} \right) B_{i+1,k-1}(x) + \left(\frac{x - x_i}{x_{i+k-1} - x_i} \right) B_{i,k-1}(x).$$

Capitolo 4

Curve e Superfici B-splines

4.1 Curve e Superfici

I due metodi piú comuni per rappresentare curve e superfici in geometria sono mediante **equazioni implicite** e **funzioni parametriche**. L'equazione implicita di una curva che giace sul piano xy ha la forma $\mathbf{f}(\mathbf{x},\mathbf{y})=\mathbf{0}$. Tale equazione descrive una relazione implicita tra le coordinate x e y dei punti che giacciono sulla curva. Nella forma parametrica, ognuna delle coordinate di un punto sulla curva é rappresentata separatamente come una funzione esplicita di un parametro indipendente $\mathbf{C}(\mathbf{u})=(\mathbf{x}(\mathbf{u}),\mathbf{y}(\mathbf{u}))$, $a \leq u \leq b$. Sebbene l'intervallo $[a,b]$ sia arbitrario, di solito viene normalizzato a $[0,1]$. Una superficie viene definita da una equazione implicita della forma $\mathbf{f}(\mathbf{x},\mathbf{y},\mathbf{z})=\mathbf{0}$, mentre una rappresentazione parametrica é data da $\mathbf{S}(\mathbf{u},\mathbf{v})=(\mathbf{x}(\mathbf{u},\mathbf{v}),\mathbf{y}(\mathbf{u},\mathbf{v}),\mathbf{z}(\mathbf{u},\mathbf{v}))$. Sono richiesti due parametri per definire una superficie, tenendo u fisso e variando v si generano le rette latitudinali, mentre tenendo fisso v e variando u si generano le rette longitudinali. Indicando con $S_u(u, v) = (x_u(u, v), y_u(u, v), z_u(u, v))$ e $S_v(u, v) = (x_v(u, v), y_v(u, v), z_v(u, v))$ rispettivamente le derivate parziali di $\mathbf{S}(\mathbf{u},\mathbf{v})$ in u e in v , si hanno le velocità lungo le rette latitudinali e longitudinali. In ogni punto della superficie dove il vettore prodotto $S_u \times S_v$ non si annulla, il vettore normale unitario, N , é dato da

$$N = \frac{S_u \times S_v}{|S_u \times S_v|}.$$

Chiaramente, permettendo alle funzioni delle coordinate $x(u),y(u)$ e $z(u)$ di essere arbitrarie, otteniamo una grande varietà di curve. La situazione ideale é di restringere il campo a funzioni che:

- siano in grado di rappresentare con precisione tutte le curve di cui abbiamo bisogno;

- siano facilmente, efficientemente e accuratamente trattate dal computer, in particolare:
 - il calcolo dei punti e delle derivate sulle curve sia efficiente;
 - la trattazione numerica sia relativamente insensibile agli errori di fluttuazione dei punti;
 - richiedano poca memoria di riserva;
- siano semplici e matematicamente ben comprensibili.

Una classe di funzioni largamente utilizzate é quella dei polinomi, anche se non tutte le curve e le superficie si possono rappresentare con precisione. Due metodi comuni di esprimere funzioni polinomiali sono le **curve con base di potenze** e le **curve di Bézier**. Una curva con base di potenze di grado n é data da

$$C(u) = (x(u), y(u), z(u)) = \sum_{i=0}^n a_i u^i, \quad 0 \leq u \leq 1. \quad (4.1.1)$$

Gli $a_i = (x_i, y_i, z_i)$ sono vettori, da cui

$$x(u) = \sum_{i=0}^n x_i u^i \quad y(u) = \sum_{i=0}^n y_i u^i \quad z(u) = \sum_{i=0}^n z_i u^i$$

. In forma matriciale si ha

$$C(u) = [a_0, a_1, \dots, a_n] \begin{bmatrix} 1 \\ u \\ \vdots \\ u^n \end{bmatrix} = [a_i]^T [u^i].$$

Differenziando l'equazione (4.1.1) si ottiene

$$a_i = \frac{C^{(i)}(u)|_{u=0}}{i!}$$

dove $C^{(i)}(u)|_{u=0}$ é l' i -esima derivata di $C(u)$ per $u=0$. Le $n+1$ funzioni $\{u^i\}$ si chiamano funzioni di base e gli $\{a_i\}$ i coefficienti della rappresentazione con base di potenze. Dato u_0 , il punto $C(u_0)$ su una curva con base di potenze si calcola molto piú efficientemente usando il **metodo di Horner**:

- per grado 1: $C(u_0) = a_1 u_0 + a_0$
- per grado 2: $C(u_0) = (a_2 u_0 + a_1) u_0 + a_0$
- \vdots
- per grado n: $C(u_0) = ((\dots (a_n u_0 + a_{n-1}) u_0 + a_{n-2}) u_0 + \dots + a_0$

ALGORITMO DI HORNER

Input : a, n, u_0

Output : C

$C = a(n);$

per $i = n - 1$ a 0

$C = C * u_0 + a(i);$

4.2 Curve di Bézier

Finché si utilizzano polinomi per le coordinate delle funzioni, le curve con basi di potenze e le curve di Bézier sono matematicamente equivalenti, ma le curve di Bézier sono migliori nelle applicazioni geometriche. Le curve con base di potenze hanno alcuni difetti: sono artificiose per il disegno di modelli interattivi, gli algoritmi per calcolarle sono piú algebrici che geometrici e numericamente hanno una forma poco pratica, l'errore si propaga se i coefficienti variano molto in grandezza (vedi [8] al cap.I). Il metodo di Bézier risolve questi difetti. Una **curva di Bézier di grado n** é definita da

$$C(u) = \sum_{i=0}^n B_{i,n}(u) P_i, \quad 0 \leq u \leq 1 \quad (4.2.2)$$

Le funzioni di base, $\{B_{i,n}(u)\}$, sono i classici **polinomi di Bernstein di grado n**, dati da

$$B_{i,n}(u) = \frac{n!}{i!(n-i)!} u^i (1-u)^{n-i} \quad (4.2.3)$$

I coefficienti geometrici di questa forma, $\{P_i\}$, sono chiamati **punti di controllo**. Si noti che deve essere $u \in [0, 1]$.

Per $n=1$ dall'equazione (4.2.3) si ha $B_{0,1} = 1 - u$ e $B_{1,1}(u) = u$, e l'equazione (4.2.2) prende la forma $C(u) = (1 - u)P_0 + uP_1$. Questo é un segmento lineare da P_0 a P_1 .

Per $n=2$ dalle equazioni (4.2.2) e (4.2.3) si ha $C(u) = (1 - u)^2P_0 + 2u(1 - u)P_1 + u^2P_2$. Questo é un arco parabolico da P_0 a P_2 . Si noti che:

- il poligono formato da $\{P_0, P_1, P_2\}$, chiamato **poligono di controllo**, approssima la forma della curva piuttosto bene;
- $P_0 = C(0)$ e $P_2 = C(1)$;
- le direzioni delle tangenti alla curva agli estremi sono parallele a $P_1 - P_0$ e $P_2 - P_1$;
- la curva é contenuta nel triangolo formato da $P_0P_1P_2$.

Per $n=3$ si ha $C(u) = (1 - u)^3P_0 + 3u(1 - u)^2P_1 + 3u^2(1 - u)P_2 + u^3P_3$. Si noti che:

- i poligoni di controllo approssimano le forme delle curve ;
- $P_0 = C(0)$ e $P_3 = C(1)$;
- **proprietá della copertura convessa**: le curve sono contenute nella copertura convessa dei loro punti di controllo di definizione;
- **proprietá di ‘variation diminishing’**: nessun segmento lineare interseca una curva piú volte di quanto esso intersechi il poligono di controllo della curva stessa (per le curve tridimensionali al posto di segmento si parla di piano);
- all'inizio ($u=0$) la curva gira nella stessa direzione di $P_0P_1P_2$. Per $u=1$ gira nella direzione di $P_1P_2P_3$;
- un ‘cappio’ nel poligono di controllo puó o meno implicare un ‘cappio’ nella curva.

In aggiunta alle proprietá giá citate, le curve di Bézier sono invarianti rispetto alle usuali trasformazioni come rotazione, traslazione, e scala, cioè si applica la trasformazione alla curva applicandola al poligono di controllo. In ogni schema di rappresentazione di una curva o di una superficie, la scelta delle funzioni di base determina le caratteristiche dello schema. Le funzioni di base $\{B_{i,n}(u)\}$ hanno alcune proprietá:

1. **non negativitá**: $\{B_{i,n}(u)\} \geq 0$ per ogni i, n e $0 \leq u \leq 1$;
2. **partizione dell'unitá**: $\sum_{i=0}^n B_{i,n}(u) = 1$ per ogni $0 \leq u \leq 1$;

3. $B_{0,n}(0) = B_{n,n}(1) = 1$;
4. $B_{i,n}(u)$ raggiungono esattamente un massimo sull'intervallo $[0,1]$, cioè per $u = i/n$;
5. **simmetria**: per ogni n , l'insieme dei polinomi $\{B_{i,n}(u)\}$ é simmetrico rispetto a $u = 1/2$;
6. **definizione ricorsiva**: $B_{i,n}(u) = (1 - u)B_{i,n-1}(u) + uB_{i-1,n-1}(u)$; definiamo $B_{i,n}(u) \equiv 0$ se $i < 0$ o $i > n$;
7. **derivate**:

$$B'_{i,n}(u) = \frac{dB_{i,n}(u)}{du} = n(B_{i-1,n-1}(u) - B_{i,n-1}(u))$$

con

$$B_{-1,n-1}(u) \equiv B_{n,n-1}(u) \equiv 0;$$

Dalla proprietà 6 si ricava un semplice algoritmo per il calcolo dei valori dei polinomi di Bernstein ad un valore fissato u .

ALGORITMO DEI POLINOMI DI BERNSTEIN

Input : i, n, u

Output : B

per $j = 0$ a n

$temp(j) = 0.0$;

$temp(n - i) = 1.0$;

$u_1 = 1.0 - u$;

per $k = 1$ a n

per $j = n$ a k passo -1

$temp(j) = u_1 * temp(j) + u * temp(j - 1)$;

$B = temp(n)$;

Esiste poi un algoritmo piú complesso per calcolare tutti i polinomi di Bernstein insieme (si veda [8] pag. ALG. A1.3). Utilizzando la proprietà 7 é facile ricavare l'espressione generale per la derivata della curva di Bézier:

$$\begin{aligned}
 C'(u) &= \frac{d\left(\sum_{i=0}^n B_{i,n}(u)P_i\right)}{du} = \sum_{i=0}^n B'_{i,n}(u)P_i \\
 &= \sum_{i=0}^n n(B_{i-1,n-1}(u) - B_{i,n-1}(u))P_i \\
 &= nB_{-1,n-1}(u)P_0 + n\sum_{i=0}^{n-1} B_{i,n-1}(u)P_{i+1} \\
 &\quad - n\sum_{i=0}^{n-1} B_{i,n-1}(u)P_i - nB_{n,n-1}(u)P_n = \\
 &= n\sum_{i=0}^{n-1} B_{i,n-1}(u)(P_{i+1} - P_i)
 \end{aligned} \tag{4.2.4}$$

Il penultimo passaggio si deve al fatto che $B_{-1,n-1}(u) \equiv B_{n,n-1}(u) \equiv 0$ (dal punto 7 qui sopra). Dalla (4.2.2) si ottengono facilmente le formule per le derivate agli estremi della curva di Bézier:

$$\begin{aligned} C'(0) &= n(P_1 - P_0) & C''(0) &= n(n-1)(P_0 - 2P_1 + P_2) \\ C'(1) &= n(P_n - P_{n-1}) & C''(1) &= n(n-1)(P_n - 2P_{n-1} + P_{n-2}) \end{aligned} \quad (4.2.5)$$

Si nota dalle equazioni (4.2.2) e (4.2.3) che:

- la derivata di una curva di Bézier di grado n è una curva di Bézier di grado $(n-1)$;
- le espressioni delle derivate agli estremi $u=0$ e $u=1$ sono simmetriche (dalla simmetria delle funzioni di base);
- la k -esima derivata agli estremi dipende (geometricamente parlando) solo dai $k+1$ punti di controllo agli estremi.

Sia $n=2$ e $C(u) = \sum_{i=0}^2 B_{i,2}(u)P_i$. Allora

$$\begin{aligned} C(u) &= (1-u)^2 P_0 + 2u(1-u)P_1 + u^2 P_2 \\ &= (1-u) \underbrace{((1-u)P_0 + uP_1)}_{\text{lineare}} + u \underbrace{((1-u)P_1 + uP_2)}_{\text{lineare}} \end{aligned}$$

Allora, $C(u)$ si ottiene come un'interpolazione lineare delle due curve di Bézier di primo grado, in particolare, ogni punto su $C(u)$ si ottiene da tre interpolazioni lineari. Fissando $u = u_0$ e ponendo $P_{1,0} = (1-u_0)P_0 + u_0P_1$, $P_{1,1} = (1-u_0)P_1 + u_0P_2$, e $P_{2,0} = (1-u_0)P_{1,0} + u_0P_{1,1}$, si ha che $C(u_0) = P_{2,0}$. Indicando in generale la curva di Bézier di grado n con $C_n(P_0, \dots, P_n)$, si ha

$$C_n(P_0, \dots, P_n) = (1-u)C_{n-1}(P_0, \dots, P_{n-1}) + uC_{n-1}(P_1, \dots, P_n) \quad (4.2.6)$$

Questo deriva dalla proprietà 6 e fissando $u = u_0$ e indicando P_i con $P_{0,i}$ l'equazione (4.2.6) da un'algoritmo ricorsivo per calcolare il punto $C(u_0) = P_{n,0}(u_0)$ su una curva di Bézier di grado n , cioè

$$P_{k,i}(u_0) = (1-u_0)P_{k-1,i}(u_0) + u_0P_{k-1,i+1}(u_0) \quad \text{per} \quad \begin{cases} k = 1, \dots, n \\ i = 0, \dots, n-k \end{cases} \quad (4.2.7)$$

L'equazione (4.2.7) viene detta **Algoritmo di De Casteljaou**.

ALGORITMO DI DE CASTELJAU

Input : P, n, u
Output : C (punto)
 per $i = 0$ a n
 $Q(i) = P(i)$;
 per $k = 1$ a n
 per $i = 0$ a $n - k$
 $Q(i) = (1.0 - u) * Q(i) + u * Q(i + 1)$;
 $C = Q(0)$;

4.3 Curve di Bézier Razionali

Le curve di Bézier razionali sono casi speciali di curve con B-splines razionali. Anche se i polinomi offrono molti vantaggi, esiste un numero importante di tipi di curve e superfici che non possono essere rappresentate con precisione utilizzando i polinomi, ad esempio cerchi, ellissi, iperboli, cilindri, cono, sfere, ecc. Si sa dalla matematica classica che tutte le coniche possono essere rappresentate mediante **funzioni razionali**, definite come frazioni di due polinomi del tipo:

$$x(u) = \frac{X(u)}{W(u)} \quad y(u) = \frac{Y(u)}{W(u)} \quad (4.3.8)$$

dove $X(u), Y(u)$ e $W(u)$ sono polinomi e ognuna delle funzioni delle coordinate ha lo stesso denominatore. Allora si definisce una **curva di Bézier razionale di grado n** mediante

$$C(u) = \frac{\sum_{i=0}^n B_{i,n}(u)w_i P_i}{\sum_{i=0}^n B_{i,n}(u)w_i} \quad 0 \leq u \leq 1 \quad (4.3.9)$$

I $P_i = (x_i, y_i, z_i)$ e i $B_{i,n}(u)$ sono gli stessi mentre i w_i sono scalari chiamati **pesi**. Allora $W(u) = \sum_{i=0}^n B_{i,n}(u)w_i$ é la funzione denominatore comune. Eccetto in casi specifici, si assume che $w_i > 0$ per ogni i . Questo assicura che $W(u) > 0$ per ogni $u \in [0, 1]$. Si scrive

$$C(u) = \sum_{i=0}^n R_{i,n}(u)P_i \quad 0 \leq u \leq 1 \quad (4.3.10)$$

dove

$$R_{i,n}(u) = \frac{B_{i,n}(u)w_i}{\sum_{j=0}^n B_{j,n}(u)w_j}.$$

Le $R_{i,n}(u)$ sono le funzioni di base razionali per questa curva. Le $R_{i,n}(u)$ hanno proprietá che possono essere facilmente derivate dall'equazione (4.3.9) e dalle proprietá corrispondenti di $B_{i,n}(u)$:

- 8 **non negativitá**: $R_{i,n}(u) \geq 0$ per ogni i, n e $0 \leq u \leq 1$;
- 9 **partizione dell'unitá**: $\sum_{i=0}^n R_{i,n}(u) = 1$ per ogni $0 \leq u \leq 1$;
- 10 $R_{0,n}(0) = R_{n,n}(1) = 1$;
- 11 $R_{i,n}(u)$ raggiunge esattamente un massimo sull'intervallo $[0, 1]$;
- 12 se $w_i = 1$ per ogni i , allora $R_{i,n}(u) = B_{i,n}(u)$ per ogni i , cioé i $B_{i,n}(u)$ sono casi speciali di $R_{i,n}(u)$; perché, se $w_i = 1$ per ogni i , $R_{i,n}(u) = B_{i,n}(u) / \sum_{j=0}^n B_{j,n}(u)$, ma, per la proprietá di partizione dell'unitá, $\sum_{j=0}^n B_{j,n}(u) = 1$, da cui la conclusione;

Queste producono le seguenti proprietá geometriche delle curve razionali di Bézier:

- 13 **proprietá della copertura convessa**: le curve sono contenute in coperture convesse dei loro punti di controllo di definizione (i P_i);
- 14 **invarianza per trasformazioni**: rotazioni, traslazioni e scale sono applicate alla curva applicandole ai punti di controllo;
- 15 **proprietá di 'variation diminishing'**: la stessa delle curve di Bézier polinomiali;
- 16 **interpolazione agli estremi**: $C(0) = P_0$ e $C(1) = P_n$;

- 17 la k-esima derivata per $u=0$ ($u=1$) dipende dal primo (l'ultimo) dei $k+1$ punti di controllo e pesi, in particolare, $C''(0)$ e $C''(1)$ sono paralleli a $P_1 - P_0$ e $P_n - P_{n-1}$, rispettivamente;
- 18 le curve di Bézier polinomiali sono casi speciali di curve di Bézier razionali.

Curve razionali con funzioni coordinate nella forma (4.3.8) (con denominatore comune) hanno un'elegante interpretazione geometrica che produce efficienti e compatte elaborazioni di riserve di dati. L'idea é di usare *coordinate omogenee* per rappresentare una curva razionale nello spazio n -dimensionale come una curva polinomiale nello spazio $(n+1)$ -dimensionale. Si comincia con un punto nello spazio Euclideo tridimensionale, $P = (x, y, z)$. Allora P viene scritto come $P^w = (wx, wy, wz, w) = (X, Y, Z, W)$ nello spazio quadridimensionale, $w \neq 0$. P si ottiene da P^w dividendo tutte le coordinate per la quarta coordinata, W , cioè proiettando P^w dall'origine sull'iperpiano $W=1$ (per il caso bidimensionale, $p=(x,y)$). Questa applicazione, indicata con H , é un'applicazione prospettica con centro nell'origine

$$P = H \{P^w\} = H \{(X, Y, Z, W)\} = \begin{cases} \left(\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W}\right) & \text{se } W \neq 0 \\ \text{direzione di } (X, Y, Z) & \text{se } W = 0 \end{cases} \quad (4.3.11)$$

Si noti che per arbitrari x, y, z, w_1, w_2 , dove $w_1 \neq w_2$

$$H \{P^{w_1}\} = H \{(w_1x, w_1y, w_1z, w_1)\} = (x, y, z) = H \{(w_2x, w_2y, w_2z, w_2)\} = H \{P^{w_2}\}.$$

Ora dato un insieme di punti di controllo, $\{P_i\}$, e pesi, $\{w_i\}$, si possono costruire i punti di controllo pesati, $P_i^w = (w_i x_i, w_i y_i, w_i z_i, w_i)$. Ora si definisce la curva di Bézier non razionale (polinomiale) nello spazio quadridimensionale

$$C^w(u) = \sum_{i=0}^n B_{i,n}(u) P_i^w \quad (4.3.12)$$

Ora, applicando l'applicazione prospettica, H , a $C^w(u)$ si ottiene la corrispondente curva di Bézier razionale dell'equazione (4.3.9), cioè copiando le funzioni delle coordinate dell'equazione (4.3.12), si ottiene

$$\begin{aligned} X(u) &= \sum_{i=0}^n B_{i,n}(u) w_i x_i & Y(u) &= \sum_{i=0}^n B_{i,n}(u) w_i y_i \\ Z(u) &= \sum_{i=0}^n B_{i,n}(u) w_i z_i & W(u) &= \sum_{i=0}^n B_{i,n}(u) w_i \end{aligned}$$

Ponendo la curva nel piano tridimensionale si ottiene

$$x(u) = \frac{X(u)}{W(u)} = \frac{\sum_{i=0}^n B_{i,n}(u)w_i x_i}{\sum_{i=0}^n B_{i,n}(u)w_i}$$

$$y(u) = \frac{Y(u)}{W(u)} = \frac{\sum_{i=0}^n B_{i,n}(u)w_i y_i}{\sum_{i=0}^n B_{i,n}(u)w_i}$$

$$z(u) = \frac{Z(u)}{W(u)} = \frac{\sum_{i=0}^n B_{i,n}(u)w_i z_i}{\sum_{i=0}^n B_{i,n}(u)w_i}$$

usando la notazione vettoriale

$$\begin{aligned} C(u) = (x(u), y(u), z(u)) &= \frac{\sum_{i=0}^n B_{i,n}(u)w_i(x_i, y_i, z_i)}{\sum_{i=0}^n B_{i,n}(u)w_i} \\ &= \frac{\sum_{i=0}^n B_{i,n}(u)w_i P_i}{\sum_{i=0}^n B_{i,n}(u)w_i} \end{aligned}$$

4.4 Superfici Prodotto di Tensori

La curva $C(u)$ é una funzione valutata su vettori di un parametro. É un'applicazione (deformazione) di un segmento rettilineo nello spazio Euclideo tridimensionale. Una superficie é una funzione valutata su vettori di due parametri, u e v , e rappresenta l'applicazione di una regione, \mathcal{R} , del piano uv nello spazio Euclideo tridimensionale. Quindi avrá la forma $S(u, v) = (x(u, v), y(u, v), z(u, v))$, $(u, v) \in R$. Il modo piú semplice per rappresentare superfici é lo **schema del prodotto di tensori**. É uno schema di curva bidirezionale che utilizza funzioni di base e coefficienti geometrici. Le funzioni di base sono funzioni bivariate di u e v , che si costruiscono come prodotto di funzioni di base univariate. I coefficienti geometrici sono sistemati (topologicamente) in una rete bidirezionale, $n \times m$. Allora una superficie prodotto di tensori ha la forma

$$S(u, v) = (x(u, v), y(u, v), z(u, v)) = \sum_{i=0}^n \sum_{j=0}^m f_i(u)g_j(v)b_{i,j} \quad (4.4.13)$$

dove

$$\begin{cases} b_{i,j} = (x_{i,j}, y_{i,j}, z_{i,j}) \\ 0 \leq u, v \leq 1 \end{cases}$$

Si noti che il dominio (u, v) di questa applicazione é un quadrato (in generale un rettangolo), e $S(u, v)$ ha una forma matriciale

$$S(u, v) = [f_i(u)]^T [b_{i,j}] [g_j(v)]$$

dove $[f_i(u)]^T$ é un $(1) \times (n+1)$ vettore riga, $[g_j(v)]$ é un $(m+1) \times (1)$ vettore colonna, e $[b_{i,j}]$ é una $(n+1) \times (m+1)$ matrice di punti tridimensionali. Differenziando la (4.4.13), rispetto a u e a v , si ottengono $S_u(u, v)$ e $S_v(u, v)$ rispettivamente e per (u_0, v_0) fissato $S_u(u_0, v_0) = C'_{v_0}(u_0)$ e $S_v(u_0, v_0) = C'_{u_0}(v_0)$. Il vettore normale $N = \frac{S_u \times S_v}{|S_u \times S_v|}$. Le superfici non razionali di Bézier si ottengono prendendo una rete bidirezionale di punti di controllo e prodotti di polinomi di Bernstein univariati

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u)B_{j,m}(v)P_{i,j} \quad 0 \leq u, v \leq 1 \quad (4.4.14)$$

Per $u = u_0$ fissato

$$\begin{aligned} C_{u_0}(v) &= S(u_0, v) = \sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u_0)B_{j,m}(v)P_{i,j} \\ &= \sum_{j=0}^m B_{j,m}(v) \left(\sum_{i=0}^n B_{i,n}(u_0)P_{i,j} \right) \\ &= \sum_{j=0}^m B_{j,m}(v)Q_j(u_0), \end{aligned} \quad (4.4.15)$$

dove

$$Q_j(u_0) = \sum_{i=0}^n B_{i,n}(u_0)P_{i,j}, \quad j = 0, \dots, m$$

é una curva di Bézier che giace sulla superficie. Analogamente,

$$C_{v_0}(v) = \sum_{i=0}^n B_{i,n}(u) Q_i(v_0)$$

é una **u isocurva** che giace sulla superficie. Come nel caso delle curve, le superfici di Bézier godono di alcune proprietà, in particolare:

- **non negatività:** $B_{i,n}(u)B_{j,m}(v) \geq 0$ per ogni i,j,u,v ;
- **partizione dell'unità:** $\sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u)B_{j,m}(v) = 1$ per ogni u e v ;
- $S(u,v)$ é contenuto nella copertura convessa dei suoi punti di controllo;
- invarianza per trasformazioni;
- la superficie interpola i quattro punti di controllo agli angoli;
- quando triangolata, la rete di controllo forma un'approssimazione piana poliedrica della superficie.

É interessante notare che non c'è la proprietà di 'variation diminishing' per la superfici di Bézier. L'algoritmo di De Casteljau é facilmente estendibile al calcolo dei punti sulla superficie di Bézier.

In riferimento all'equazione (4.4.15), sia (u_0, v_0) fissato. Per j_0 fissato, $Q_{j_0}(u_0) = \sum_{i=0}^n B_{i,n}(u_0)P_{i,j_0}$ é il punto ottenuto applicando l'algoritmo di De Casteljau alla j_0 -esima riga di punti di controllo, cioè $\{P_{i,j_0}\}$, $i=0, \dots, n$. Quindi, applicando l'algoritmo $(m+1)$ volte si ottiene $C_{u_0}(v)$ e applicandolo ancora una volta a $C_{u_0}(v)$ in $v = v_0$ si ottiene $C_{u_0}(v_0) = S(u_0, v_0)$. Questo procedimento richiede

$$\frac{n(n+1)(m+1)}{2} + \frac{m(m+1)}{2}$$

interpolazioni lineari. Per simmetria, possiamo calcolare $C_{v_0}(u)$ per primo ($n+1$ applicazioni di De Casteljau) e poi calcolare $C_{v_0}(u_0) = S(u_0, v_0)$. Questo richiede

$$\frac{m(m+1)(n+1)}{2} + \frac{n(n+1)}{2}$$

interpolazioni lineari. Allora, se $n > m$ si calcola $C_{v_0}(u)$ per primo, poi $C_{v_0}(u_0)$, altrimenti, si calcola $C_{u_0}(v)$ per primo, poi $C_{u_0}(v_0)$.

Si definisce una *superficie razionale di Bézier* come una proiezione prospettica di una superficie polinomiale di Bézier quadridimensionale

$$S^w(u, v) = \sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u)B_{j,m}(v)P_{i,j}^w \quad (4.4.16)$$

e

$$S(u, v) = H \{S^w(u, v)\} = \frac{\sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u) B_{j,m}(v) w_{i,j} P_{i,j}}{\sum_{i=0}^n \sum_{j=0}^m B_{i,n}(u) B_{j,m}(v) w_{i,j}} = \sum_{i=0}^n \sum_{j=0}^m R_{i,j}(u, v) P_{i,j} \quad (4.4.17)$$

dove

$$R_{i,j}(u, v) = \frac{B_{i,n}(u) B_{j,m}(v) w_{i,j}}{\sum_{r=0}^n \sum_{s=0}^m B_{r,n}(u) B_{s,m}(v) w_{r,s}} \quad (4.4.18)$$

Si noti che le $R_{i,j}(u, v)$ sono funzioni razionali, ma non sono prodotti di altre funzioni di base. Quindi, $S(u, v)$ non é una superficie prodotto di tensori, ma $S^w(u, v)$ lo é. Assumendo $w_{i,j} > 0$ per ogni i e j , le proprietá elencate precedentemente per le superfici di Bézier non razionali (e il prodotto di funzioni $B_{i,n}(u) B_{j,m}(v)$) si estendono naturalmente alle superfici razionali di Bézier. Inoltre, se $w_{i,j} = 1$ per ogni i e j , allora $R_{i,j}(u, v) = B_{i,n}(u) B_{j,m}(v)$, e la superficie corrispondente é non razionale.

4.5 Curve e Superfici Polinomiali o Razionali a Trattti

Le curve costituite da un solo segmento polinomiale o razionale sono spesso inadeguate. I loro difetti sono:

- si richiede un grado elevato per soddisfare un gran numero di restrizioni, per esempio, si richiede grado $(n-1)$ per far passare una curva polinomiale di Bézier attraverso n punti. Tuttavia, un grado elevato delle curve é inefficiente da trattare e numericamente instabile;
- un grado elevato é richiesto per adattarsi accuratamente a figure complesse;
- una curva (superficie) di segmento singolo non é adatta per il disegno interattivo di figure, benché le curve di Bézier possono essere modellate attraverso i loro punti di controllo (e pesi), il controllo non é sufficientemente locale (Vedi [8] cap. II).

Come si é già detto la soluzione é di utilizzare curve (superfici) che siano **polinomiali e razionali a tratti**. Allora vogliamo una rappresentazione di una curva del tipo

$$C(u) = \sum_{i=0}^n f_i(u)P_i \quad (4.5.19)$$

dove i P_i sono i *punti di controllo*, e le $\{f_i(u), i = 0, \dots, n\}$ sono *funzioni polinomiali a tratti* che formano una base dello spazio vettoriale di tutte le funzioni polinomiali a tratti del grado e della continuit  desiderati (per una sequenza fissata di **punti di rottura** (*'breakpoints'*), $U = \{u_i\}, 0 \leq i \leq m$). Si noti che la continuit  é determinata dalle funzioni di base, da cui i punti di controllo possono essere modificati senza alterare la continuit  della curva. Inoltre, le $\{f_i\}$ devono avere le solite propriet  analitiche che garantiscono che le curve definite dall'equazione (4.5.19) hanno propriet  geometriche simili a quelle delle curve di B zier, cio  copertura convessa, 'variation diminishing', invarianza per trasformazioni. Un'altra propriet  importante che si chiede nelle nostre funzioni di base é quella del **supporto locale**, che implica che ogni $f_i(u)$ sia diversa da zero solo su un numero limitato di sottointervalli, non sull'intero dominio, $[u_0, u_n]$. Dato che P_i é moltiplicato per $f_i(u)$, un movimento di P_i ha effetto sulla forma della curva solo sul sottointervallo dove $f_i(u)$ é diversa da zero. Infine costruiamo curve razionali a tratti, con le funzioni polinomiali di base a tratti

$$C^w(u) = \sum_{i=0}^n f_i(u)P_i^w \quad (4.5.20)$$

e superfici prodotto di tensori non razionali e razionali

$$\begin{aligned} S(u, v) &= \sum_{i=0}^n \sum_{j=0}^m f_i(u)g_j(v)P_{i,j} \\ S^w(u, v) &= \sum_{i=0}^n \sum_{j=0}^m f_i(u)g_j(v)P_{i,j}^w \end{aligned} \quad (4.5.21)$$

Dei vari modi di definire le B-splines riprendiamo la formula di ricorrenza dovuta a de Boor, Cox e Mansfield (vedi anche [5] cap. X) perch  é la pi  utile per l'implementazione al computer. Sia $U = \{u_0, \dots, u_m\}$ una sequenza non decrescente di numeri reali, cio  $u_i \leq u_{i+1}, i=0, \dots, m-1$. Le u_i si chiamano **nodi**, e U é il **vettore dei nodi**. L' i -esima funzione B-spline di grado p (*ordine* $p+1$), indicata con $N_{i,p}(u)$, é definita da

$$N_{i,0}(u) = \begin{cases} 1 & \text{se } u_i \leq u \leq u_{i+1} \\ 0 & \text{altrimenti} \end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \quad (4.5.22)$$

Si noti che

- $N_{i,0}(u)$ é una funzione ‘step’, uguale a zero ovunque tranne sull’intervallo semiaperto $u \in [u_i, u_{i+1})$;
- per $p > 0$, $N_{i,p}(u)$ é una combinazione lineare di due funzioni di base di grado $(p-1)$;
- il calcolo di un insieme di funzioni di base richiede di specificare un vettore dei nodi, U , e il grado, p ;
- l’equazione (4.5.22) puó produrre il quoziente $0/0$, definiamo questo quoziente zero;
- le $N_{i,p}(u)$ sono polinomi a tratti, definiti sull’intero asse reale, generalmente solo l’intervallo $[u_0, u_m]$ é interessante;
- l’intervallo semiaperto, $[u_i, u_{i+1})$, é chiamato l’**i-esima ampiezza** (**‘span’**) dei nodi e puó avere lunghezza zero, perché i nodi non devono essere necessariamente distinti;
- il calcolo delle funzioni di grado p genera una tabella triangolare troncata

$$\begin{array}{cccc} N_{0,0} & & & \\ N_{1,0} & N_{0,1} & & \\ N_{2,0} & N_{1,1} & N_{0,2} & N_{0,3} \\ N_{3,0} & N_{2,1} & N_{1,2} & N_{1,3} \\ N_{4,0} & N_{3,1} & N_{2,2} & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Le B-splines hanno una serie di proprietá importanti che le rendono piú interessanti dal punto di vista geometrico per curve e superfici. Poniamo il grado sia p e il vettore dei nodi sia $U = \{u_0, \dots, u_m\}$.

1. $N_{i,p}(u) = 0$ se u é fuori dall’intervallo $[u_i, u_{i+p+1})$ (proprietá del supporto locale).
2. in ogni ‘span’ dei nodi, $[u_j, u_{j+1})$, al massimo $p+1$ degli $N_{i,p}$ sono diversi da zero, cioé le funzioni $N_{j-p,p}, \dots, N_{j,p}$.

3. $N_{i,p} \geq 0$ per ogni i, p e u (non negativit ). Si dimostra per induzione su p .   chiaramente vero per $p=0$, poniamo sia vero per $p-1$, $p \geq 0$, con i e u arbitrari. Per la definizione (4.5.22) e per la propriet  1, $N_{i,p-1}(u) = 0$ se $u \notin [u_i, u_{i+p})$. Ma $u \in [u_i, u_{i+p})$ implica

$$\frac{u - u_i}{u_{i+p} - u_i}$$

  non negativo. Per ipotesi $N_{i,p-1}(u)$   non negativo, quindi il primo termine della (4.5.22)   non negativo. Lo stesso vale per il secondo termine, e quindi gli $N_{i,p}(u)$ sono non negativi.

4. per un arbitraria ampiezza ('span') dei nodi, $[u_i, u_{i+1})$, $\sum_{j=i-p}^i N_{j,p}(u) = 1$ per ogni $u \in [u_i, u_{i+1})$ (partizione dell'unit ).
5. tutte le derivate di $N_{i,p}(u)$ esistono nell'interno di uno 'span' dei nodi (dove   un polinomio). In un nodo $N_{i,p}(u)$   $p-k$ volte continuamente differenziabile, dove k   la molteplicit  del nodo. Da cui crescendo il grado cresce la continuit , e crescendo la molteplicit  del nodo decresce la continuit .
6. eccetto per il caso $p=0$, $N_{i,p}(u)$ raggiunge esattamente un valore massimo.

4.6 Derivate delle Funzioni B-Spline

La derivata di una funzione di base   data da

$$N'_{i,p} = \frac{p}{u_{i+p} - u_i} N_{i,p-1}(u) - \frac{p}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \quad (4.6.23)$$

Lo dimostriamo per induzione su p . Per $p=1$, $N_{i,p-1}$ e $N_{i+1,p-1}$ sono 0 o 1, quindi $N'_{i,p}$  

$$\frac{1}{u_{i+1} - u_i} \text{ o } - \frac{1}{u_{i+2} - u_{i+1}}.$$

Ora assumiamo che l'equazione (4.6.23) sia vera per $p-1$, $p > 1$. Usando la regola del prodotto, $(fg)' = f'g + fg'$, per derivare le funzione di base

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u)$$

si ottiene

$$N'_{i,p} = \frac{1}{u_{i+p} - u_i} N_{i,p-1} + \frac{u - u_i}{u_{i+p} - u_i} N'_{i,p-1} - \frac{1}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1} + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N'_{i+1,p-1} \quad (4.6.24)$$

Con vari passaggi e utilizzando la formula di Cox-de Boor (4.5.22) si arriva alla dimostrazione. Ora sia $N_{i,p}^{(k)}$ la k-esima derivata di $N_{i,p}(u)$, ripetute differenziazioni della (4.6.23) producono la formula generale

$$N_{i,p}^{(k)}(u) = p \left(\frac{N_{i,p-1}^{(k-1)}}{u_{i+p} - u_i} - \frac{N_{i+1,p-1}^{(k-1)}}{u_{i+p+1} - u_{i+1}} \right) \quad (4.6.25)$$

l'equazione (4.6.26) é un'altra generalizzazione della (4.6.23), calcola la k-esima derivata di $N_{i,p}(u)$ in termini delle funzioni $N_{i,p-k}, \dots, N_{i+k,p-k}$

$$N_{i,p}^{(k)} = \frac{p!}{(p-k)!} \sum_{j=0}^k a_{k,j} N_{i+j,p-k} \quad (4.6.26)$$

con

$$a_{0,0} = 1$$

$$a_{k,0} = \frac{a_{k-1,0}}{u_{i+p-k+1} - u_i}$$

$$a_{k,j} = \frac{a_{k-1,j} - a_{k-1,j-1}}{u_{i+p+j-k+1} - u_{i+j}} \quad j = 1, \dots, k-1$$

$$a_{k,k} = \frac{-a_{k-1,k-1}}{u_{i+p+1} - u_{i+k}}$$

Osservazioni sulla (4.6.26):

- k non deve superare p (tutte le derivate piú alte sono zero);
- i denominatori coinvolti nelle differenze di nodi possono diventare zero; il quoziente in questo caso é definito essere zero.

Per completezza, una formula aggiuntiva per calcolare le derivate delle funzioni B-splines é

$$N_{i,p}^{(k)} = \frac{p}{p-k} \left(\frac{u-u_i}{u_{i+p}-u_i} N_{i,p-1}^{(k)} + \frac{u_{i+p+1}-u}{u_{i+p+1}-u_{i+1}} N_{i+1,p-1}^{(k)} \right) \quad k = 0, \dots, p-1 \quad (4.6.27)$$

L'equazione (4.6.27) da la k-esima derivata di $N_{i,p}(u)$ in termini delle k-esime derivate di $N_{i,p-1}$ e $N_{i+1,p-1}$.

4.7 Ulteriori Proprietá delle Funzioni di Base

Sia $\{u_j\}$, $0 \leq j \leq k$, un insieme strettamente crescente di punti di rottura ('breakpoints'). L'insieme di tutte le funzioni polinomiali a tratti di grado p su $\{u_j\}$ che sono C^{r_j} continue in $u = u_j$ formano uno spazio vettoriale, \mathcal{V} ($-1 \leq r_j \leq p$). Se non si impongono vincoli di continuitá ($r_j = -1$ per ogni j), allora la dimensione di \mathcal{V} (indicata con $\dim \mathcal{V}$) é uguale a $k(p+1)$. Ogni vincolo di continuitá decresce la dimensione di uno, perció

$$\dim \mathcal{V} = k(p+1) - \sum_{j=0}^k (r_j + 1) \quad (4.7.28)$$

Dalla proprietá 5 di pag. 67, otteniamo le funzioni B-splines di grado p con nodi $\{u_j\}$, e con la continuitá desiderata, ponendo la molteplicitá appropriata ai nodi, s_j , dove $s_j = p - r_j$. Da cui si usa un vettore dei nodi della forma

$$U = \{ \underbrace{u_0, \dots, u_0}_{s_0}, \underbrace{u_1, \dots, u_1}_{s_1}, \dots, \underbrace{u_k, \dots, u_k}_{s_k} \}.$$

e poniamo

$$m = \left(\sum_{j=0}^k s_j \right) - 1$$

Allora chiaramente ci sono m funzioni di grado zero, $N_{i,0}$, m-1 funzioni di grado 1, $N_{i,1}$, e in generale, m-p funzioni grado p, $N_{i,p}$, che hanno la continuitá desiderata, $r_j = p - s_j$. Da cui gli $N_{i,p}$ sono contenuti in \mathcal{V} . Sostituendo $s_j = p - r_j$ nell'equazione (4.7.28) si ottiene

$$\begin{aligned}
\dim \mathcal{V} &= k(p+1) - \sum_{j=0}^k (p - s_j + 1) \\
&= k(p+1) - (k+1)p + \sum_{j=0}^k s_j - (k+1) \\
&= -p - 1 + \sum_{j=0}^k s_j \\
&= m - p
\end{aligned}$$

Allora, il numero di funzioni B-splines di grado p su U é uguale a $\dim \mathcal{V}$. Giustificiamo ora il termine funzioni di *base* mostrando che le $N_{i,p}$ sono linearmente indipendenti, cioè formano una base per lo spazio vettoriale \mathcal{V} . La dimostrazione é per induzione su p .

Consideriamo, ora, i vettori dei nodi, chiaramente una volta fissato il grado, il vettore dei nodi determina completamente le funzioni $N_{i,p}(u)$. Ci sono diversi tipi di vettori dei nodi. Vettori dei nodi detti **non periodici** o aperti hanno la forma

$$U = \{ \underbrace{a, \dots, a}_{p+1}, u_{p+1}, \dots, u_{m-p-1}, \underbrace{b, \dots, b}_{p+1} \} \quad (4.7.29)$$

cioé, il primo e l'ultimo nodo hanno molteplicitá $p+1$. Per vettori dei nodi non periodici si hanno due proprietá aggiuntive delle funzioni di base:

- 7 un vettore dei nodi della forma

$$U = \{ \underbrace{0, \dots, 0}_{p+1}, \underbrace{1, \dots, 1}_{p+1} \}$$

produce i polinomi di Bernstein di grado p ;

- 8 sia $m+1$ il numero di nodi. Allora ci sono $n+1$ funzioni di base dove $n=m-p-1$; $N_{0,p}(a) = 1$ e $N_{n,p}(b) = 1$. $N_{0,p}(a) = 1$ segue dal fatto che $N_{0,0}, \dots, N_{p-1,0} = 0$, e questo implica $N_{0,p}(a) = N_{p,0}(a) = 1$. Dalla proprietá 4 pag. 67 segue che $N_{i,p}(a) = 0$ per $i \neq 0$ e $N_{i,p}(b) = 0$ per $i \neq n$.

Un vettore dei nodi $U = \{u_0, \dots, u_m\}$ si dice **uniforme** se tutti i nodi interni sono equispaziati, cioè se esiste un numero reale, d , tale che $d = u_{i+1} - u_i$ per ogni $p \leq i \leq m - p - 1$, altrimenti si dice **non uniforme**.

4.8 Definizione e Proprietá delle Curve B-splines

Una **curva B-spline di grado p** é definita da

$$C(u) = \sum_{i=0}^n N_{i,p}(u)P_i \quad a \leq u \leq b \quad (4.8.30)$$

dove i $\{P_i\}$ sono **punti di controllo** e i $\{N_{i,p}(u)\}$ sono le funzioni di base B-splines di grado p definite sul vettore dei nodi non periodico e non uniforme

$$U = \{ \underbrace{a, \dots, a}_{p+1}, u_{p+1}, \dots, u_{m-p-1}, \underbrace{b, \dots, b}_{p+1} \}$$

(m+1 nodi). A meno che fissato altrimenti, assumiamo che a=0 e b=1. Il poligono formato dai $\{P_i\}$ é chiamato il **poligono di controllo**. Tre passi sono richiesti per calcolare un punto su una curva B-spline per un valore fissato di u:

1. trovare lo ‘span’ dei nodi in cui u giace;
2. calcolare le funzioni di base non zero;
3. moltiplicare i valori delle funzioni di base non zero per i corrispondenti punti di controllo.

Alcune proprietá delle curve B-splines, seguono da quelle delle funzioni $N_{i,p}(u)$. Sia C(u) definito dall’equazione (4.8.30)

1. se n=p e $U = \{0, \dots, 0, 1, \dots, 1\}$, allora C(u) é una curva di Bézier;
2. C(u) é una curva polinomiale a tratti (dato che $N_{i,p}(u)$ sono polinomi a tratti); il grado, p, il numero dei punti di controllo, n+1, e il numero dei nodi, m+1, sono correlati da

$$m = n + p + 1 \quad (4.8.31)$$

3. **interpolazione agli estremi:** $C(0) = P_0$ e $C(1) = P_n$;
4. **invarianza affine:** una trasformazione affine é applicata alla curva applicandola ai punti di controllo. Sia r un punto in \mathcal{E}^3 (spazio Euclideo tridimensionale). Una *trasformazione affine*, indicata con Φ , porta \mathcal{E}^3 in \mathcal{E}^3 e ha la forma

$$\Phi(r) = Ar + v$$

dove A é una matrice 3×3 e v é una vettore. Le trasformazioni affini includono traslazioni, rotazioni, scale e tagli ('shears'). La proprietá di invarianza affine per le curve B-splines segue dalla proprietá della partizione dell'unitá di $N_{i,p}(u)$. Quindi, sia $r = \sum \alpha_i p_i$, dove $p_i \in \mathcal{E}^3$ e $\sum \alpha_i = 1$. Allora

$$\begin{aligned}\Phi(r) &= \Phi\left(\sum \alpha_i p_i\right) = A\left(\sum \alpha_i p_i\right) + v = \sum \alpha_i A p_i + \sum \alpha_i v \\ &= \sum \alpha_i (A p_i + v) = \sum \alpha_i \Phi(p_i)\end{aligned}$$

5. **proprietá forte della copertura convessa:** la curva é contenuta nelle copertura convessa del suo poligono di controllo; infatti, se $u \in [u_i, u_{i+1})$, $p \leq i < m - p - 1$, allora $C(u)$ é nella copertura convessa dei punti di controllo P_{i-p}, \dots, P_i . Questo segue dalle proprietá di non negativitá e partizione dell'unitá degli $N_{i,p}(u)$, e la proprietá che $N_{j,p}(u) = 0$ per $j < i - p$ e $j > i$ quando $u \in [u_i, u_{i+1})$.
6. **schema di modificazione locale:** muovendo P_i cambia $C(u)$ solo nell'intervallo $[u_i, u_{i+p+1})$. Questo segue dal fatto che $N_{i,p}(u) = 0$ per $u \notin [u_i, u_{i+p+1})$.
7. il poligono di controllo rappresenta un'approssimazione lineare a tratti della curva; questa approssimazione é migliorata dall'inserimento di un nodo o dall'aumento del grado. Come regola generale, piú basso é il grado, piú vicino la curva B-spline segue il poligono di controllo.
8. muovendosi lungo la curva da $u=0$ a $u=1$, le funzioni $N_{i,p}(u)$ agiscono come interruttori; quando u passa vicino a un nodo, una $N_{i,p}(u)$ (e quindi il corrispondente P_i) si spegne, e la seguente si accende.
9. **proprietá di 'variation diminishing':** nessun piano ha piú intersezioni con la curva che con il poligono di controllo (per le curve bidimensionali si cambia la parola piano con retta);
10. La continuitá e la differenziabilitá di $C(u)$ seguono da quelle di $N_{i,p}(u)$ (dato che $C(u)$ é esattamente una combinazione lineare degli $N_{i,p}(u)$). Quindi, $C(u)$ é infinitamente differenziabile nell'interno degli intervalli dei nodi, ed é almeno $p-k$ volte continuamente differenziabile in un nodo di molteplicitá k .
11. é possibile e, a volte utile, utilizzare punti di controllo multipli (coincidenti).

4.9 Le Derivate di una Curva B-spline

Sia $C^{(k)}(u)$ la k -esima derivata di $C(u)$. Se u é fissato, possiamo ottenere $C^{(k)}(u)$ calcolando la k -esima derivata delle funzioni di base. In particolare

$$C^{(k)}(u) = \sum_{i=0}^n N_{i,p}^{(k)}(u) P_i \quad (4.9.32)$$

Mentre se non si fissa u , si differenzia formalmente la curva B-spline di grado p

$$C(u) = \sum_{i=0}^n N_{i,p} P_i$$

definita sul vettore dei nodi

$$U = \{ \underbrace{0, \dots, 0}_{p+1}, u_{p+1}, \dots, u_{m-p-1}, \underbrace{1, \dots, 1}_{p+1} \}$$

Dall'equazione (4.9.32) e (4.6.23) otteniamo

$$\begin{aligned} C'(u) &= \sum_{i=0}^n N'_{i,p}(u) P_i \\ &= \sum_{i=0}^n \left(\frac{p}{u_{i+p} - u_i} N_{i,p-1}(u) - \frac{p}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \right) P_i \\ &= \left(p \sum_{i=-1}^{n-1} N_{i+1,p-1}(u) \frac{P_{i+1}}{u_{i+p+1} - u_{i+1}} \right) \\ &\quad - \left(p \sum_{i=0}^n N_{i+1,p-1}(u) \frac{P_i}{u_{i+p+1} - u_{i+1}} \right) \\ &= p \frac{N_{0,p-1}(u) P_0}{u_p - u_0} + p \sum_{i=0}^{n-1} N_{i+1,p-1}(u) \frac{P_{i+1} - P_i}{u_{i+p+1} - u_{i+1}} - p \frac{N_{n+1,p-1}(u) P_n}{u_{n+p+1} - u_{n+1}} \end{aligned}$$

Il primo e l'ultimo termine sono $0/0$, che é 0 per definizione. Quindi

$$C'(u) = p \sum_{i=0}^{n-1} N_{i+1,p-1}(u) \frac{P_{i+1} - P_i}{u_{i+p+1} - u_{i+1}} = \sum_{i=0}^{n-1} N_{i+1,p-1}(u) Q_i$$

dove

$$Q_i = p \frac{P_{i+1} - P_i}{u_{i+p+1} - u_{i+1}} \quad (4.9.33)$$

Ora sia U' il vettore dei nodi ottenuto togliendo il primo e l'ultimo dei nodi da U , cioè

$$U = \{ \underbrace{0, \dots, 0}_p, u_{p+1}, \dots, u_{m-p-1}, \underbrace{1, \dots, 1}_p \} \quad (4.9.34)$$

(U' ha m-1 nodi). Allora é facile verificare che la funzione $N_{i+1,p-1}(u)$, calcolato su U, é uguale a $N_{i,p-1}(u)$ calcolato su U'. Quindi

$$C'(u) = \sum_{i=0}^{n-1} N_{i,p-1}(u) Q_i \quad (4.9.35)$$

dove i Q_i sono definiti dall'equazione (4.9.33) e gli $N_{i,p-1}(u)$ sono calcolati su U'. Da cui $C'(u)$ é una curva B-spline di grado p-1.

Dato che $C'(u)$ é una curva B-spline, applichiamo le equazioni dalla (4.9.33) alla (4.9.35) ricorsivamente per ottenere derivate piú alte. Ponendo $P_i^{(0)} = P_i$, scriviamo

$$C(u) = C^{(0)}(u) = \sum_{i=0}^n N_{i,p}(u) P_i^{(0)}.$$

Allora

$$C^{(k)}(u) = \sum_{i=0}^{n-k} N_{i,p-k}(u) P_i^{(k)} \quad (4.9.36)$$

con

$$P_i^{(k)} = \begin{cases} P_i & k = 0 \\ \frac{p-k+1}{u_{i+p+1}-u_{i+k}} (P_{i+1}^{(k-1)} - P_i^{(k-1)}) & k > 0 \end{cases}$$

e

$$U^{(k)} = \{ \underbrace{0, \dots, 0}_{p-k+1}, u_{p+1}, \dots, u_{m-p-1}, \underbrace{1, \dots, 1}_{p-k+1} \}. \quad (4.9.37)$$

4.10 Definizione e Proprietá di una Superficie B-spline

Una superficie B-spline si ottiene prendendo una rete bidirezionale di punti di controllo, due vettori dei nodi e i prodotti delle funzioni B-spline univariate

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u) N_{j,q}(v) P_{i,j} \quad (4.10.38)$$

con

$$U = \{ \underbrace{0, \dots, 0}_{p+1}, u_{p+1}, \dots, u_{r-p-1}, \underbrace{1, \dots, 1}_{p+1} \}$$

$$V = \{ \underbrace{0, \dots, 0}_{q+1}, v_{q+1}, \dots, v_{s-q-1}, \underbrace{1, \dots, 1}_{q+1} \}$$

U ha r+1 nodi e V ne ha s+1. L'equazione (4.8.31) prende la forma

$$r = n + p + 1 \quad e \quad s = m + q + 1 \quad (4.10.39)$$

Le proprietá del tensore prodotto delle funzioni base seguono dalle corrispondenti proprietá delle funzioni di base univariate elencate nel paragrafo 4.4

- 12 **non negativitá:** $N_{i,p}(u)N_{j,q}(v) \geq 0$ per ogni i,j,p,q,u,v;
- 13 **partizione dell'unitá:** $\sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u)N_{j,q}(v) = 1$ per ogni $(u, v) \in [0, 1] \times [0, 1]$;
- 14 se n=p, m=q, $U = \{0, \dots, 0, 1, \dots, 1\}$ e $V = \{0, \dots, 0, 1, \dots, 1\}$, allora $N_{i,p}(u)N_{j,q}(v) = B_{i,n}(u)B_{j,m}(v)$ per ogni i,j; cioé prodotti di funzioni B-splines degenerano a prodotti di polinomi di Bernstein;
- 15 $N_{i,p}(u)N_{j,q}(v) = 0$ se (u, v) é fuori dal rettangolo $[u_i, u_{i+p+1}) \times [v_j, v_{j+q+1})$;
- 16 in ogni rettangolo dato, $[u_{i_0}, u_{i_0+1}) \times [v_{j_0}, v_{j_0+1})$, al massimo $(p+1)(q+1)$ funzioni di base non sono zero, in particolare le $N_{i,p}N_{j,q}$ per $i_0 - p \leq i \leq i_0$ e $j_0 - q \leq j \leq j_0$;
- 17 se $p > 0$ e $q > 0$, allora $N_{i,p}N_{j,q}$ raggiungono esattamente un valore massimo;

- 18 Internamente al rettangolo formato dalle linee dei nodi u e v , dove la funzione é un polinomio bivariato, tutte le derivate parziali di $N_{i,p}N_{j,q}$ esistono; nel nodo u (nodo v) essa é $p-k$ ($q-k$) volte derivabile nella direzione di u (v), dove k é la molteplicitá del nodo.

Le superfici B-spline hanno le seguenti proprietá:

- 19 se $n=p$, $m=q$, $U = \{0, \dots, 0, 1, \dots, 1\}$ e $V = \{0, \dots, 0, 1, \dots, 1\}$, allora $S(u,v)$ é una superficie di Bézier (segue dalla 14);
- 20 la superficie interpola i quattro punti di controllo d'angolo: $S(0,0) = P_{0,0}$, $S(1,0) = P_{n,0}$, $S(0,1) = P_{0,m}$ e $S(1,1) = P_{n,m}$; questo segue dalla 13 e dall'identitá

$$N_{0,p}(0)N_{0,q}(0) = N_{n,p}(1)N_{0,q}(0) = N_{0,p}(0)N_{m,q}(1) = N_{n,p}(1)N_{m,q}(1) = 1$$

- 21 **invarianza affine**: una trasformazione affine é applicata alla superficie applicandola ai punti di controllo; questa segue dalla 13;
- 22 **proprietá forte della copertura convessa**: se $(u, v) \in [u_{i_0}, u_{i_0+1}] \times [v_{j_0}, v_{j_0+1}]$, allora $S(u,v)$ é nella copertura convessa dei punti di controllo $P_{i,j}$, $i_0 - p \leq i \leq i_0$ e $j_0 - q \leq j \leq j_0$; questa segue dalle 12,13 e 16;
- 23 se triangolata, la rete di controllo forma un'approssimazione piana a tratti della superficie; come nel caso delle curve, piú piccolo é il grado meglio é l'approssimazione;
- 24 **schema locale di modificazione**: se $P_{i,j}$ viene mosso ciò influisce sulla superficie solo nel rettangolo $[u_i, u_{i+p+1}] \times [v_j, v_{j+q+1}]$; questa segue dalla 15;
- 25 la continuitá e la differenziabilitá di $S(u,v)$ segue da quella delle funzioni di base. In particolare, $S(u,v)$ é $p-k$ ($q-k$) volte differenziabile nella direzione di u (v) nel nodo u (v) di molteplicitá k .

Non c'è proprietá di 'variation diminishing' per le superfici B-splines. Le curve isoparametriche (isocurve) su $S(u,v)$ si ottengono in maniera analoga a quelle delle superfici di Bézier. Fissato $u = u_0$

$$\begin{aligned} C_{u_0}(v) &= S(u_0, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u_0)N_{j,q}(v)P_{i,j} \\ &= \sum_{j=0}^m N_{j,q}(v) \left(\sum_{i=0}^n N_{i,p}(u_0)P_{i,j} \right) = \sum_{j=0}^m N_{j,q}(v)Q_j(u_0) \end{aligned} \quad (4.10.40)$$

dove

$$Q_j(u_0) = \sum_{i=0}^n N_{i,p}(u_0) P_{i,j}$$

Analogamente per $v = v_0$ fissato:

$$C_{v_0}(u) = \sum_{i=0}^n N_{i,p}(u) Q_i(v_0),$$

dove

$$Q_i(v_0) = \sum_{j=0}^m N_{j,q}(v_0) P_{i,j}, \quad (4.10.41)$$

é una u isocurva su $S(u,v)$. $C_{u_0}(v)$ é una curva B-spline di grado q su V e $C_{v_0}(u)$ é una curva B-spline di grado p su U. Il punto $S(u_0, v_0)$ é l'intersezione tra $C_{u_0}(v)$ e $C_{v_0}(u)$.

4.11 Derivate di una Superficie B-spline

Sia (u,v) fissato. Generalmente, si é interessati a calcolare tutte le derivate parziali di $S(u,v)$ fino ad un ordine d incluso, cioè

$$\frac{\partial^{k+l}}{\partial^k u \partial^l v} S(u, v) \quad 0 \leq k + l \leq d \quad (4.11.42)$$

Come per le curve, otteniamo queste derivate calcolando le derivate delle funzioni di base. In particolare

$$\frac{\partial^{k+l}}{\partial^k u \partial^l v} S(u, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}^{(k)} N_{j,q}^{(l)} P_{i,j} \quad (4.11.43)$$

Differenziamo formalmente $S(u,v)$. Rispetto a u noi abbiamo

$$S_u(u, v) = \frac{\partial}{\partial u} S(u, v) = \sum_{j=0}^m N_{j,q}(v) \left(\frac{\partial}{\partial u} \sum_{i=0}^n N_{i,p}(u) P_{i,j} \right) \quad (4.11.44)$$

$$= \sum_{j=0}^m N_{j,q}(v) \left(\frac{\partial}{\partial u} C_j(u) \right)$$

dove

$$C_j(u) = \sum_{i=0}^n N_{i,p}(u) P_{i,j} \quad j = 0, \dots, m$$

sono curve B-splines. Applicando l'equazione (4.9.35) a ognuno dei $C_j(u)$ e sostituendolo nell'equazione (4.11.44), otteniamo

$$S_u(u, v) = \sum_{i=0}^{n-1} \sum_{j=0}^m N_{i,p-1}(u) N_{j,q}(v) P_{i,j}^{(1,0)} \quad (4.11.45)$$

dove

$$P_{i,j}^{(1,0)} = p \frac{P_{i+1,j} - P_{i,j}}{u_{i+p+1} - u_{i+1}}$$

$$U^{(1)} = \left\{ \underbrace{0, \dots, 0}_p, u_{p+1}, \dots, u_{r-p-1}, \underbrace{1, \dots, 1}_p \right\}$$

$$V^{(0)} = V$$

Analogamente

$$S_v(u, v) = \sum_{i=0}^n \sum_{j=0}^{m-1} N_{i,p}(u) N_{j,q-1}(v) P_{i,j}^{(0,1)} \quad (4.11.46)$$

dove

$$P_{i,j}^{(0,1)} = q \frac{P_{i,j+1} - P_{i,j}}{v_{j+q+1} - v_{j+1}}$$

$$U^{(0)} = U$$

$$V^{(1)} = \left\{ \underbrace{0, \dots, 0}_q, v_{q+1}, \dots, v_{s-q-1}, \underbrace{1, \dots, 1}_q \right\}$$

Applicando prima la (4.11.45), poi la (4.11.46) si ottiene

$$S_{uv}(u, v) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} N_{i,p-1}(u) N_{j,q-1}(v) P_{i,j}^{(1,1)} \quad (4.11.47)$$

dove

$$P_{i,j}^{(1,1)} = q \frac{P_{i,j+1}^{(1,0)} - P_{i,j}^{(1,0)}}{v_{j+q+1} - v_{j+1}}$$

e $U^{(1)}$ e $V^{(1)}$ sono già stati definiti. In generale

$$\frac{\partial^{k+l}}{\partial^k u \partial^l v} S(u, v) = \sum_{i=0}^{n-k} \sum_{j=0}^{m-l} N_{i,p-k} N_{j,q-l} P_{i,j}^{(k,l)} \quad (4.11.48)$$

dove

$$P_{i,j}^{(k,l)} = (q-l+1) \frac{P_{i,j+1}^{(k,l-1)} - P_{i,j}^{(k,l-1)}}{v_{j+q+1} - v_{j+1}}$$

Usando le equazioni dalla (4.11.45) alla (4.11.48), si ricavano le utili formule per le derivate agli angoli. Per es. all'angolo $(u,v)=(0,0)$ si ha

$$S_u(0, 0) = P_{0,0}^{(1,0)} = \frac{p}{u_{p+1}} (P_{1,0} - P_{0,0})$$

$$S_v(0, 0) = P_{0,0}^{(0,1)} = \frac{q}{v_{q+1}} (P_{0,1} - P_{0,0}) \quad (4.11.49)$$

$$S_{uv}(0, 0) = P_{0,0}^{(1,1)} = \frac{q}{v_{q+1}} (P_{0,1}^{(1,0)} - P_{0,0}^{(1,0)})$$

$$\frac{pq}{u_{p+1} v_{q+1}} (P_{1,1} - P_{0,1} - P_{1,0} + P_{0,0})$$

Ora sia $u_0 = 0$ e $v_0 = 0$. Dalle proprietà delle funzioni di base é facile vedere che le isocurve $C_{u_0}(v)$ e $C_{v_0}(u)$ sono date da

$$C_{u_0}(v) = \sum_{j=0}^m N_{j,q}(v) P_{0,j} \quad C_{v_0}(u) = \sum_{i=0}^n N_{i,p}(u) P_{i,0}$$

Capitolo 5

NURBS

5.1 Definizione e Proprietá delle Curve NURBS

Una **curva NURBS** (acronimo di **Non Uniform Rational B-Spline**) di **grado p** é definita da

$$C(u) = \frac{\sum_{i=0}^n N_{i,p}(u)w_i P_i}{\sum_{i=0}^n N_{i,p}(u)w_i} \quad a \leq u \leq b \quad (5.1.1)$$

dove i $\{P_i\}$ sono i **punti di controllo** (che formano il **poligono di controllo**), i $\{w_i\}$ sono i **pesi** e gli $\{N_{i,p}(u)\}$ sono le **funzioni di base B-splines di grado p** definite su un vettore di nodi non periodico e non uniforme

$$U = \{ \underbrace{a, \dots, a}_{p+1}, u_{p+1}, \dots, u_{m-p-1}, \underbrace{b, \dots, b}_{p+1} \}$$

A meno che altrimenti stabilito, assumiamo che $a=0$, $b=1$ e $w_i > 0$ per ogni i . Ponendo

$$R_{i,p}(u) = \frac{N_{i,p}(u)w_i}{\sum_{j=0}^n N_{j,p}(u)w_j} \quad (5.1.2)$$

possiamo riscrivere l'equazione (5.1.1) nella forma

$$C(u) = \sum_{i=0}^n R_{i,p}(u)P_i \quad (5.1.3)$$

Gli $\{R_{i,p}(u)\}$ sono le **funzioni razionali di base** e sono funzioni razionali a tratti su $u \in [0, 1]$.

Gli $R_{i,p}(u)$ hanno le seguenti proprietà derivate dall'equazione (4.2.2) e corrispondenti alle proprietà degli $N_{i,p}(u)$:

1. **non negatività:** $R_{i,p}(u) \geq 0$ per ogni i, p e $u \in [0, 1]$;
2. **partizione dell'unità:** $\sum_{i=0}^n R_{i,p}(u) = 1$ per ogni $u \in [0, 1]$;
3. $R_{0,p}(0) = R_{n,p}(1) = 1$;
4. per $p > 0$ tutti gli $R_{i,p}(u)$ raggiungono esattamente un massimo sull'intervallo $u \in [0, 1]$;
5. **supporto locale:** $R_{i,p}(u) = 0$ per $u \notin [u_i, u_{i+p+1})$. Inoltre in ogni 'span' dei nodi al massimo $p+1$ degli $R_{i,p}(u)$ non sono zero (in generale $R_{i-p,p}(u), \dots, R_{i,p}(u)$ non sono zero in $[u_i, u_{i+p+1})$);
6. tutte le derivate di $R_{i,p}(u)$ esistono nell'interno dello 'span' dei nodi, dove sono una funzione razionale con denominatore non zero. In un nodo, $R_{i,p}(u)$ é $p-k$ volte continuamente differenziabile, dove k é la molteplicitá del nodo;
7. se $w_i = 1$ per ogni i , allora $R_{i,p}(u) = N_{i,p}(u)$ per ogni i ; cioè gli $N_{i,p}(u)$ sono casi speciali degli $R_{i,p}(u)$. In effetti per ogni $a \neq 0$, se $w_i = a$ per ogni i allora $R_{i,p}(u) = N_{i,p}(u)$ per ogni i .

Queste proprietà hanno come conseguenza le seguenti importanti caratteristiche geometriche delle curve NURBS:

8. $C(0) = P_0$ e $C(1) = P_n$; questo segue dalla 3;
9. **invarianza affine:** una trasformazione affine é applicata alla curva applicandola ai punti di controllo; le curve NURBS sono anche invarianti rispetto alle proiezioni prospettiche, fatto importante in 'computer graphics';
10. **proprietá forte della copertura convessa:** se $u \in [u_i, u_{i+1})$, allora $C(u)$ giace dentro la copertura convessa dei punti di controllo P_{i-p}, \dots, P_i ; questo segue dalle proprietà 1, 2 e 5;

11. $C(u)$ é infinitamente differenziabile nell'interno degli 'spans' dei nodi ed é p-k volte differenziabile in un nodo di molteplicitá k;
12. **proprietá di 'variation diminishing'**: nessun piano ha piú intersezioni con la curva che con il poligono di controllo (per curve bidimensionali si cambia 'piano' con 'retta');
13. una curva NURBS senza nodi interni é una curva razionale di Bézier, poiché gli $N_{i,p}(u)$ si riducono a $B_{i,p}(u)$; confrontando le equazioni (5.1.2) e (5.1.3) con l'equazione (4.3.10), insieme alla proprietá 7, implica che le curve NURBS contengono le B-splines non razionali e le curve di Bézier razionali e non razionali come casi speciali;
14. **approssimazione locale**: se il punto di controllo P_i viene mosso, o il peso w_i viene cambiato, ciò ha effetto solo sulla porzione della curva sull'intervallo $u \in [u_i, u_{i+p+1})$; questo segue dalla 5.

L'effetto qualitativo ottenuto aumentando (diminuendo) un singolo peso w_i , dato $u \in [u_i, u_{i+p+1})$, é di muovere il punto $C(u)$ piú vicino (piú lontano) da P_i e quindi la curva é attirata verso (spinta lontano da) P_i . Inoltre, il movimento di $C(u)$ per u fissato é lungo una retta. Come per il caso delle curve di Bézier, le coordinate omogenee offrono un efficiente metodo per rappresentare le curve NURBS. Sia H l'applicazione prospettica data dall'equazione (4.3.11), per un insieme dato di punti di controllo, $\{P_i\}$, e pesi, $\{w_i\}$, costruiamo i punti di controllo pesati, $P_i^w = (w_i x_i, w_i y_i, w_i z_i, w_i)$. Allora definiamo curva B-spline non razionale (polinomiale a tratti) nello spazio quadridimensionale come

$$C^w(u) = \sum_{i=0}^n N_{i,p}(u) P_i^w \quad (5.1.4)$$

Applicando l'applicazione prospettica, H , a $C^w(u)$ si ottiene la corrispondente curva B-spline razionale (a tratti nello spazio tridimensionale)

$$\begin{aligned} C(u) &= H \{C^w(u)\} = H \left\{ \sum_{i=0}^n N_{i,p}(u) P_i^w \right\} \\ &= \frac{\sum_{i=0}^n N_{i,p} w_i P_i}{\sum_{i=0}^n N_{i,p} w_i} = \sum_{i=0}^n R_{i,p} P_i. \end{aligned}$$

5.2 Derivate di una Curva NURBS

Sia

$$C(u) = \frac{w(u)C(u)}{w(u)} = \frac{A(u)}{w(u)}$$

dove $A(u)$ é la funzione stimata sul vettore le cui coordinate sono le prime tre coordinate di $C^w(u)$ ($A(u)$ é il numeratore dell'equazione (5.1.1)). Allora

$$\begin{aligned} C'(u) &= \frac{w(u)A'(u) - w'(u)A(u)}{w(u)^2} \\ &= \frac{w(u)A'(u) - w'(u)w(u)C(u)}{w(u)^2} = \frac{A'(u) - w'(u)C(u)}{w(u)} \end{aligned} \quad (5.2.5)$$

Poiché $A(u)$ e $w(u)$ rappresentano le coordinate di $C^w(u)$, otteniamo le loro derivate prime usando le equazioni dalla (4.9.33) alla (4.9.35). Calcoliamo le derivate di ordine piú alto differenziando $A(u)$ usando la regola di Leibnitz

$$\begin{aligned} A^{(k)}(u) &= (w(u)C(u))^{(k)} = \sum_{i=0}^k \binom{k}{i} w^{(i)}(u)C^{(k-i)}(u) \\ &= w(u)C^{(k)}(u) + \sum_{i=1}^k \binom{k}{i} w^{(i)}(u)C^{(k-i)}(u) \end{aligned}$$

da cui otteniamo

$$C^{(k)}(u) = \frac{A^{(k)}(u) - \sum_{i=1}^k \binom{k}{i} w^{(i)}(u)C^{(k-i)}(u)}{w(u)} \quad (5.2.6)$$

La (5.2.6) da la k-esima derivata di $C(u)$ in termini della k-esima derivata di $A(u)$ e la prima attraverso le (k-1)-esime derivate di $C(u)$ e $w(u)$.

5.3 Definizione e Proprietá delle Superfici NURBS

Una superficie NURBS di grado p nella direzione di u e grado q nella direzione di v é una funzione razionale a tratti bivariata, valutata su un vettore, della forma

$$S(u, v) = \frac{\sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u) N_{j,q}(v) w_{i,j} P_{i,j}}{\sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u) N_{j,q}(v) w_{i,j}} \quad 0 \leq u, v \leq 1 \quad (5.3.7)$$

I $\{P_{i,j}\}$ formano una rete di controllo bidirezionale, i $\{w_{i,j}\}$ sono i pesi e $\{N_{i,p}(u)\}$ e $\{N_{j,q}(v)\}$ sono le funzioni di base B-splines non razionali definite sui vettori dei nodi

$$U = \left\{ \underbrace{0, \dots, 0}_{p+1}, u_{p+1}, \dots, u_{r-p-1}, \underbrace{1, \dots, 1}_{p+1} \right\}$$

$$V = \left\{ \underbrace{0, \dots, 0}_{q+1}, v_{q+1}, \dots, v_{s-q-1}, \underbrace{1, \dots, 1}_{q+1} \right\}$$

dove $r=n+p+1$ e $s=m+q+1$. Introducendo le funzioni di base razionali a tratti

$$R_{i,j}(u, v) = \frac{N_{i,p}(u) N_{j,q}(v) w_{i,j}}{\sum_{k=0}^n \sum_{l=0}^m N_{k,p}(u) N_{l,q}(v) w_{k,l}} \quad (5.3.8)$$

la superficie dell'equazione (5.3.7) si può scrivere

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^m R_{i,j}(u, v) P_{i,j} \quad (5.3.9)$$

Le importanti proprietá delle funzioni $R_{i,j}(u, v)$ sono le stesse di quelle per le funzioni di base non razionali, $N_{i,p}(u)N_{j,q}(v)$:

- **15 non negativitá:** $R_{i,j}(u, v) \geq 0$ per ogni i, j, u e v ;
- **16 partizione dell'unitá:** $\sum_{i=0}^n \sum_{j=0}^m R_{i,j}(u, v) = 1$ per ogni $(u, v) \in [0, 1] \times [0, 1]$;

- **17 supporto locale:** $R_{i,j}(u, v) = 0$ se (u, v) é fuori dal rettangolo dato da $[u_i, u_{i+p+1}) \times [v_j, v_{j+q+1})$;
- **18** in ogni rettangolo della forma $[u_{i_0}, u_{i_0+1}) \times [v_{j_0}, v_{j_0+1})$, al massimo $(p+1)(q+1)$ funzioni di base non sono zero, in particolare gli $R_{i,j}(u, v)$ per $i_0 - p \leq i \leq i_0$ e $j_0 - q \leq j \leq j_0$ non sono zero;
- **19 estremi:** se $p > 0$ e $q > 0$, allora $R_{i,j}(u, v)$ raggiunge esattamente un valore massimo;
- **20** $R_{0,0}(0, 0) = R_{n,0}(1, 0) = R_{0,m}(0, 1) = R_{n,m}(1, 1) = 1$;
- **21 differenziabilit :** internamente al rettangolo formato dalle linee dei nodi u e v , tutte le derivate parziali di $R_{i,j}(u, v)$ esistono. nel nodo u (nodo v) é $p-k$ ($q-k$) volte differenziabile nella direzione di u (v), dove k é la molteplicit  del nodo;
- **22** se tutti i $w_{i,j} = a$ per ogni $0 \leq i \leq n$, $0 \leq j \leq m$, e $a \neq 0$, allora $R_{i,j}(u, v) = N_{i,p}(u)N_{j,q}(v)$ per ogni i, j .

Le propriet  dalla 15 alla 22 danno le seguenti importanti propriet  geometriche delle superfici NURBS:

- **23 interpolazione nei punti agli angoli:** $S(0, 0) = P_{0,0}$, $S(1, 0) = P_{n,0}$, $S(0, 1) = P_{0,m}$ e $S(1, 1) = P_{n,m}$;
- **24 invarianza affine:** una trasformazione affine é applicata alla superficie applicandola ai punti di controllo;
- **25 propriet  forte della copertura convessa:** assumiamo $w_{i,j} \geq 0$ per ogni i, j . Se $(u, v) \in [u_{i_0}, u_{i_0+1}) \times [v_{j_0}, v_{j_0+1})$, allora $S(u, v)$ é nella copertura convessa dei punti di controllo $P_{i,j}$, $i_0 - p \leq i \leq i_0$ e $j_0 - q \leq j \leq j_0$;
- **26 modificazione locale:** se $P_{i,j}$ viene mosso, o $w_{i,j}$ viene cambiato, ci  influisce sulla forma della superficie solo nel rettangolo $[u_i, u_{i+p+1}) \times [v_j, v_{j+q+1})$;
- **27** le superfici B-splines non razionali, di B zier e razionali di B zier sono speciali casi di superfici NURBS;
- **28 differenziabilit :** $S(u, v)$ é $p-k$ ($q-k$) volte differenziabile rispetto a u (v) nel nodo u (nodo v) di molteplicit  k ;

Si ricorda che non c'  la propriet  di 'variation diminishing' per le superfici NURBS. Si pu  usare sia il movimento dei punti di controllo che la modifica dei pesi per cambiare localmente la forma delle superfici NURBS. Qualitativamente l'effetto su una superficie, dato $(u, v) \in [u_i, u_{i+p+1}) \times [v_j, v_{j+q+1})$,  : crescendo (decrecendo) $w_{i,j}$, il punto $S(u, v)$ si muove pi  vicino a (pi  lontano da) $P_{i,j}$, e da ci  la superficie   tirata verso (spinta via da) $P_{i,j}$. Come nel caso delle curve, il movimento di $S(u, v)$   lungo una retta.

É conveniente rappresentare una superficie NURBS utilizzando le coordinate omogenee, cioè

$$S^w(u, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u) N_{j,q}(v) P_{i,j}^w \quad (5.3.10)$$

dove $P_{i,j}^w = (w_{i,j}x_{i,j}, w_{i,j}y_{i,j}, w_{i,j}z_{i,j}, w_{i,j})$. Allora $S(u, v) = H \{S^w(u, v)\}$. Applicando le equazioni (4.10.38) e (4.10.39), otteniamo curve isoparametriche su una superficie NURBS. Prima, fissato $u = u_0$

$$C_{u_0}^w(v) = S^w(u_0, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u_0) N_{j,q}(v) P_{i,j}^w \quad (5.3.11)$$

$$\begin{aligned} &= \sum_{j=0}^m N_{j,q}(v) \left(\sum_{i=0}^n N_{i,p}(u_0) P_{i,j}^w \right) \\ &= \sum_{j=0}^m N_{j,q}(v) Q_j^w(u_0) \end{aligned}$$

dove

$$Q_j^w(u_0) = \sum_{i=0}^n N_{i,p}(u_0) P_{i,j}^w$$

Analogamente

$$C_{v_0}^w(u) = \sum_{i=0}^n N_{i,p}(u) Q_i^w(v_0) \quad (5.3.12)$$

dove

$$Q_i^w(v_0) = \sum_{j=0}^m N_{j,q}(v_0) P_{i,j}^w$$

$C_{u_0}^w(v)$ ($C_{v_0}^w(u)$) é una curva NURBS di grado q (p) sul vettore dei nodi V (U). Il punto $S^w(u_0, v_0)$ giace sull'intersezione di $C_{u_0}^w(v)$ e $C_{v_0}^w(u)$. Proiettando si ha

$$C_{u_0}(v) = H \{C_{u_0}^w(v)\} = H \{S^w(u_0, v)\} = S(u_0, v)$$

$$C_{v_0}(u) = H \{C_{v_0}^w(u)\} = H \{S^w(u, v_0)\} = S(u, v_0) \quad (5.3.13)$$

5.4 Derivate di una Superficie NURBS

Le derivate di $S^w(u, v)$ sono calcolate utilizzando le equazioni dalla (4.10.41) alla (4.11.47). Ora si ricavano le formule per le derivate di $S(u, v)$ in termini di quelle di $S^w(u, v)$. Sia

$$S(u, v) = \frac{w(u, v)S(u, v)}{w(u, v)} = \frac{A(u, v)}{w(u, v)}$$

dove $A(u, v)$ é il numeratore di $S(u, v)$ (eq. (5.3.7)). Allora

$$S_\alpha(u, v) = \frac{A_\alpha(u, v) - w_\alpha(u, v)S(u, v)}{w(u, v)} \quad (5.4.14)$$

dove α indica entrambi u o v . In generale

$$\begin{aligned} A^{(k,l)} &= [(wS)^k]^l = \left(\sum_{i=0}^k \binom{k}{i} w^{(i,0)} S^{(k-i,0)} \right)^l \\ &= \sum_{i=0}^k \binom{k}{i} \sum_{j=0}^l \binom{l}{j} w^{(i,j)} S^{(k-i,l-j)} \\ &= w^{(0,0)} S^{(k,l)} + \sum_{i=1}^k \binom{k}{i} w^{(i,0)} S^{(k-i,l)} + \sum_{j=1}^l \binom{l}{j} w^{(0,j)} S^{(k,l-j)} \\ &\quad + \sum_{i=1}^k \binom{k}{i} \sum_{j=1}^l \binom{l}{j} w^{(i,j)} S^{(k-i,l-j)} \end{aligned}$$

e segue che

$$\begin{aligned} S^{(k,l)} &= \frac{1}{w} \left(A^{(k,l)} - \sum_{i=1}^k \binom{k}{i} w^{(i,0)} S^{(k-i,l)} \right. \\ &\quad \left. - \sum_{j=1}^l \binom{l}{j} w^{(0,j)} S^{(k,l-j)} - \sum_{i=1}^k \binom{k}{i} \sum_{j=1}^l \binom{l}{j} w^{(i,j)} S^{(k-i,l-j)} \right) \quad (5.4.15) \end{aligned}$$

Dall'equazione (5.4.15) otteniamo

$$S_{uv} = \frac{A_{uv} - w_{uv}S - w_u S_v - w_v S_u}{w} \quad (5.4.16)$$

$$S_{uu} = \frac{A_{uu} - 2w_u S_u - w_{uu} S}{w} \quad (5.4.17)$$

$$S_{vv} = \frac{A_{vv} - 2w_v S_v - w_{vv} S}{w} \quad (5.4.18)$$

Dalle equazioni (4.11.47),(5.4.14) e (5.4.15)

$$S_u(0,0) = \frac{p}{u_{p+1}} \frac{w_{1,0}}{w_{0,0}} (P_{1,0} - P_{0,0}) \quad (5.4.19)$$

$$S_v(0,0) = \frac{q}{v_{q+1}} \frac{w_{0,1}}{w_{0,0}} (P_{0,1} - P_{0,0}) \quad (5.4.20)$$

$$S_{uv}(0,0) = \frac{pq}{w_{0,0} u_{p+1} v_{q+1}} \left(w_{1,1} P_{1,1} - \frac{w_{1,0} w_{0,1}}{w_{0,0}} (P_{1,0} + P_{0,1}) + \left(\frac{2w_{1,0} w_{0,1}}{w_{0,0}} - w_{1,1} \right) P_{0,0} \right) \quad (5.4.21)$$

5.5 Alcuni Esempi di Utilizzo delle NURBS

Vorrei mostrare nelle sezioni seguenti alcune delle formule matematiche utilizzate per creare una figura e per modificarne alcuni aspetti. In questo caso si vede come costruire un cilindro e come muovere un punto di controllo o modificare uno dei pesi per cambiare forma a tale cilindro.

5.5.1 Come Costruire una Rappresentazione NURBS di un Cilindro

Sia W un vettore di lunghezza unitaria e $C(u) = \sum_{i=0}^n R_{i,p}(u)P_i$ sia una curva NURBS di grado p , sul vettore dei nodi U , con pesi w_i . Voglio rappresentare un cilindro generale, $S(u,v)$, ottenuto spostando $C(u)$ e distanza d lungo W . Denotando il parametro per la direzione di spostamento con v , $0 \leq v \leq 1$, chiaramente $S(u,v)$ deve soddisfare due condizioni:

- per \bar{u} fissato, $S(\bar{u}, v)$ é un segmento rettilineo da $C(\bar{u})$ a $C(\bar{u}) + dW$;
- per \bar{v} fissato

$$S(u, \bar{v}) = C(u) + \bar{v}dW = \sum_{i=0}^n R_{i,p}(P_i + \bar{v}dW)$$

(proprietá di invarianza traslazionale).

La rappresentazione desiderata é

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^1 R_{i,p;j,1}(u, v)P_{i,j} \quad (5.5.22)$$

sui vettori dei nodi U e V , dove $V=0,0,1,1$ e U é il vettore dei nodi di $C(u)$. I punti di controllo sono dati da $P_{i,0} = P_i$ e $P_{i,1} = P_i + dW$ e i pesi sono $w_{i,0} = w_{i,1} = w_i$. Un **cilindro regolare circolare** si ottiene traslando il cerchio NURBS a distanza d lungo un vettore normale al piano del cerchio. Un **arco quadratico razionale di Bézier** ha la forma

$$\begin{aligned} C(u) &= \frac{(1-u)^2 w_0 P_0 + 2u(1-u)w_1 P_1 + u^2 w_2 P_2}{(1-u)^2 w_0 + 2u(1-u)w_1 + u^2 w_2} \\ &= R_{0,2}(u)P_0 + R_{1,2}(u)P_1 + R_{2,2}P_2 \end{aligned} \quad (5.5.23)$$

e ad esempio una parametrizzazione di arco circolare puó essere

$$C(u) = \left(\frac{1-u^2}{1+u^2}, \frac{2u}{1+u^2} \right) \quad (5.5.24)$$

su $u \in [0, 1]$. Questo arco é centrato nell'origine, ha raggio 1 e percorre un angolo di 90° . Unendo quattro di questi archi insieme, usando nodi doppi, si ottiene il cerchio intero.

5.5.2 Riposizionare un Punto di Controllo

Sia $C(u) = \sum_{i=0}^n R_{i,p}(u)P_i$ una curva B-spline razionale o non razionale. Vogliamo riposizionare un arbitrario punto di controllo, P_k , $0 \leq k \leq n$. Indichiamo con \hat{P}_k il nuovo punto di controllo traslato, il vettore di traslazione con $V = \hat{P}_k - P_k$. Allora la nuova curva, $\hat{C}(u)$, é data da

$$\hat{C}(u) = R_{0,p}(u)P_0 + \dots + R_{k,p}(u)(P_k + V) + \dots + R_{n,p}(u)P_n = C(u) + R_{k,p}(u)V \quad (5.5.25)$$

L'equazione (5.5.25) esprime una traslazione funzionale di tutti i punti della curva, $C(u)$, per cui $u \in [u_k, u_{k+p+1})$; tutti i punti fuori da questo intervallo rimangono invariati. La traslazione massima avviene al massimo della funzione $R_{k,p}(u)$. Se supponiamo fissato \bar{u} con $\bar{u} \in [u_k, u_{k+p+1})$ e $P = C(\bar{u})$, se vogliamo muovere P a distanza d nella direzione di V riposizionando P_k , dobbiamo determinare α tale che

$$\hat{P}_k = P_k + \alpha V \quad (5.5.26)$$

provochi la traslazione voluta. Sia $\hat{P} = \hat{C}(\bar{u})$, chiaramente

$$\left| \hat{P} - P \right| = d = \alpha |V| R_{k,p}(\bar{u}) \quad (5.5.27)$$

da cui

$$\alpha = \frac{d}{|V| R_{k,p}(\bar{u})} \quad (5.5.28)$$

Dal punto di vista grafico, ci sono due applicazioni per il riposizionamento di un punto di controllo:

- nel primo caso si può specificare il vettore di direzione e la distanza e selezionare il punto sul poligono di controllo; se il punto scelto é un punto di controllo esistente il sistema semplicemente lo trasla e ridisegna la curva, altrimenti il sistema trasforma il punto in un punto di controllo attraverso l'inverso dell'inserimento di un nodo (si veda [8] cap. V par. 5.2) quindi lo trasla e ridisegna la curva;

- nel secondo caso si può scegliere un punto della curva, P (da cui il sistema ricava \bar{u}), quindi scegliere il punto \hat{P} in cui P sarà spostato (da cui il sistema ricava d e V); al sistema non resta che scegliere un punto di controllo P_k da traslare per produrre il movimento desiderato.

5.5.3 Modificare un Peso

Sia $C(u) = \sum_{i=0}^n R_{i,p}(u)P_i$ una curva NURBS e sia \bar{u} un parametro fissato tale che $\bar{u} \in [u_k, u_{k+p+1})$. Come ho già detto nel par. 5.1 per le curve e nel par. 5.3 per le superfici, l'effetto della modificazione di un peso w_k è di muovere il punto $C(\bar{u})$ lungo una retta definita da P_k e dal punto origine, $C(\bar{u})$. Quindi, mentre il riposizionamento di un punto di controllo ha l'effetto di una traslazione su una curva, la modifica di un peso ha un effetto di proiezione prospettica, dove tutti i punti della curva nel dominio coinvolto sono tirati (spinti) lungo rette che si incontrano nel punto di controllo corrispondente al peso modificato. Si dimostra ciò derivando l'equazione di $C(u)$ rispetto a w_k , cioè che un punto arbitrario della curva, $C(u)$, si muove lungo una retta in direzione di P_k quando w_k aumenta di valore. Per la dimostrazione e per un approccio geometrico vedere [8] a pagg. 520-526. Ovviamente il discorso è simile per quanto riguarda la modifica di un peso in una superficie. Altri strumenti di modifica di curve e superfici si trovano ai paragrafi 11.4-11.5 del cap.XI di [8].

5.5.4 Aumentare il Grado di una Curva

Sia $C_p^w(u) = \sum_{i=0}^n N_{i,p}(u)P_i^w$ una curva NURBS di grado p sul vettore dei nodi U . Poiché $C_p^w(u)$ è una curva polinomiale a tratti, è possibile elevare il suo grado a $p+1$, cioè devono esistere punti di controllo Q_i^w e un vettore dei nodi \hat{U} tali che

$$C_p^w(u) = C_{p+1}^w(u) = \sum_{i=0}^{\hat{n}} N_{i,p+1}(u)Q_i^w \quad (5.5.29)$$

$C_{p+1}^w(u)$ e $C_p^w(u)$ sono la stessa curva, sia geometricamente che parametricamente, $C_{p+1}^w(u)$ è semplicemente $C_p^w(u)$ incorporata in uno spazio di dimensione più elevata. 'Aumento del grado' si riferisce al processo (algoritmo) per calcolare gli sconosciuti Q_i^w e \hat{U} . In

particolare se C_1, \dots, C_n , ($n \geq 2$) é una sequenza di curve NURBS con la proprietá che il punto finale di C_i coincide con il punto iniziale di C_{i+1} , allora le curve possono essere combinate in una curva NURBS singola. Il primo passo é di elevare il grado delle curve allo stesso valore. Per i dettagli si veda [8] pag. 193 e seguenti.

5.6 Creazione di una Figura con Maya

Le superfici NURBS in Maya possono essere create in vari modi, partendo da una curva e facendola ruotare intorno ad un asse ('revolve'), disegnando piú curve e creando una superficie che le congiunge ('lofting') oppure utilizzando le superfici giá inserite (primitive), ad esempio un cilindro. Con il comando *Create/NURBS* ho creato un cilindro di raggio 1 cm con 8 isocurve verticali e 13 orizzontali alto 9 cm totali (vedi par. 5.5.1).

L'ho traslato di 4,5 cm lungo l'asse Y perché fosse appoggiato sul piano XZ. La traslazione é una trasformazione affine $F(X) = X + v$, dove X mi da le coordinate di un punto della curva, e v é un vettore. Le NURBS sono invarianti per trasformazioni affini.

Invarianza affine: una trasformazione affine é applicata alla curva applicandola ai punti di controllo. Sia r un punto in \mathcal{E}^3 (spazio Euclideo tridimensionale). Una *trasformazione affine*, indicata con Φ , porta \mathcal{E}^3 in \mathcal{E}^3 e ha la forma

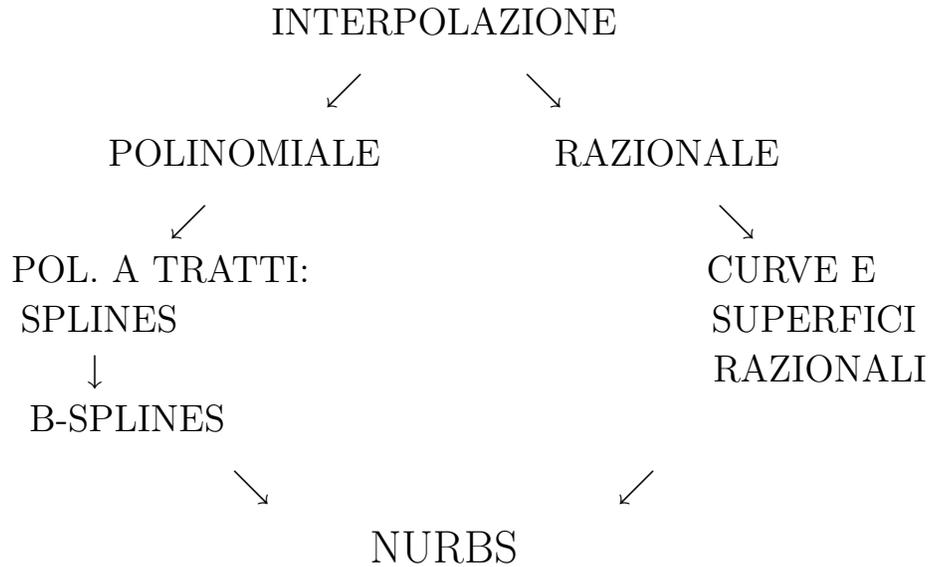
$$\Phi(r) = Ar + v$$

dove A é una matrice 3×3 e v é un vettore. Le trasformazioni affini includono traslazioni, rotazioni, scale e tagli ('shears'). La proprietá di invarianza affine per le curve B-splines segue dalla proprietá della partizione dell'unitá di $N_{i,p}(u)$. Quindi, sia $r = \sum \alpha_i p_i$, dove $p_i \in \mathcal{E}^3$ e $\sum \alpha_i = 1$. Allora

$$\begin{aligned} \Phi(r) &= \Phi\left(\sum \alpha_i p_i\right) = A\left(\sum \alpha_i p_i\right) + v = \sum \alpha_i A p_i + \sum \alpha_i v \\ &= \sum \alpha_i (A p_i + v) = \sum \alpha_i \Phi(p_i) \end{aligned}$$

Ho quindi suddiviso la superficie del cilindro in 10 parti, separandole lungo le isocurve 0,2,4,6 verticali e la 6 e la 8 orizzontali. Ho tagliato la porzione di superficie tra le isocurve 6 e 8 orizzontali e le 0 e 2 verticali. Ho creato un altro cilindro di raggio 0,5 con 8 isocurve verticali e 7 orizzontali, alto 6 cm totali, ruotato di 65 lungo Z. La rotazione é una trasformazione affine del tipo $F(X) = LX + v$, dove X e v sono gli stessi usati nella traslazione, mentre L é la matrice identitá in cui é stata sostituita una sottomatrice di ordine 2 con una matrice del tipo $A = \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$. Quindi l'ho traslato di 2,5 lungo X e 3,8 lungo Y. Ho suddiviso la superficie in 4 parti divise lungo le isocurve. Selezionando una superficie alla volta e in particolare i punti di controllo sul confine verso il primo cilindro, li ho spostati fino a creare un collegamento con la superficie dello stesso

colore sul primo cilindro (algoritmo di riposizionamento di un punto e modifica di un peso par. 5.5.2 e 5.5.3). Eventualmente si possono sistemare i punti di controllo con piú precisione inserendo le loro coordinate direttamente in una tabella (il programma provvede e collocarli nella giusta posizione). Infine esiste un comando (*Stitch*) per, diciamo, ‘cucire’ i punti di controllo insieme in modo da eliminare eventuali fessure (algoritmo di aumento del grado di una curva par. 5.5.4).



Bibliografia

- [1] Bevilacqua, Bini, Capovani, Menchi, METODI NUMERICI, Zanichelli, 1992
- [2] Galligani, ELEMENTI DI ANALISI NUMERICA, Calderini, 1986
- [3] Quarteroni, Sacco, Saleri, MATEMATICA NUMERICA, Springer, 2000
- [4] Stoer, Burlish, INTRODUCTION TO NUMERICAL ANALYSIS, Springer, 1992
- [5] Carl de Boor, A PRACTICAL GUIDE TO SPLINES, Springer, 1978
- [6] Stefano De Marchi, FUNZIONI SPLINES UNIVARIATE, Forum, 1999
- [7] G.Farin, D. Hansford, THE GEOMETRY TOOLBOX, Peters, 1998
- [8] L.Piegl, W.Tiller, THE NURBS BOOK, Springer, 1997
- [9] Sito Scilab: <http://scilabsoft.inria.fr/>
- [10] J.Kundert, P.L.Gibbs, MAYA 5 - GUIDA COMPLETA, Apogeo, 2004