Laurea specialistica

Splines in **R**

Data una sequenza di punti $x_0 < x_1 < \ldots < x_n$ in **R** (*nodi*), e una corrispondente sequenza di valori $y_0, y_1, \ldots, y_n$, si dice *spline* una funzione interpolatrice che nei punti x_i assume il valore y_i e che negli intervalli $]x_i, x_{i+1}[$ è un polinomio di grado *k* e verifica opportune proprietà di regolarità agli estremi degli intervalli (ad esempio, si può richiedere la continuità della funzione e delle sue derivate fino alla (*k*-1)-esima).
Le spline più usate sono quelle di grado 1 e 3. Il nome spline in origine indicava delle travi flessibili che si curvavano per adattarsi a determinati vincoli (ad esempio nella costruzione delle chiglie delle navi). Analogamente, queste funzioni interpolanti hanno la proprietà di adattarsi in maniera ottimale (secondo criteri ben definiti) ai vincoli imposti dai valori assegnati.

Spline lineari

Una spline lineare è continua in $[x_0, x_n]$ e lineare negli intervalli aperti $]x_i, x_{i+1}[$. Ovviamente i punti di ascissa x_i sono in generale punti angolosi.
Le spline lineari con nodi in $x_0, \ldots, x_n$

- sono uno spazio lineare (ossia le loro combinazioni lineari sono ancora dello stesso tipo);
- sono descritte da $2n$ coefficienti (2 per ogni intervallo $]x_i, x_{i+1}[$)
- sono soggette a *n*-1 vincoli lineari (continuità nei punti $x_i \mid i = 1, \ldots, n-1$: $a_i x_i + b_i = a_{i+1} x_i + b_{i+1}$)

Quindi sono uno spazio lineare (*n*+1)-dim. Ogni spline lineare è univocamente determinata dai suoi valori negli *n*+1 punti $x_0, \ldots, x_n$. Una possibile scelta di funzioni di base per questo spazio è $\{\varphi_j(x) : \varphi_j(x_k) = \delta_{jk}\ ,\ j,k = 0, \ldots, n\}$; $\varphi_j(x)$ continua e lineare a tratti. Ovviamente $\varphi_j(x)$ è identicamente nulla fuori dell'intervallo $[x_{j-1}, x_{j+1}]$. Segue che una qualunque spline lineare con nodi in $x_0, \ldots, x_n$ si può esprimere come $s(x) = \sum_{k=0}^{n} y_k \varphi_k(x)$, dove y_k sono i valori assunti da s nei nodi.
La derivata di una spline lineare è costante a tratti e non è definita nei punti x_j ; essa è a quadrato integrabile in $[x_0, x_n]$.

TEOR: Fra tutte le funzioni continue e con derivata a quadrato integrabile in $[x_0, x_n]$, tali che $f(x_j) = y_j$, la spline lineare è quella che minimizza $\int_{x_0}^{x_n} | f'(x) |^2 \, dx$

DIM: sia $s(x)$ la spline lineare, $f(x)$ una qualsiasi altra funzione interpolante: $s(x_j) = f(x_j) = y_j$. Sia $g = f - s \Rightarrow g(x_j) = 0$. Allora $\int_{x_{j-1}}^{x_j} | f'|^2 = \int_{x_{j-1}}^{x_j} | s'|^2 + \int_{x_{j-1}}^{x_j} | g'|^2 + 2\int_{x_{j-1}}^{x_j} s'g'$.

Essendo s' costante nell'intervallo, si ha $\int_{x_{j-1}}^{x_j} s'g' = \text{cost.}(g(x_j) - g(x_{j-1})) = 0$. Quindi, al variare di f fra le funzioni interpolanti, il minimo di $\int_{x_{j-1}}^{x_j} | f'|^2$ per ogni j , e quindi il minimo di $\int_{x_0}^{x_n} | f'|^2$, si ha quando $g(x)=\text{cost.}$, dove necessariamente cost.=0 , dato che $g(x_j)=0$. Di conseguenza, si ha il minimo per $f=s$.

Spline cubiche

Una spline cubica è un polinomio di terzo grado in ciascuno degli intervalli aperti $]x_i, x_{i+1}[$, e nei nodi è continua con derivate prima e seconda continue.
Le spline cubiche con nodi in $x_0,\dots,x_n$ sono uno spazio lineare, e sono descritte da $4n$ coefficienti (4 per ciascun polinomio di terzo grado). Per individuare univocamente una spline cubica occorrono quindi $4n$ condizioni. Di queste, $2n$ derivano dall'imporre che ciascun polinomio di terzo grado assuma i valori assegnati y_i, y_{i+1} agli estremi dell'intervallo $]x_i, x_{i+1}[$; altre $2(n\text{-}1)$ condizioni derivano dall'imporre la continuità delle derivate prime e seconde nei nodi interni. Per avere univocità occorre imporre altre 2 condizioni. Solitamente si impongono condizioni al contorno, scegliendo fra 2 possibili opzioni:

i) $s''(x_0)=s''(x_n)=0$

ii) $s'(x_0)=a$, $s'(x_n)=b$ (a e b arbitrari).

Le spline cubiche con nodi in $x_0,\dots,x_n$ che verificano la condizione i) costituiscono uno spazio lineare (n+1)-dim.; è possibile scegliere come funzioni di base le spline cubiche $\varphi_j(x)$ che, oltre a verificare la i), assumono nei nodi i valori $\varphi_j(x_k)=\delta_{jk}$. Allora, come nel caso lineare, le spline cubiche con gli stessi nodi hanno l'espressione $s(x)=\sum_{k=0}^{n} y_k\varphi_k(x)$, dove y_k sono i valori assunti da s nei nodi. Si noti che, a differenza del caso lineare, le $\varphi_j(x)$ non sono identicamente nulle in alcun intervallo. Se invece non si impone la condizione i), lo spazio è (n+3)-dim.; come base si possono scegliere n+1 spline $\phi_j(x)$ con $\phi_j(x_k)=\delta_{jk}$ e in più con le condizioni al contorno $\phi_j'(x_0)=\phi_j'(x_n)=0$, e inoltre 2 funzioni ψ_0,ψ_n che si annullano in tutti i nodi, con $\psi_0'(x_0)=\psi_n'(x_n)=1$, $\psi_0'(x_n)=\psi_n'(x_0)=0$. In questo caso, una spline con gli stessi nodi che verifica la condizione ii) si esprime come $s(x)=\sum_{k=0}^{n} y_k\phi_k(x)+a\psi_0(x)+b\psi_1(x)$.

E' possibile costruire una base di spline che si annullano identicamente al di fuori di un intervallo finito. Queste funzioni possono essere costruite prendendo un intervallo unione di 4 intervalli consecutivi delimitati da nodi, e imponendo agli estremi le condizioni al contorno $s(x_j)=s'(x_j)=s''(x_j)=s(x_{j+4})=s'(x_{j+4})=s''(x_{j+4})=0$. Questa scelta è particolarmente comoda se i nodi sono equidistanti, perché in questo caso le funzioni di base si possono ottenere una dall'altra per traslazione. Ad esempio, scegliendo come nodi -2,-1,0,1,2 una funzione così fatta ha la forma $s(x)=\begin{cases}(x+2)^3 & , \quad -2\le x\le -1\\ -3x^3-6x^2+4 & , \quad -1<x\le 0\\ s(-x) & , \quad 0<x\le 2\end{cases}$, che è simmetrica rispetto a x=0. Se si ha un intervallo con n+1 nodi equidistanti, bisogna avere n+3 funzioni di base; oltre alle funzioni $s_j(x),\ j=0,\dots n$ simmetriche rispetto ai nodi bisogna aggiungere 2 funzioni, $s_{-1}(x)$ simmetrica rispetto a $x_{-1}<x_0$ e $s_{n+1}(x)$ simmetrica rispetto a $x_{n+1}>x_n$.

Analogamente a quanto fatto per le spline lineari, si può provare che, fra tutte le funzioni derivabili fino al secondo ordine e con derivata seconda a quadrato integrabile in $[x_0, x_n]$, con $f(x_k) = y_k$, e con date condizioni al contorno, la spline cubica è quella che rende minimo $\int_{x_0}^{x_n} | f''(x)|^2 \, dx$.

DIM: Sia $g = f - s$, dove s è la spline, f qualsiasi altra funzione interpolante. Allora $\int_{x_{j-1}}^{x_j} | f''|^2 = \int_{x_{j-1}}^{x_j} | s''|^2 + \int_{x_{j-1}}^{x_j} | g''|^2 + 2\int_{x_{j-1}}^{x_j} s'' g''$. Integrando per parti, si ottiene $\int_{x_{j-1}}^{x_j} s'' g'' = -\int_{x_{j-1}}^{x_j} s''' g' + s'' g' \big|_{x_{j-1}}^{x_j}$.

Essendo s''' costante e $g(x_{j-1}) = g(x_j) = 0$, il primo addendo è nullo; sommando su tutti gli intervalli si ottiene $\int_{x_0}^{x_n} s'' g'' = s''(x_n) g'(x_n) - s''(x_0) g'(x_0) = 0$, dato che, o è verificata la condizione al contorno i), oppure, se è verificata la ii) con gli stessi valori per f e s, $g'(x_0) = g'(x_n) = 0$. Quindi il minimo si realizza per $\int_{x_0}^{x_n} | g''|^2 = 0$, che è certamente verificata per $f = s$.

NOTA: e' proprio a questa proprietà, che potrebbe essere definita di minimizzazione globale della curvatura, che è dovuto il termine *spline*, con riferimento al significato originario sopra menzionato.

Viene ora delineata una procedura per il calcolo di un spline cubica, nel caso semplice, ma facilmente generalizzabile, che i nodi siano equidistanti, con $x_j - x_{j-1} = 1$.

Sia $M_j = s''(x_j)$. Allora in $[x_{j-1}, x_j]$ $s''(x) = M_{j-1}(x_j - x) + M_j (x - x_{j-1})$, dato che s'' è lineare negli intervalli fra i nodi, e, di conseguenza

$$
\begin{aligned}
s'(x) &= -\tfrac{1}{2} M_{j-1}(x_j - x)^2 + \tfrac{1}{2} M_j (x - x_{j-1})^2 + A_j \\
s(x) &= \tfrac{1}{6} M_{j-1}(x_j - x)^3 + \tfrac{1}{6} M_j (x - x_{j-1})^3 + A_j x + B_j
\end{aligned}
.
$$

Imponendo nei nodi i valori dati, si ottiene

$$
\begin{aligned}
y_j &= \tfrac{1}{6} M_j + A_j x_j + B_j \\
y_{j-1} &= \tfrac{1}{6} M_{j-1} + A_j x_{j-1} + B_j
\end{aligned}
\Rightarrow A_j = y_j - y_{j-1} - \tfrac{1}{6}(M_j - M_{j-1})
$$

; noto A_j, si calcola immediatamente B_j.

Ora, imponendo la continuità delle derivate prime nei nodi, si ottiene

$$
\begin{aligned}
&\tfrac{1}{2} M_j + y_j - y_{j-1} - \tfrac{1}{6}(M_j - M_{j-1}) = -\tfrac{1}{2} M_j + y_{j+1} - y_j - \tfrac{1}{6}(M_{j+1} - M_j) \\
&\quad \Rightarrow M_{j-1} + 4M_j + M_{j+1} = 6(y_{j+1} - 2y_j + y_{j-1}) \equiv d_j \quad \text{(noto)}
\end{aligned}
.
$$

Si ottiene quindi un sistema lineare di *n-1* equazioni in *n+1* incognite, la cui forma matriciale è

$$
\begin{pmatrix}
1 & 4 & 1 & 0 & \vdots & \vdots & 0 \\
0 & 1 & 4 & 1 & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & 0 & 1 & 4 & 1
\end{pmatrix}
\begin{pmatrix}
M_0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ M_n
\end{pmatrix}
=
\begin{pmatrix}
d_1 \\ \vdots \\ \vdots \\ \vdots \\ d_{n-1}
\end{pmatrix}.
$$

Evidentemente questo sistema è sotto-determinato. Per renderlo univocamente risolubile, occorre introdurre le condizioni al contorno. Nel caso i) basta imporre $M_0 = M_n = 0$, eliminandole come incognite, ed eliminando anche la prima e l'ultima colonna della matrice; nel caso ii) si introducono i valori noti di $s'(x_0)$ e $s'(x_n)$, aggiungendo 2 equazioni della forma $\begin{array}{c} 2M_0 + M_1 = d_0 \\ M_{n-1} + 2M_n = d_n \end{array}$ (dove, utilizzando l'espressione di $s'(x)$, si ottiene $d_0 = 6(y_1 - y_0 - s'(x_0))$, $d_n = 6(s'(x_n) - y_n + y_{n-1})$), che si traducono dell'aggiunta di una riga in cima e una in fondo alla matrice. Una volta ricavati M_j, $s(x)$ è completamente determinata.

La regolarità della struttura della matrice e la presenza di molti zeri la rendono facilmente trattabile dal punto di vista numerico anche quando le dimensioni sono grandi.
NOTA: E' facile verificare che, nel caso particolare in cui $y_j = 0$ in tutti i nodi, ed eventualmente, nel caso ii), $a = b = 0$, la soluzione è $s(x) \equiv 0$.

Interpolazione approssimata con spline

Data una sequenza di valori y_k associati a punti x_k, non necessariamente si richiede che la funzione interpolante $y(x)$ verifichi esattamente $y(x_k) = y_k$. Infatti può darsi che i valori y_k siano il risultato di misure affette da errore, e che proprio questi errori comportino un andamento ingiustificatamente oscillante per la funzione interpolante. E' quindi ragionevole scegliere una funzione interpolante più liscia, accontentandosi che i valori dati siano assunti solo approssimativamente.
In generale la funzione interpolante viene scelta in uno spazio lineare generato da una certa base di funzioni, ed i coefficienti vengono determinati ad esempio con il criterio dei minimi quadrati: $\sum_k | y_k - \sum_j c_j \phi_j (x_k) |^2 = \min$.
Si può per esempio scegliere una base di splines, con nodi non necessariamente coincidenti con i punti x_k. Scegliendo i nodi equidistanti e spline di base nulle fuori di un determinato intervallo si ottengono due vantaggi: le spline di base si ottengono l'una dall'altra per traslazione, e ciascun valore y_k ha effetto soltanto sui coefficienti delle spline non nulle nel punto corrispondente x_k, che sono in piccolo numero. Se però i punti x_k sono sparsi irregolarmente, con densità molto diverse in diversi intervalli, si possono verificare due tipi opposti di inconvenienti: se i supporti delle spline sono piccoli, può accadere che qualcuno di essi non contenga punti x_k, e di conseguenza il coefficiente della spline corrispondente è indeterminato; se viceversa sono grandi, le spline variano lentamente, e possono quindi dare approssimazioni inaccettabili dove i valori y_k variano rapidamente in piccoli intervalli.

Una possibile strada per ovviare a questo inconveniente consiste nell'usare supporti di spline non omogenei, ma più piccoli là dove si ha maggiore densità di punti. Si può ad esempio scegliere per i nodi inizialmente un determinato passo piuttosto grande, poi dimezzarlo via via là dove l'approssimazione risulta troppo grossolana, finché non si ottiene un'approssimazione accettabile. Questo procedimento, detto *multirisoluzione*, porta a costruire una base di spline della forma $\phi_{jk}(x) = \phi(2^j x - k)$. Si noti che gli spazi generati dalle spline con supporti più ampi sono sottospazi di quelli generati dalle spline con supporti più piccoli.

Un procedimento alternativo, detto *regolarizzazione di Tychonov*, consiste nel modificare il principio dei minimi quadrati imponendo nell'espressione da minimizzare dei vincoli aggiuntivi sulla funzione cercata, in modo da ottenere unicità della soluzione nel caso che la semplice applicazione dei minimi quadrati porti all'indeterminazione di qualche coefficiente: $\sum_k | y_k - f(x_k) |^2 + \mu K(f) = \min$, dove $f = \sum c_j \phi_j$, K è un opportuno funzionale (ad esempio, per le spline lineari si sceglie $K(f)=\int |f'(x)|^2\,dx$, per quelle cubiche $K(f)=\int |f''(x)|^2\,dx$;

μ è un peso che regola l'influenza della regolarizzazione rispetto alla ricerca di una buona approssimazione. Se μ è piccolo (ad esempio $\mu=0.1$), la presenza di $K(f)$ nella quantità da minimizzare è poco rilevante e l'algoritmo cerca soprattutto di ottenere una buona approssimazione; se invece μ è grande (ad esempio $\mu=10$), nell'algoritmo prevale la ricerca di una buona regolarizzazione ($K(f)$ piccolo) piuttosto che di una buona approssimazione.

ESEMPIO: Si consideri l'intervallo $[-1,1]$ suddiviso in 4 intervalli di uguale ampiezza, con nodi in –1 , -1/2 , 0 , 1/2 , 1. Una base di spline lineari con questi nodi è data da

$$\varphi_j(x)=\begin{cases}1+2(x-\frac{j}{2}) & \frac{j\text{-}1}{2}\le x<\frac{j}{2}\\ 1-2(x-\frac{j}{2}) & \frac{j}{2}\le x\le\frac{j+1}{2}\\ 0 & x\notin\left[\frac{j\text{-}1}{2},\frac{j+1}{2}\right]\end{cases}\qquad ,\qquad j=-2,-1,0,1,2$$

Una spline lineare ha la seguente espressione generale:

$$s(x)=\begin{cases}c_1\big(1-2(x+1)\big)+c_2\big(1+2(x+\frac{1}{2})\big)=-c_1+2c_2-2(c_1-c_2)x & -1\le x<-\frac{1}{2}\\ c_2\big(1-2(x+\frac{1}{2})\big)+c_3\big(1+2x\big)=c_3-2(c_2-c_3)x & -\frac{1}{2}\le x<0\\ c_3\big(1-2x\big)+c_4\big(1+2(x-\frac{1}{2})\big)=c_3-2(c_3-c_4)x & 0\le x<\frac{1}{2}\\ c_4\big(1-2(x-\frac{1}{2})\big)+c_5\big(1+2(x-1)\big)=2c_4-c_5-2(c_4-c_5)x & \frac{1}{2}\le x\le 1\end{cases}$$

Di conseguenza $$s'(x)=\begin{cases}-2(c_1-c_2) & -1\le x<-\frac{1}{2}\\ -2(c_2-c_3) & -\frac{1}{2}\le x<0\\ -2(c_3-c_4) & 0\le x<\frac{1}{2}\\ -2(c_4-c_5) & \frac{1}{2}\le x\le 1\end{cases}$$

e $$\int_{-1}^{1} s'^2\,dx=2(c_1-c_2)^2+2(c_2-c_3)^2+2(c_3-c_4)^2+2(c_4-c_5)^2\equiv 2\mathbf{c}^T\mathbf{Bc}$$

dove $$\mathbf{B}=\begin{pmatrix}1&-2&0&0&0\\0&2&-2&0&0\\0&0&2&-2&0\\0&0&0&2&-2\\0&0&0&0&1\end{pmatrix}$$

Dato un insieme di valori y_i campionati in corrispondenza di punti $x_i\in[-1,1]$, per applicare l'interpolazione approssimata con regolarizzazione di Tychonov si tenga innanzitutto presente che la spline interpolante ha la forma **y=Ac** , dove i coefficienti che compaiono nelle diverse righe della matrice **A** dipendono dal particolare intervallo in cui cade il corrispondente punto x_i . Ad esempio, se $x_k = 1/3$, che cade nell'intervallo [0,1/2], si ha $y_k = c_3 - \frac{2}{3}(c_3 - c_4) = \frac{1}{3}c_3 + \frac{2}{3}c_4$.

L'uso del metodo dei minimi quadrati porta alla minimizzazione dell'espressione $(\mathbf{y} - \mathbf{Ac})^T (\mathbf{y} - \mathbf{Ac}) + 2\alpha \mathbf{c}^T \mathbf{Bc}$, da cui si ottiene

$$- 2\mathbf{A}^T (\mathbf{y} - \mathbf{Ac}) + 2\alpha(\mathbf{B} + \mathbf{B}^T)\mathbf{c} = 0 \;\Rightarrow\; \left(\mathbf{A}^T \mathbf{A} + \alpha(\mathbf{B} + \mathbf{B}^T)\right)\mathbf{c} = \mathbf{A}^T \mathbf{y}$$

Splines in $\mathbf{R}^2$

Dal punto di vista applicativo le spline in dim.2 sono di grande importanza nelle discipline del rilevamento perché consentono la modellizzazione di superfici lisce a partire da punti sparsi. Dal punto di vista matematico, la problematiche sono simili a quelle viste nel caso 1-dim., ovviamente con maggiori complicazioni formali. Nel seguito viene fatto un breve cenno agli aspetti principali.

Dato uno spazio lineare di spline 1-dim. con base $\{S_j(x)\}$, si può definire nel piano xy uno spazio lineare di spline 2-dim. introducendo come base $\{S_{jk}(x, y) = S_j(x)S_k(y)\}$. Se le spline 1-dim. sono cubiche, le corrispondenti spline 2-dim. sono dette *bicubiche.* I nodi costituiscono una griglia rettangolare $\{(x_j, y_k) \mid j = 0,\cdots,m \;;\; k = 0,\cdots,n\}$. Le spline bicubiche sono C^∞ all'interno di ciascuno dei rettangoli del reticolo; sui lati di tali rettangoli è in ogni caso garantita la continuità di tutte le derivate parziali prime e seconde, e inoltre di $S'''_{xxy}, S'''_{xyy}, S^{(iv)}_{xxyy}$.
Per avere la determinazione univoca di una spline interpolante, è necessario fornire, oltre ai valori nei nodi, anche delle condizioni al contorno. Quelle più usate sono:

i) valori dati di $S'_x(x_0, y_k), S'_x(x_m, y_k), S'_y(x_j, y_0), S'_y(x_j, y_n) \mid j = 0,\cdots,m \;;\; k = 0,\cdots,n$; $S''_{xy}(x_0, y_0), S''_{xy}(x_m, y_0), S''_{xy}(x_0, y_n), S''_{xy}(x_m, y_n)$;

ii) $S''_{xx}(x_0, y_k) = S''_{xx}(x_m, y_k) = 0 ; S''_{yy}(x_j, y_0) = S''_{yy}(x_j, y_n) = 0 \mid j = 0,\cdots,m \;;\; k = 0,\cdots,n$; $S^{(iv)}_{xxyy}(x_0, y_0) = S^{(iv)}_{xxyy}(x_0, y_n) = S^{(iv)}_{xxyy}(x_m, y_0) = S^{(iv)}_{xxyy}(x_m, y_n) = 0$.

In totale, il numero di condizioni da imporre è $2(m+1) + 2(n+1) + 4 = 2m + 2n + 8$, uguale alla differenza fra il numero delle funzioni di base, $(m+3)(n+3)$, e il numero dei valori imposti ai nodi, $(m+1)(n+1)$.

La formulazione dei problemi di interpolazione approssimata con il metodo dei minimi quadrati è del tutto analoga al caso 1-dim.

Appendice - Metodo alternativo per la definizione e la costruzione delle spline (B-splines)

Data una sequenza di nodi $u_0,\cdots,u_m$, $u_i \le u_{i+1}$ (non necessariamente tutti distinti), si definisce

$$N_{i,0}(u) = \chi_{[u_i,u_{i+1}[}(u)$$

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

Si intende che, se $u_i = u_{i+1}$, la corrispondente $N_{i,0}$ non è definita, e quindi viene omessa nella formula di ricorrenza.

Sono verificate le seguenti proprietà:

- $N_{i,p}(u) = 0$ per $u \notin \lfloor u_i, u_{i+p+1} \lfloor$
- in ogni $[u_i, u_{i+1}[$ sono non nulle al più p+1 fra le $N_{i,p}$.
- $N_{i,p}(u) \ge 0$
- $\sum_i N_{i,p}(u) = 1$
- in un nodo di molteplicità k (ossia quando $u_i = u_{i+1} = \cdots = u_{i+k-1}$) $N_{i,p}(u)$ è continua con le sue derivate fino all'ordine p-k; inoltre, queste derivate sono nulle in u_i e u_{i+p+1} , dato che $N_{i,p}$ è identicamente nulla al di fuori di $\lfloor u_i, u_{i+p+1} \lfloor$.
- $N_{i,p}(u)$, $p \ne 0$ ha un solo punto di massimo;

NOTA: Le $N_{i,p}$ sono polinomi di grado p in ogni intervallo fra 2 nodi consecutivi. Si assuma ora che i nodi siano tutti distinti. Allora una qualsiasi funzione che è un polinomio di grado p nell'intervallo fra 2 nodi consecutivi, che è continua con le sue derivate fino all'ordine p-1 in tutti i nodi e che è identicamente nulla al di fuori di un intervallo $\lfloor u_i, u_{i+p+1} \lfloor$ è definita da $(p+1)^2$ coefficienti (p+1 in ognuno dei p+1 intervalli che costituiscono $\lfloor u_i, u_{i+p+1} \lfloor$) e deve verificare $p(p+2) = (p+1)^2 - 1$ condizioni (ossia p condizioni – la continuità della funzione e delle sue derivate fino all'ordine p-1 – in ciascuno dei p+2 punti $u_i,\cdots,u_{i+p+1}$). Rimane quindi un grado di libertà, ossia la funzione deve coincidere con $N_{i,p}(u)$ a meno di una costante moltiplicativa.

ESERCIZI:

i) data la sequenza di nodi 0,0,0,0,1,2,3,4,4,4,4 , costruire $N_{0,3}, N_{1,3}, N_{2,3}, N_{3,3}$

ii) data la sequenza di nodi 0,0,0,0,1,1,2,3,4,4,4,4 , costruire $N_{0,3}, N_{1,3}, N_{2,3}, N_{3,3}, N_{4,3}$

Come si è visto, le spline cubiche costruite su m intervalli con estremi nei nodi $u_0,\cdots,u_m$ costituiscono uno spazio vettoriale di dimensione m+3. Un possibile modo di costruire una base consiste nell'introdurre una sequenza di nodi coincidenti con $u_0,\cdots,u_m$ in cui il primo e l'ultimo sono ripetuti p+1=4 volte (come nell'esercizio i)), e nel costruire su questa sequenza tutte le possibili $N_{i,3}$. Poiché ciascuna di queste funzioni è non identicamente nulla su 4 intervalli consecutivi (eventualmente di ampiezza nulla) delimitati dai nodi, si verifica che il loro numero è esattamente m+3 (la prima è non nulla nell'intervallo comprendente i nodi u_0, u_0, u_0, u_0, u_1, le altre si ottengono scalando i nodi via via di 1, fino all'ultima che comprende i nodi $u_{m-1}, u_m, u_m, u_m, u_m$).

I nodi interni con molteplicità maggiore di 1 vengono introdotti se si vuole costruire una funzione interpolante per cui si richiede in determinati punti un regolarità inferiore (ossia un più piccolo numero di derivate continue) a quella delle spline ordinarie.

Interpolazione di curve e superfici in forma parametrica

Finora le spline sono state esaminate come strumento per costruire funzioni interpolanti a partire da valori noti in un insieme discreto di punti (sulla retta o nel piano). Il risultato sono funzioni $y(x)$ o $z(x, y)$ che possono poi essere rappresentate graficamente nel piano xy o nello spazio 3-dim. xyz. Più in generale, è possibile utilizzare le spline per l'interpolazione di curve nel piano o nello spazio e di superfici nello spazio, che non necessariamente possono essere espresse globalmente come grafici di funzioni.

Per fissare le idee, si consideri nello spazio 3-dim. una sequenza di punti $\mathbf{r}_i \equiv (x_i, y_i, z_i)$, $i = 0, \cdots, m$, che si suppone di voler interpolare con un arco di curva espresso parametricamente: $\mathbf{r} = \mathbf{r}(u) \equiv (x(u), y(u), z(u))$. A tale scopo è necessario:

- stabilire l'intervallo di variabilità del parametro u : $[u_0, u_m]$, con $\mathbf{r}(u_0) = \mathbf{r}_0$, $\mathbf{r}(u_m) = \mathbf{r}_m$;
- fissare in questo intervallo i valori dei parametri corrispondenti ai nodi: $u_i : \mathbf{r}(u_i) = \mathbf{r}_i$.

A questo punto, ogni componente di $\mathbf{r}$ viene interpolata con una spline definita in $[u_0, u_m]$ con nodi in u_i . Poiché si possono usare le $N_{i,p}$ sopra definite come base per lo spazio delle spline di grado p , si avrà $\begin{array}{l} x(u) = \sum a_i N_{i,p}(u) \\ y(u) = \sum b_i N_{i,p}(u) \\ z(u) = \sum c_i N_{i,p}(u) \end{array}$, ossia $\mathbf{r}(u) = \sum \mathbf{v}_i N_{i,p}(u)$, dove $\mathbf{v}_i \equiv \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix}$ è un vettore, che corrisponde a un punto nello spazio 3-dim.

I punti $\mathbf{v}_i$ sono detti *punti di controllo*. Il fatto rilevante è che, in virtù delle proprietà $N_{i,p}(u) \geq 0$, $\sum_i N_{i,p}(u) = 1$, l'arco di curva $\mathbf{r}(u)$ è tutto contenuto nell'*inviluppo convesso* dei punti di controllo (l'inviluppo convesso di un insieme di punti è un poliedro convesso – ossia tale che, se contiene 2 punti, contiene anche tutto il segmento che li congiunge – con vertici in alcuni di tali punti, e gli altri al suo interno).

Poiché l'operazione sopra descritta non è univoca (la scelta di $[u_0, u_m]$ e degli u_i è in larga misura arbitraria) è possibile giocare su questi gradi di libertà per modellare la curva interpolante.

Un discorso analogo può essere fatto per una superficie nello spazio. In questo caso lo spazio parametrico è 2-dim., e una porzione di superficie è descritta dalle equazioni $\begin{array}{l} x = x(u, v) \\ y = y(u, v) \\ z = z(u, v) \end{array}$. Per fissare il dominio di variabilità dei parametri e la posizione dei nodi in questo dominio, si definiscono 2 sequenze $u_0, \cdots, u_m$, $v_0, \cdots, v_n$; il dominio è il rettangolo con vertici in

(u_0, v_0), (u_0, v_n), (u_m, v_0), (u_m, v_n) , e i nodi, ossia i punti di campionamento dei valori da interpolare, sono parametrizzati da (u_i, v_j). Come funzioni di base si possono scegliere i prodotti $N_{i,p}(u) N_{j,p}(v)$.