MODELLAZIONE DI OGGETTI

Luigi Mussio
Politecnico di Milano – DICA
Piazza L. da Vinci, 32 – 20133 Milano
Tel. 02–2399–6501, Fax. 02–2399–6602
e–mail luigi.mussio@polimi.it

RIASSUNTO

La modellazione di oggetti si occupa della ricostruzione di linee, superfici e corpi 3D. Talvolta le linee chiuse sono contorni di figure e le superfici chiuse sono superfici esterne di corpi 3D (per queste ultime, manca, in italiano, una parola unificante, fra le tante che identificano situazioni diverse). Il metodo degli elementi finiti e le funzioni spline costituiscono metodi classici per dette modellazioni. Inoltre con particolare riferimento ai casi più complessi di modellazione, metodi discreti di triangolazione o tetraedrazione rappresentano passi preliminari alla ricostruzione di linee e superfici mediante forme parametriche. Infine alcuni esempi mostrano modellazioni solide e ricostruzione di superfici, ed un'ulteriore parte conclusiva presenta le trasformazioni elementari piane e nello spazio 3D.

PARTE I – INTERPOLAZIONE CON FUNZIONI SPLINE[1]

[1] Il metodo degli elementi finiti è il più antico ed il più classico fra i metodi d'interpolazione, a supporto fisso, operanti nel dominio dei valori argomentali. Esso si basa sulle seguenti operazioni:

- scomposizione del dominio dei valori argomentali in intervalli (od intorni bi, tri – multidimensionali) finiti;
- interpolazione dei valori argomentali, compresi in ciascun intervallo (o intorno) finito, con la combinazione lineare di quattro (o sedici, sessantaquattro, oppure altra potenza di quattro con esponente uguale alla dimensione dell'intorno) opportune funzioni polinomiali di terzo grado;
- stima dei coefficienti delle combinazioni lineari, dei valori interpolati, corrispondenti alle osservazioni o meno, e calcolo della documentazione statistica di corredo.

Una funzione monodimensionale polinomiale di terzo grado ha tre zeri (o uno soltanto), come pure un massimo relativo, un minimo relativo ed un punto di flesso a tangente inclinata (o un unico punto di flesso a tangente orizzontale).
Una funzione semplice è data dal prodotto della bisettrice del $1°$ e $3°$ quadrante: $y = x$, con il prodotto notevole dell'algebra: $y = (x+1)(x-1) = x^2 - 1$, ovvero: $y = x(x^2 - 1)$. La sostituzione delle coordinate con nuove coordinate, così definite: $x = x / \sqrt{3}$ e $y = 2y / 3\sqrt{3}$, permette di esprimere la stessa funzione, in forma standard, fra il massimo relativo ed il minimo relativo, aventi coordinate $(-1, 1)$ e $(1, -1)$:

$$y = \frac{x(x^2 - 3)}{2}$$

In quest'ultima funzione, si effettuano separatamente due nuove sostituzioni di coordinate, al fine di ottenere le prime due funzioni base per l'applicazione del metodo degli elementi finiti. La prima nuova sostituzione di coordinate, così definite: $x = 2x - 1$ e $y = 2y - 1$, trasporta in un quadrato unitario, appartenente al primo quadrante e con un vertice nell'origine: $(0 \le x \le 1,\ 0 \le y \le 1)$, la funzione dispari compresa fra il massimo relativo ed il minimo relativo:

$$y = 2x^3 - 3x^2 + 1 \qquad \text{(I)}$$

la seconda nuova sostituzione di coordinate, così definite: $x = 3x - 2$ e $y = 27y / 2 - 1$, trasporta, ancora nel suddetto quadrato unitario, il settore parabolico cubico, con vertice nel massimo relativo, compreso fra il minimo relativo ed il punto del ramo discendente avente la stessa ordinata:

$$y = x^3 - 2x^2 + x \qquad \text{(II)}$$

A riguardo, si noti, come il settore parabolico cubico, a differenza di quello quadratico (individuato su una conica a parabola), non sia una funzione pari. Le altre due funzioni base per l'applicazione del metodo degli elementi finiti sono ottenute per simmetria. In particolare, dalla prima funzione base per simmetria assiale: $x = 1 - x$ e $y = y$, si ottiene la funzione:

1. Costruzione delle funzioni spline

Le funzioni spline di ordine superiore sono costituite, a partire da quella di ordine zero, con successive convoluzioni, cioè proprio con gli stessi passaggi formali impiegati nel calcolo della funzione densità di probabilità della somma di due variabili casuali uniformemente distribuite. Il punto di partenza è nel primo caso una funzione di spline di ordine zero a supporto compatto (Fig. 1.2 – caso a), formalmente identica alla distribuzione rettangolare definita sullo stesso supporto.

$y = -2x^3 + 3x^2$ (III)

mentre dalla seconda funzione per simmetria raggiata: $x = 1 - x$ e $y = -y$, si ottiene la funzione:

$y = x^3 - x^2$ (IV)

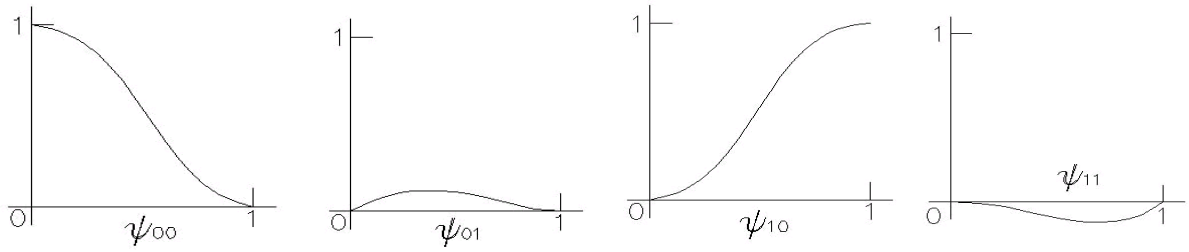


Figura 1.1 – Funzioni base per il metodo degli elementi finiti

Come evidente, tutte le funzioni sono continue con derivata prima continua e pertanto appartengono alla spazio funzionale $C_1(R)$, in quanto presentano discontinuità agli estremi nella derivata seconda (Fig. 1.1).
Le funzioni base bidimensionali (e tridimensionali o multidimensionali) sono ottenute, grazie, alla decomposizione ortogonale di dette funzioni in un intorno quadrato (o in un intorno cubico o ipercubico), moltiplicando a due, a due (a tre, a tre od a più, a più) ciascuna funzione base monodimensionale per ciascuna funzione base monodimensionale (due o più, in numero uguale alla dimensione dell'intorno meno uno), lungo due (tre o più) direzioni ortogonali distinte. In particolare, le sedici funzioni base bidimensionali, definite in un cubo unitario $(0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le z \le 1)$, appartenente al primo ottante, con un vertice nell'origine, hanno espressione:

$z = (2x^3 - 3x^2 + 1)(2y^3 - 3y^2 + 1)$	(I.1)	$z = (-2x^3 + 3x^2)(2y^3 - 3y^2 + 1)$	(III.1)
$z = (2x^3 - 3x^2 + 1)(y^3 - 3y^2 + 1)$	(I.2)	$z = (-2x^3 + 3x^2)(y^3 - 3y^2 + y)$	(III.2)
$z = (2x^3 - 3x^2 + 1)(-2y^3 + 3y^2)$	(I.3)	$z = (-2x^3 + 3x^2)(-2y^3 + 3y^2)$	(III.3)
$z = (2x^3 - 3x^2 + 1)(y^3 - y^2)$	(I.4)	$z = (-2x^3 + 3x^2)(y^3 - y^2)$	(III.4)
$z = (x^3 - 3x^2 + x)(2y^3 - 3y^2 + 1)$	(II.1)	$z = (x^3 - x^2)(2y^3 - 3y^2 + 1)$	(IV.1)
$z = (x^3 - 3x^2 + x)(y^3 - 3y^2 + y)$	(II.2)	$z = (x^3 - x^2)(y^3 - 3y^2 + y)$	(IV.2)
$z = (x^3 - 3x^2 + x)(-2y^3 + 3y^2)$	(II.3)	$z = (x^3 - x^2)(-2y^3 + 3y^2)$	(IV.3)
$z = (x^3 - 3x^2 + x)(y^3 - y^2)$	(II.4)	$z = (x^3 - x^2)(y^3 - y^2)$	(IV.4)

Le funzioni base tri – multidimensionali sono omesse, per motivi di brevità, essendo identiche le modalità della loro costituzione. Resta da sottolineare la loro appartenenza sempre a spazi funzionali (bi, tri – multidimensionali) $C_1(R)$, grazie anche alla continuità delle derivate parziali seconde miste (come noto, date dal prodotto di derivate parziali prime, lungo direzioni ortogonali distinte, a loro volta, come già detto, funzioni continue).

La generica equazione d'osservazione per l'interpolazione di un dato valore argomentale k, compreso in un certo intervallo finito l, con la combinazione lineare di quattro funzioni base monodimensionali ha espressione:

$$y_k = \sum_i a_i^{(l)} F_i\left(\left(x_k - x_l^{min}\right) / \left(x_l^{max} - x_l^{min}\right)\right) \qquad \text{con} \quad i = \text{I, II, III, IV}$$

La generica equazione d'osservazione per l'interpolazione di un dato valore argomentale k, compreso in un certo intorno quadrato l, con la combinazione lineare di sedici funzioni base bidimensionali ha espressione:

$$z_k = \sum_i \sum_j a_{ij}^{(l)} F_{ij}\left(\left(x_k - x_l^{min}\right) / \left(x_l^{max} - x_l^{min}\right), \left(y_k - y_l^{min}\right) / \left(y_l^{max} - y_l^{min}\right)\right) \qquad \text{con} \quad i = \text{I, II, III, IV} \quad \text{e} \quad j = 1, 2, 3, 4$$

La generica equazione d'osservazione per l'interpolazione di un dato valore argomentale, compreso in un certo intorno cubico (o ipercubico), con la combinazione lineare di sessantaquattro (o altra potenza di quattro) funzioni base tridimensionali (o multidimensionali) è omessa, per motivi di brevità, essendo identiche le modalità della sua costruzione.

Si richiama di seguito l'enunciato del teorema (o limite) centrale della statistica, in quanto utile a giustificare l'analogia proposta. Sia $\{x_i\}$ una successione di variabili casuali indipendenti, aventi tutte la stessa distribuzione, media μ e varianza σ^2, allora la successione: $s_n = \sum_{i=1}^{n} x_i$, tende asintoticamente in legge alla variabile casuale x, avente media $n\mu$, varianza $n\sigma^2$ e distribuzione normale: $s_n \cong N\left(n\mu, n\sigma^2\right)$, qualunque sia la distribuzione delle variabili casuali x_i.

Si ricordi che si ha convergenza in legge di una successione di variabili casuali indipendenti $\{x_i\}$ ad un'altra variabile casuale x, quando la successione delle funzioni di distribuzione $\{F_n(x)\}$ ha limite la funzione di distribuzione $F_n(x)$ ovvero: $\lim_{n\to\infty}\{F_n(x)\} = F(x)$, quasi ovunque in x, cioè a meno di un insieme di misura nulla (un numero finito o un'infinità numerabile di punti).

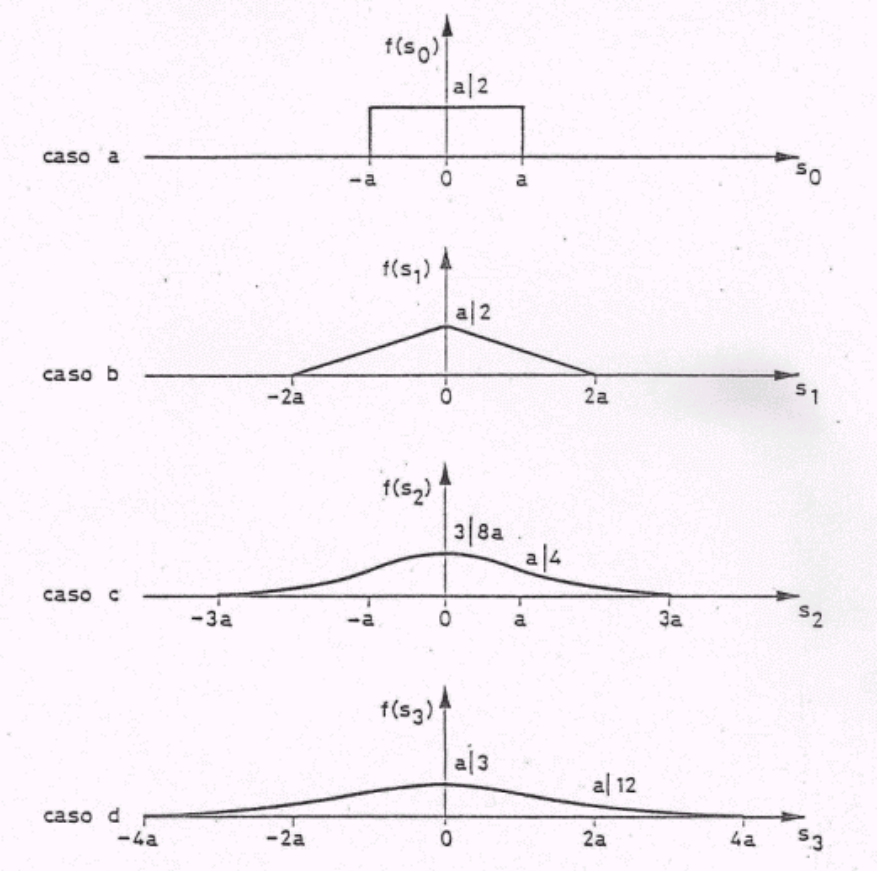


Fig. 1.2 – Funzioni spline di ordine 0, 1, 2, 3

La dimostrazione del teorema è omessa, essendo inutile allo sviluppo dell'argomento. Si mostra invece il calcolo della funzione densità di probabilità della somma di due, tre, quattro variabili casuali indipendenti, nel caso particolare di distribuzione rettangolare, perché analogo alla costruzione di funzioni spline di ordine uno, due, tre. Si dimostra innanzitutto che la funzione densità di probabilità della somma delle due variabili casuali indipendenti è esprimibile tramite un integrale di convoluzione.

Siano x e y due variabili casuali indipendenti e sia: $s = x + y$, la variabile casuale corrispondente alla loro somma. Funzione densità di probabilità $f_s(s)$ è ricavata dalle funzioni densità di probabilità $f_x(x)$ e $f_y(y)$ utilizzano il teorema di conservazione delle probabilità elementari:

$$f_s(s)ds = \int_A f_x(x) f_y(y) dA = \int_{-\infty}^{+\infty} f_x(x) f_y(y) dy dx$$

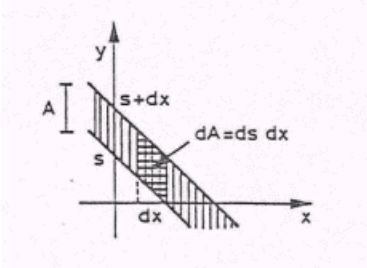


Fig. 1.3 – Somma di variabili casuali indipendenti

Si noti a proposito come l'indipendenza delle variabili casuali sia indispensabile, per poter esprimere la probabilità elementare del prodotto come prodotto delle probabilità elementari marginali, e che la corrispondenza di una regione infinita A all'intervallo infinitesimo ds richiede la formulazione integrale del praccitato teorema e che l'equivalenza: $dA = dydx$, è geometricamente evidente con riferimento alla Fig. 1.3. Infine semplificando, entrambi i membri, per ds e sostituendo: $y = s - x$ (essendo unitario il determinante Jacobiano corrispondente alla sostituzione della variabile d'integrazione), si ha:

$$f_s(s) = \int_{-\infty}^{+\infty} f_x(x) f_y(s-x) dx$$

che rappresenta l'espressione formale dell'integrale di convoluzione.

La funzione densità di probabilità della somma di due variabili casuali indipendenti, ovvero la funzione spline di ordine uno a supporto compatto (Fig. 1.2 – caso b), è ottenuta applicando l'integrale di convoluzione alle funzioni mostrate in Fig. 1.4, nelle posizioni 1 e 2.
Infatti essendo: $f(s_0) \neq 0$, per $-a \leq s_0 \leq a$ ed analogamente: $f(s_1 - s_0) \neq 0$, per $-a \leq s_1 - s_0 \leq a$, ovvero: $s_0 - a \leq s_1 \leq s_0 + a$, come pure: $s_1 - a \leq s_0 \leq s_1 + a$, tenendo conto di entrambe le disuguaglianze, si ha per la seconda: $-2a \leq s_1 \leq 0$ (posizione 1), e $0 \leq s_1 \leq 2a$ (posizione 2) e per il prodotto ai fini del calcolo dell'integrale di convoluzione:

$$f(s_1) = 0 \qquad s_1 < -2a$$

$$f(s_1) = \frac{1}{4a^2} \int_{-a}^{s_1 * a} ds_0 = \frac{2a + s_1}{4a^2} \qquad -2a \leq s_1 \leq 0$$

$$f(s_1) = \frac{1}{4a^2}\int_{s_1-a}^{a} ds_0 = \frac{2a - s_1}{4a^2} \qquad 0 \le s_1 \le 2a$$

$$f(s_1) = 0 \qquad s_1 > 2a$$

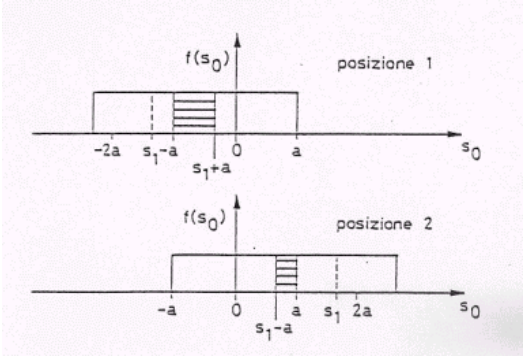


Fig. 1.4 – Costruzione delle funzioni spline di ordine 1

La funzione densità di probabilità della somma di tre variabili casuali indipendenti, ovvero la funzione spline di ordine due a supporto compatto (Fig. 1.2 – caso c), è ottenuta applicando l'integrale di convoluzione alle funzioni mostrate in Fig. 1.5, nelle posizioni 1, 2 e 3, tenendo conto della proprietà associativa della somma e, di conseguenza, degli integrali di convoluzione (pertanto sommare tre funzioni coincide con sommare inizialmente le prime due e successivamente alla loro somma anche la terza).

Infatti essendo: $f(s_1) \neq 0$, per $-2a \le s_1 \le 0$ e $0 \le s_1 \le 2a$, ed analogamente: $f(s_2 - s_1) \neq 0$, per $-a \le s_2 - s_1 \le a$, ovvero: $s_1 - a \le s_2 \le s_1 + a$, come pure: $s_2 - a \le s_1 \le s_2 + a$, tenendo conto di entrambe le disuguaglianze, si ha per la seconda: $-3a \le s_2 \le -a$ (posizione 1), $-a \le s_2 \le +a$ (posizione 2) e $-a \le s_2 \le 3a$ (posizione 3), e per il prodotto, ai fini del calcolo dell'integrale di convoluzione:

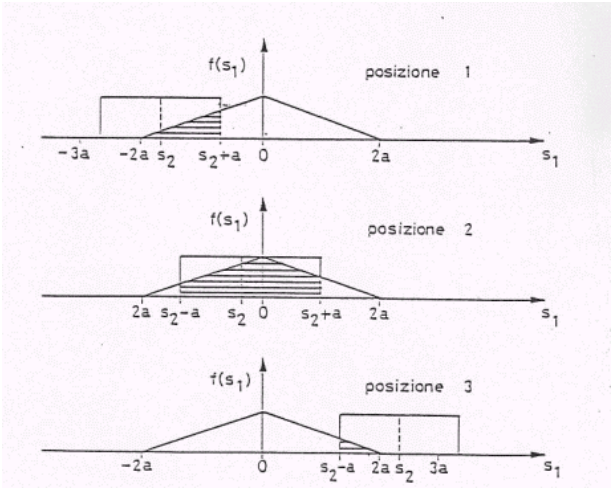


Fig. 1.5 – Costruzione delle funzioni spline di ordine 2

$$f(s_2) = 0 \qquad s_2 < -3a$$

$$f(s_2)=\frac{1}{8a^3}\int_{-2a}^{s_2*a}(2a+s_1)ds_1=\frac{(3a+s_2)^2}{16a^3} \qquad -3a\le s_2\le -a$$

$$f(s_2)=\frac{1}{8a^3}\int_{s_2-a}^{0}(2a+s_1)ds_1+\frac{1}{8a^3}\int_{0}^{s_2+a}(2a-s_1)ds_1=\frac{3}{8a}-\frac{s_2^2}{8a^3} \qquad -a\le s_2\le a$$

$$f(s_2)=\frac{1}{8a^3}\int_{s_2-a}^{2a}(2a-s_1)ds_1=\frac{(3a-s_2)^2}{16a^3} \qquad a\le s_2\le 3a$$

$$f(s_2)=0 \qquad s_2>3a$$

La funzione densità di probabilità della somma di quattro variabili casuali indipendenti, ovvero la funzione spline di ordine tre a supporto compatto (Fig. 1.2 – caso d), è ottenuta applicando l'integrale di convoluzione alle funzioni mostrate in Fig. 1.6, nelle posizioni 1, 2, 3 e 4. Si noti che, sempre per la sopraccitata proprietà associativa della somma, la stessa funzione potrebbe essere ottenuta, applicando l'integrale di convoluzione alla funzione spline di ordine uno con se stessa.

Infatti essendo: $f(s_2)\neq 0$, per $-3a\le s_2\le -a$, $-a\le s_2\le a$ e $a\le s_2\le 3a$, ed analogamente: $f(s_3-s_2)\neq 0$, per $-a\le s_3-s_2\le a$, ovvero: $s_2-a\le s_3\le s_2+a$, come pure: $s_3-a\le s_2\le s_3+a$, tenendo conto di entrambe le disuguaglianze, si ha per la seconda: $-4a\le s_3\le -2a$ (posizione 1), $-2a\le s_3\le 0$ (posizione 2), $0\le s_3\le 2a$ (posizione 3) e $2a\le s_3\le 4a$ (posizione 4), e per il prodotto, ai fini del calcolo dell'integrale di convoluzione:

$$f(s_3)=0 \qquad s_3<-4a$$

$$f(s_3)=\frac{1}{16a^4}\int_{-3a}^{s_3*a}\frac{(3a+s_2)^2}{2}ds_2=\frac{(4a+s_3)^3}{96a^4} \qquad -4a\le s_3\le -2a$$

$$f(s_3)=\frac{1}{16a^4}\int_{s_3-a}^{-a}\frac{(3a+s_2)^2}{2}ds_2+\frac{1}{16a^4}\int_{-a}^{s_3+a}(3a^2-s_2^2)ds_2=\frac{(4a+s_3)^3-4(2a+s_3)^3}{96a^4} \qquad -2a\le s_3\le 0$$

$$f(s_3)=\frac{1}{16a^4}\int_{s_3-a}^{a}(3a^2-s_2^2)ds_2+\frac{1}{16a^4}\int_{a}^{s_3+a}\frac{(3a-s_2)^2}{2}ds_2=\frac{(4a-s_3)^3-4(2a-s_3)^3}{96a^4} \qquad 0\le s_3\le 2a$$

$$f(s_3)=\frac{1}{16a^4}\int_{s_3-a}^{3a}\frac{(3a-s_2)^2}{2}ds_2=\frac{(4a-s_3)^3}{96a^4} \qquad 2a\le s_3\le 4a$$

$$f(s_3)=0 \qquad s_3>4a$$

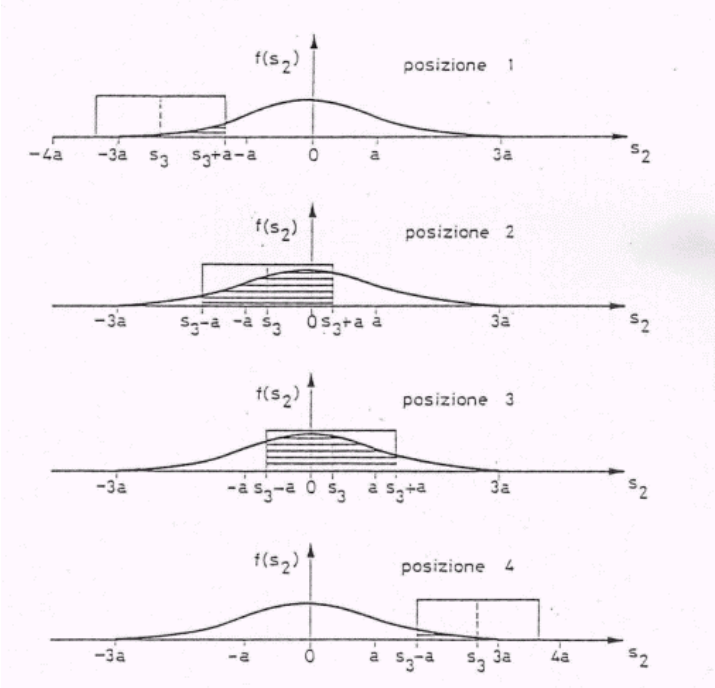


Fig. 1.6 Costruzione della funzione spline di ordine 3

Ad ogni passo di convoluzione si ha un aumento dell'estensione del supporto compatto ed un passaggio a spazi funzionali $C^n\ (R)$, dove l'esponente n è sempre incrementato. Ciò vale a conferma del teorema centrale della statistica, essendo la distribuzione normale definita su tutto l'asse reale ed infinitamente derivabile. Nel prosieguo, si riportano anche le espressioni delle funzioni corrispondenti alla derivata prima della funzione spline di ordine due e delle derivate prima e seconda della funzione spline di ordine tre.

A riguardo, si ricordi che uno spazio funzionale $C^n(R)$, è uno spazio di funzioni continue con le loro derivate, fino all'ordine n, su tutto l'asse reale R.

Le funzioni corrispondenti alla derivata prima della funzione spline di ordine due sono

$f'(s_2) = (3a + s_2)/8a^3$ (posizione 1)

$f'(s_2) = -s_2/4a$ (posizione 2)

$f'(s_2) = -(3a - s_2)/8a^3$ (posizione 3)

In particolare si ha:

$f'(-3a) = 0$

$f'(a) = -1/4a^2$

$f'(3a) = 0$

Le funzioni corrispondenti alla derivata prima della funzione spline di ordine tre sono:

$$f'(s_3) = (4a + s_3)^2 / 32a^4 \qquad \text{(posizione 1)}$$

$$f'(s_3) = \left[(4a + s_3)^2 - 4(2a + s_3)^2\right] / 32a^4 \qquad \text{(posizione 2)}$$

$$f'(s_3) = -\left[(4a - s_3)^2 - 4(2a - s_3)^2\right] / 32a^4 \qquad \text{(posizione 3)}$$

$$f'(s_3) = -(4a - s_3)^2 / 32a^4 \qquad \text{(posizione 4)}$$

In particolare, si ha:

$$f'(-4a) = 0$$
$$f'(-2a) = 1/2a^2$$
$$f'(0) = 0$$
$$f'(2a) = -1/2a^2$$
$$f'(4a) = 0$$

Le funzioni corrispondenti alla derivata seconda della funzione spline di ordine tre sono:

$$f''(s_3) = (4a + s_3)/16a^4 \qquad \text{(posizione 1)}$$

$$f''(s_3) = [(4a + s_3) - 4(2a + s_3)]/16a^4 \qquad \text{(posizione 2)}$$

$$f''(s_3) = [(4a - s_3) - 4(2a - s_3)]/16a^4 \qquad \text{(posizione 3)}$$

$$f''(s_3) = (4a - s_3)/16a^4 \qquad \text{(posizione 4)}$$

In particolare, si ha:

$$f''(-4a) = 0$$
$$f''(-2a) = 1/8a^3$$
$$f''(0) = -1/4a^3$$
$$f''(2a) = 1/8a^3$$
$$f''(4a) = 0$$

Resta da osservare che le funzioni corrispondenti:

- alla derivata prima della funzione spline di ordine uno;
- alla derivata seconda della funzione spline di ordine due;
- alla derivata terza della funzione spline di ordine tre,

sono tutte costanti e, di conseguenza, le derivate successive identicamente nulle.

2. Proprietà geometriche delle funzioni spline

Ad ogni passo di convoluzione, si ha un aumento tanto dell'estensione del supporto compatto, quando del grado della funzione spline, ovvero dell'ordine di derivabilità della stessa. In particolare:

- la funzione spline di ordine zero, a supporto compatto in $[-a, a]$, è solo definita ovunque;
- la funzione spline di ordine uno, a supporto compatto in $[-2a, 2a]$, appartiene a $C^0(R)$, ovvero è definita ovunque ed è ovunque continua, ma la sua derivata prima non è ovunque continua;
- la funzione spline di ordine due, a supporto compatto in $[-3a, 3a]$, appartiene a $C^1(R)$, ovvero è definita ovunque ed è ovunque continua con derivata prima anch'essa ovunque continua, ma la sua derivata seconda non è ovunque continua;
- la funzione spline di ordine tre, a supporto compatto in $[-4a, 4a]$, appartiene a $C^2(R)$, ovvero è definita ovunque ed è ovunque continua con derivate prima e seconda anch'esse ovunque continue.

Si ricordi poi che la combinazione lineare di funzioni definite è pure una funzione definita; inoltre gli spazi funzionali $C^n(R)$ sono spazi vettoriali: ogni combinazione lineare fra funzioni interne allo spazio genera una funzione interna allo spazio. Pertanto la combinazione lineare di funzioni spline di ordine zero, o uno, o due, o tre opportunamente ridefinite con una traslazione del supporto, permette insieme l'estensione indefinita del supporto compatto e la conservazione delle proprietà geometriche già possedute dalle funzioni spline combinate.

Infine date n funzioni spline di variabili geometriche ortogonali il loro prodotto cartesiano è una funzione spline n – dimensionale. Questa funzione è a supporto compatto nell'ipervolume descritto dal prodotto dei supporti compatti dei singoli fattori e conserva estese all'ipervolume le proprietà geometriche già possedute dalle funzioni spline monodimensionali. A sua volta, la combinazione lineare di funzioni spline multidimensionali, ridefinite con opportune traslazioni del supporto, è una funzione spline, avente estensione indefinita del supporto compatto, che ancora conserva le proprietà geometriche già possedute dalle funzioni spline combinate.

Di particolare importanza è il prodotto di due (o tre) funzioni spline di variabili geometricamente ortogonali. Infatti la funzione spline risultante è una funzione spline bidimensionale, (tridimensionale) ovvero una superficie (ipersuperficie) a supporto compatto su un quadrato (un cubo) avente per lati i supporti compatti dei due (tre). Fra queste, la funzione spline bilineare (trilineare) è ottenuta come prodotto di due (tre) funzioni spline di ordine uno delle variabili cartesiane ortogonali fattori (in particolare, il supporto quadrato della funzione spline bilineare è composto da quattro celle piane) e la funzione spline bicubica (tricubica) è ottenuta come prodotto di due (tre) funzioni spline di ordine tre delle medesime variabili cartesiane ortogonali (in particolare, il supporto quadrato della funzione spline bicubica è composto da sedici celle piane).

La funzione spline bilineare è continua, ma non è ovunque derivabile e, secondo le quattro posizioni ottenute applicando la regola: $i, j \rightarrow k(i-1)+j = l$, con $i, j = 1, k$, dove $k = 2$, ha valori:

$$f(s_1,t_1)=\frac{1}{16a^4}(2a+s_1)(2a+t_1) \qquad \text{per } -2a \le s_1 \le 0 \text{ e } -2a \le t_1 \le 0$$

ovvero in posizione $1,\ 1\rightarrow 2(1-1)+1=1$

$$f(s_1,t_1)=\frac{1}{16a^4}(2a+s_1)(2a-t_1)$$ per $-2a \le s_1 \le 0$ e $0 \le t_1 \le 2a$

ovvero in posizione 1, $2 \rightarrow 2(1-1)+2=2$

$$f(s_1,t_1)=\frac{1}{16a^4}(2a-s_1)(2a+t_1)$$ per $0 \le s_1 \le 2a$ e $-2a \le t_1 \le 0$

ovvero in posizione 1, $2 \rightarrow 2(2-1)+1=3$

$$f(s_1,t_1)=\frac{1}{16a^4}(2a-s_1)(2a-t_1)$$ per $0 \le s_1 \le 2a$ e $0 \le t_1 \le 2a$

ovvero in posizione 2, $2 \rightarrow 2(2-1)+2=4$

La funzione spline bicubica è continua, con derivate prima e seconda anch'esse ovunque continue e, secondo le sedici posizioni applicando la regola: $i, j \rightarrow k(i-1)+j = l$, con $i, j = 1, k$, dove $k = 4$, ha valori:

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a+s_3)^3(4a+t_3)^3$$ per $-4a \le s_3 \le -2a$ e $-4a \le t_3 \le -2a$

ovvero in posizione 1, $1 \rightarrow 4(1-1)+1=1$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a+s_3)^3\left((4a+t_3)^3-4(2a+t_3)^3\right)$$ per $-4a \le s_3 \le -2a$ e $-2a \le t_3 \le 0$

ovvero in posizione 1, $2 \rightarrow 4(1-1)+2=2$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a+s_3)^3\left((4a-t_3)^3-4(2a-t_3)^3\right)$$ per $-4a \le s_3 \le -2a$ e $0 \le t_3 \le 2a$

ovvero in posizione 1, $3 \rightarrow 4(1-1)+3=3$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a+s_3)^3(4a-t_3)^3$$ per $-4a \le s_3 \le -2a$ e $2a \le t_3 \le 4a$

ovvero in posizione 1, $4 \rightarrow 4(1-1)+4=4$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a+s_3)^3-4(2a+t_3)^3\right)(4a-t_3)^3$$ per $-2a \le s_3 \le 0$ e $-4a \le t_3 \le -2a$

ovvero in posizione 2, $1 \rightarrow 4(2-1)+1=5$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a+s_3)^3-4(2a+t_3)^3\right)\left((4a+t_3)^3-4(2a+t_3)^3\right)$$

per $-2a \le s_3 \le 0$ e $-2a \le t_3 \le 0$

ovvero in posizione 2, $2 \rightarrow 4(2-1)+2=6$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a+s_3)^3-4(2a+t_3)^3\right)\left((4a-t_3)^3-4(2a-t_3)^3\right)$$

per $-2a \le s_3 \le 0$ e $0 \le t_3 \le 2a$

ovvero in posizione 2, $3 \rightarrow 4(2-1)+3=7$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a+s_3)^3-4(2a+s_3)^3\right)(4a-t_3)^3 \qquad \text{per } -2a\le s_3\le 0 \text{ e } 2a\le t_3\le 4a$$

ovvero in posizione 2, $4\rightarrow 4(2-1)+4=8$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a-s_3)^3-4(2a-t_3)^3\right)(4a+t_3)^3 \qquad \text{per } 0\le s_3\le 2a \text{ e } -4a\le t_3\le -2a$$

ovvero in posizione 3, $1\rightarrow 4(3-1)+1=9$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a-s_3)^3-4(2a-t_3)^3\right)\left((4a+t_3)^3-4(2a+t_3)^3\right)$$

per $0\le s_3\le 2a$ e $-2a\le t_3\le 0$

ovvero in posizione 3, $2\rightarrow 4(3-1)+2=10$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a-s_3)^3-4(2a-t_3)^3\right)\left((4a-t_3)^3-4(2a-t_3)^3\right)$$

per $0\le s_3\le 2a$ e $0\le t_3\le 2a$

ovvero in posizione 3, $3\rightarrow 4(3-1)+3=11$

$$f(s_3,t_3)=\frac{1}{9216a^8}\left((4a-s_3)^3-4(2a-s_3)^3\right)(4a-t_3)^3 \qquad \text{per } 0\le s_3\le 2a \text{ e } 2a\le t_3\le 4a$$

ovvero in posizione 3, $4\rightarrow 4(3-1)+4=12$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a-s_3)^3(4a+t_3)^3 \qquad \text{per } 2a\le s_3\le 4a \text{ e } -4a\le t_3\le -2a$$

ovvero in posizione 4, $1\rightarrow 4(4-1)+1=13$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a-s_3)^3\left((4a+t_3)^3-4(2a+t_3)^3\right) \qquad \text{per } 2a\le s_3\le 4a \text{ e } -2a\le t_3\le 0$$

ovvero in posizione 4, $2\rightarrow 4(4-1)+2=14$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a-s_3)^3\left((4a-t_3)^3-4(2a-t_3)^3\right) \qquad \text{per } 2a\le s_3\le 4a \text{ e } 0\le t_3\le -2a$$

ovvero in posizione 4, $3\rightarrow 4(4-1)+3=15$

$$f(s_3,t_3)=\frac{1}{9216a^8}(4a-s_3)^3(4a-t_3)^3 \qquad \text{per } 2a\le s_3\le 4a \text{ e } 2a\le t_3\le 4a$$

ovvero in posizione 4, $4\rightarrow 4(4-1)+4=16$

Si noti che sia la funzione spline bilineare sia la funzione spline bicubica non sono funzioni isotrope (anche se quella cubica è pressoché isotropa fino alla quota 0.005); infatti le isolinee di livello di entrambe sono costituite da particolari linee chiuse concentriche, contenute nel supporto compatto quadrato, come illustra la Fig. 2.1. La Fig. 2.2 mostra una vista assonometrica delle funzioni spline bilineare e bicubica.

Le funzioni corrispondenti alle derivate prime e seconde della funzione spline bicubica si ottengono facilmente, tenendo conto delle seguenti semplici regole di derivazione:

- le derivate prime sono prodotto della derivata prima di uno dei due fattori monodimensionali per l'altro fattore monodimensionale;
- le derivate seconde pure sono il prodotto della derivata seconda di uno dei due fattori monodimensionali per l'altro fattore monodimensionale;
- la derivata seconda mista è il prodotto delle derivate prime di entrambi i fattori monodimensionali.

Le funzioni spline si dicono poste in forma standard se: $a = ½$.

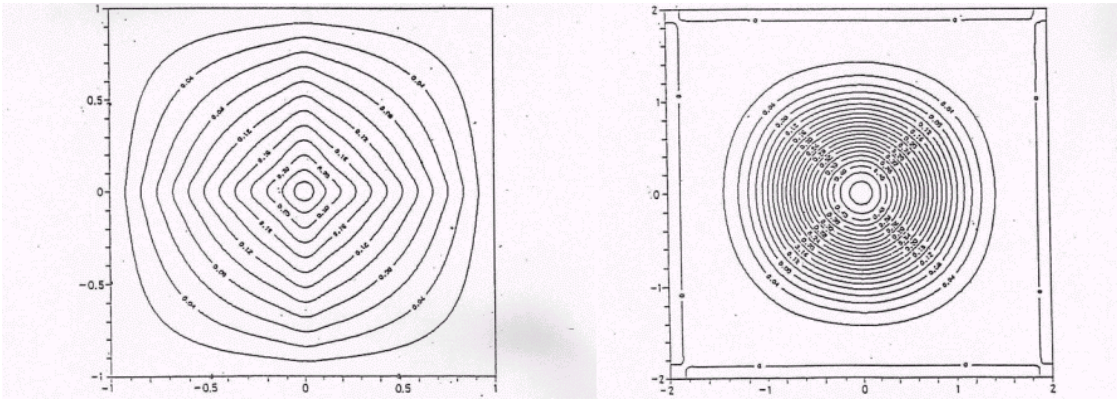


Fig. 2.1 – Curve di livello delle funzioni spline bilineare e bicubica

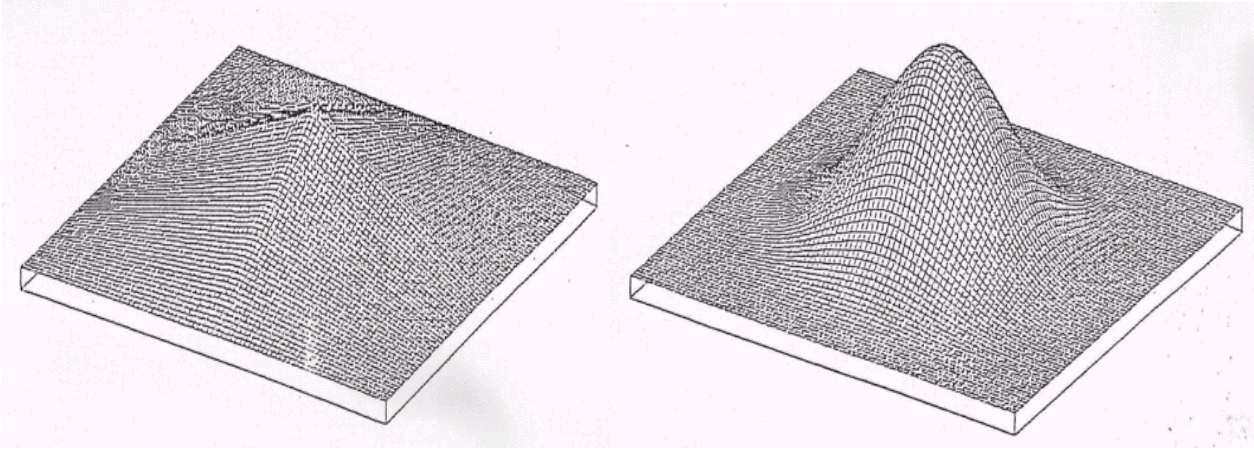


Fig. 2.2 – Assonometria delle funzioni spline bilineare e bicubica

3. Interpolazione con funzioni spline

L'interpolazione di un insieme di dati i dati, di una serie temporale (o altra sequenza monodimensionale) oppure quelli di un lattice a supporto territoriale) con un insieme di funzioni spline (lineari: caso a; cubiche: caso b; bilineari: caso c; bicubiche: caso d) richiede quali basi di dati in ingresso:

- i tempi di osservazione (t), i corrispondenti valori osservati (y_0) e l'intervallo temporale (Δt), costituente il passo d'interpolazione monodimensionale;
- le coordinate piane (x, y) dei punti di osservazione, i corrispondenti valori osservati (z_0) ed il lato del quadrato ($\Delta x = \Delta y$), costituente il passo d'interpolazione bidimensionale.

Si osservi come mentre, nel primo caso, il passo d'interpolazione può essere adattativo, nel secondo caso, il passo d'interpolazione e la finestra generata devono essere costanti, anche se la finestra generata potrebbe essere rettangolare. Si noti tuttavia che passi e finestre variabili da zona a zona possono essere realizzati, in successione, ove si proceda sequenzialmente.
In questo caso (come pure con un passo adattativo nel caso monodimensionale), tramite l'analisi di varianza (confronto fra la varianza spiegata e la varianza residua), si procede al campionamento progressivo, cioè all'acquisizione di nuove osservazioni, nelle zone dove l'approssimazione non è accettabile, mentre si procede al campionamento selettivo, cioè al diradamento delle osservazioni, nelle zone dove l'approssimazione è più che accettabile.

Alla fase di input segue la fase della classificazione dei dati, articolata nei seguenti passi:

- ricerca dei valori estremi per i parametri ordinatori dei dati;
- calcolo del numero di classi pari a:

  nel caso monodimensionale, la parte intera del quoziente della differenza, fra l'estremo superiore (t_{max}) e l'estremo inferiore (t_{min}), del parametro ordinatore sull'intervallo temporale (Δt), incrementata di un'unità: $\nu = int\left(\left(t_{max} - \left(t_{min} - \varepsilon\Delta t\right)\right)/\Delta t\right)+1$, dove ε è un numero positivo opportunamente piccolo.

  nel caso bidimensionale, il prodotto di due numeri interi: $\mu = \lambda\eta$, definiti in analogia al caso precedente come : $\lambda = int\left(\left(x_{max} - \left(x_{min} - \varepsilon\Delta x\right)\right)/\Delta x\right)+1$ e $\eta = int\left(\left(y_{max} - \left(y_{min} - \varepsilon\Delta y\right)\right)/\Delta y\right)+1$, dove λ e η sono il numero di classi (monodimensionali) rispettivamente nelle direzioni dell'ascissa e dell'ordinata, ovvero le parti intere del quoziente delle differenze, fra gli estremi superiori (x_{max}, y_{max}) e gli estremi inferiori (x_{min}, y_{min}) delle coordinate ascissa e ordinata, sul lato del quadrato ($\Delta x = \Delta y$), incrementati di un'unità, essendo ancora ε un numero positivo opportunamente piccolo.

- calcolo del numero di incognite, pari a:

  nel caso monodimensionale, un'unità più il numero delle classi: $n = \nu + 1$ (caso a), tre unità più il numero delle classi: $n = \nu + 3$ (caso b);

  nel caso bidimensionale, il prodotto di due numeri interi: $n = lh = \left(\lambda+1\right)\left(\nu+1\right)$ (caso c), e: $n = lh = \left(\lambda+3\right)\left(\nu+3\right)$ (caso d), essendo tali numeri i numeri delle classi (monodimensionali) nelle direzioni dell'ascissa e dell'ordinata, aumentati di una e tre unità, a seconda del caso considerato.

- ricerca di eventuali intervalli o quadrati vuoti, oltre al primo ed all'ultimo intervallo, limitatamente al caso b, ed a tutti i quadrati di bordo, limitatamente al caso d:

$$P \in A_i \cup A_{i+1} \qquad \text{con} \qquad i = 1, \nu - 1$$

$$P \in A_i \cup A_{i+1} \cup A_{i+\lambda} \cup A_{i+\lambda+1} \qquad \text{con} \qquad i = 1, \nu - 1$$

Infatti, nel caso a, due intervalli consecutivi vuoti non permettono di calcolare il coefficiente della funzione spline centrata sul nodo intermedio. Pertanto onde evitare una singolarità nel sistema normale, dovrebbe essere vincolato a zero tale coefficiente. Analogamente nel caso b, quattro intervalli consecutivi vuoti non permettono di calcolare il coefficiente della funzione spline centrata sul nodo intermedio e anche due intervalli consecutivi vuoti, a causa dell'enorme smorzamento sulle code della funzione spline di ordine tre, non consentono di calcolare bene lo stesso valore. Pertanto onde evitare una singolarità e comunque un malcondizionamento nel sistema normale, dovrebbe essere vincolato a zero tale coefficiente. In questo caso, anche i coefficienti della prima ed ultima funzione spline sono sempre vincolati a zero; tuttavia essi non sono cancellati dal sistema di equazioni di osservazione così che tutte le equazioni abbiano formalmente quattro coefficienti.

Nel caso c, quattro quadrati adiacenti vuoti, disposti in forma di quadrato, non permettono di calcolare il coefficiente della funzione spline con vertice sul nodo centrale. Pertanto onde evitare una singolarità nel sistema normale, dovrebbe essere vincolato a zero tale coefficiente. Analogamente nel caso d, sedici quadrati adiacenti vuoti, disposti in forma di quadrato, non permettono di calcolare il coefficiente della funzione spline con vertice sul nodo centrale e anche quattro quadrati adiacenti vuoti, disposti in forma di quadrato, a causa dell'enorme smorzamento sulle code della funzione spline bicubica, non consentono di calcolare bene lo stesso valore. Pertanto onde evitare una singolarità e comunque un malcondizionamento nel sistema normale, dovrebbe essere vincolato a zero tale coefficiente. In questo caso, anche i coefficienti di tutte le funzioni spline perimetrali (ovvero situate sui contorni) sono sempre vincolati a zero; tuttavia essi non sono cancellati dal sistema di equazioni di osservazione così che tutte le equazioni abbiano formalmente sedici coefficienti.

Si ricordi a proposito che un sistema è detto numericamente malcondizionato, quando piccole variazioni nei coefficienti, oppure nei termini noti o nei pesi, producono grosse variazioni nella soluzione, negli scarti – residui e soprattutto nelle matrici che forniscono la loro attendibilità (in termini algebrici, detto sistema è un sistema quasi singolare).

Si osservi inoltre come, limitatamente agli intervalli ed ai quadrati vuoti, sia preferibile introdurre equazioni di pseudo – osservazione alle differenze finite prime fra i coefficienti, di peso basso rispetto a quello delle equazioni di osservazione, invece di imporre i vincoli sopraccitati. Infatti questa soluzione permette di essere più aderente a quella che si può presumere possa essere il comportamento reale dei dati, benché mancanti, mentre i vincoli a zero producono inevitabilmente buchi inverosimili nella rappresentazione continua dei dati stessi. Si noti infine, come proprio l'adozione delle suddette pseudo – osservazioni, grazie al loro basso peso, permette la loro scrittura per tutto il dominio dei dati, a prescindere dall'esistenza o meno di intervalli o quadrati vuoti. Infatti dove detti vuoti esistono, queste pseudo – osservazioni garantiscono da singolarità e malcondizionamenti, mentre dove non sono presenti, proprio il basso peso non altera di molto la soluzione e la sua attendibilità.

Si riportano di seguito le equazioni di pseudo – osservazioni alle differenze finite prime fra i coefficienti delle funzioni spline, rispettivamente nel caso monodimensionale e nel caso bidimensionale:

$$b_{I+i+1} - b_{I+i-1} = 0 \qquad a_{I+i+1\,J+j} - a_{I+i-1\,J+j} = 0$$

$$a_{I+i\,J+j+1} - a_{I+i\,J+j-1} = 0$$

Questo modo di procedere è noto come l'impiego della norma ibrida di Tikhonov, in quanto la scrittura delle equazioni di pseudo – osservazione di basso peso, determina scarti – residui nelle stesse che devono essere minimizzati contemporaneamente agli scarti – residui delle equazioni di osservazione.

La scrittura dei coefficienti delle equazioni di osservazione può essere sintetizzata, caso per caso, in espressioni compatte, dove il valore osservato nel generico punto P è scomposto in un valore interpolato nello stesso punto e nello scarto – residuo dell'equazione di osservazione corrispondente. A sua volta, il valore interpolato è rappresentabile come la sommatoria dei prodotti del coefficiente incognito della funzione spline in esame per un simbolo sostitutivo dell'espressione formale della stessa funzione.
La posizione del generico punto P è definita rispettivamente dalla sua ascissa, nei casi a, b, e dalle sue coordinate piane, nei casi c, d, dove ogni parametro è sempre espresso nelle unità standard della funzione spline in esame[2].

Caso a: $\hat{y}^P = y_0^P + \hat{v}^P = \sum_{j=i}^{i+1} \hat{a}_j S^j_{j-i+1}\left(s_P^j\right) \qquad \forall P \in A_i \qquad i=1,n-1$

essendo:

y_0^P , $\hat{y}^P$ i valori osservato ed interpolato nel punto P

$\hat{v}^P = \hat{y}^P - y_0^P$ lo scarto – residuo dell'equazione di osservazione

$s_P^j = \left(t^P - t^j\right) / \Delta t$ l'ascissa del punto P, espressa nell'unità standard della funzine spline j

t^P , t^j le ascisse del punto P e del nodo della funzione spline j

inoltre: $\hat{a}_j$ il coefficiente incognito della funzione spline di ordine uno j e S^j_{j-i+1} un simbolo sostitutivo dell'espressione formale della funzione spline j in posizione $j-i+1$.

---

[2] Il numero di addendi della suddetta sommatoria è pari a:

- due e quattro, rispettivamente nei casi a, b;
- quattro e sedici, rispettivamente nei casi c, d.

Infatti solo le funzioni spline aventi valore massimo nei nodi vicini ad un dato punto sono direttamente coinvolte nell'interpolazione dello stesso. Allora con specifico riferimento ai quattro casi già illustrati in precedenza, i nodi vicini sono rispettivamente:

- il precedente ed il seguente (cioè due) la classe, ovvero l'intervallo contenente un dato punto, nel caso a;
- due precedenti e due seguenti (cioè quattro) la classe contenente un dato punto, nel caso b;
- quattro circostanti la classe, ovvero l'intorno quadrato contenente un dato punto, nel caso c;
- quattro circostanti più dodici sub – circostanti (in totale sedici) la classe contenente un dato punto, nel caso d.

Pertanto tutte le funzioni spline, aventi vertici negli altri nodi, non danno alcun contributo.

Caso b: $\hat{y}^p = y_0^p + \hat{v}^p = \sum_{j=i-1}^{i+2} \hat{a}_j S^{j}_{j-i+2\left(s_p^j\right)}$ $\forall P \in A_i$ i=2,n–2

dove tutti i simboli hanno i significati del caso precedente, pur essendo la funzione spline in uso di ordine tre.

Caso c: $\hat{z}^p = z_0^p + \hat{v}^p = \sum_{k'=i}^{i+1} \sum_{k''=j}^{j+1} \hat{a}_{(k''-1)l+k'} S^{(k''-1)l+k'}_{2(k''-j)+k'-i+1\left(s_p^{k'}, t_p^{k''}\right)}$ $\forall P \in A_i$

i=1,l–1 e j=1,h–1 (lh=n)

essendo

z_0^p, $\hat{z}^p$ i valori osservato ed interpolato nel punto P

$\hat{v}^p = \hat{z}^p - z_0^p$ lo scarto – residuo nell'equazione di osservazione

$s_p^{k'} = \left(x^p - x^{k'}\right)/\Delta x$, $t_p^{k''} = \left(y^p - y^{k''}\right)/\Delta y$ le coordinate piane del punto P, espresse nell'unità standard della spline $(k''-1)l+k'$

(x^p, y^p), $(x^{k'}, y^{k''})$ le coordinate piane del punto P e del nodo della funzione spline $(k''-1)l+k$

inoltre: $\hat{a}_{(k''-1)l+k'}$ il coefficiente incognito della spline bidimensionale di ordine uno $(k''-1)l+k$ e $S^{(k''-1)l+k'}_{2(k''-j)+k'-i+1}$ un simbolo sostitutivo dell'espressione formale della funzione spline $(k''-1)l+k$ in posizione $(k''-j)l+k'-i+1$.

Caso d: $\hat{z}^p = z_0^p + \hat{v}^p = \sum_{k'=i-1}^{i+2} \sum_{k''=j-1}^{j+2} \hat{a}_{(k''-1)l+k'} S^{(k''-1)l+k'}_{4(k''-j+1)+k'-i+2\left(s_p^{k'}, t_p^{k''}\right)}$ $\forall P \in A_i$

i=1,l–3 e j=1,h–3 (lh=n)

dove tutti i simboli hanno significati del caso precedente, pur essendo la funzione bidimensionale spline in uso di ordine tre.

Come usuale nei problemi di interpolazione tutte le osservazioni sono considerate indipendenti ed aventi la stessa precisione; pertanto tutte le osservazioni hanno peso uno, a meno di un successivo passo per il raffinamento del modello stocastico. La fase di scrittura del sistema termina con la scrittura di equazioni di pseudo – osservazione sovrappesate, se sostitutive dei vincoli che costringono a zero le ordinate dei nodi estremi, oppure le quote dei nodi di contorno, e sottopesate ovunque, per salvaguardare i nodi interni ad intervalli od a quadrati vuoti, come già illustrato in precedenza[3].

[3] Si noti come la scrittura delle equazioni di osservazione e pseudo – osservazione si traduca essenzialmente in un'imponente mole di indici (indicatori, contatori, puntatori, puntatori rovesci), atti a collegare fra loro informazioni relative a punti, classi, nodi e funzioni spline.

Le matrici disegno e normale hanno rispettivamente forma:

- a diagonali parallele, precisamente bidiagonale o tetradiagonale nei casi a, b;
- a gruppi di diagonali parallele, due volte bidiagonale e quattro volte tetradiagonale, nel casi c, d.
- a banda, con ampiezza della banda pari a due o quattro, nei casi a, b (Fig. 3.1 – casi a, b);
- a banda, con una sovrabanda pure a banda, con ampiezza della banda rispettivamente pari a due o quattro e tre o sette, nei caso c, d (Fig. 3.1 – casi c, d).

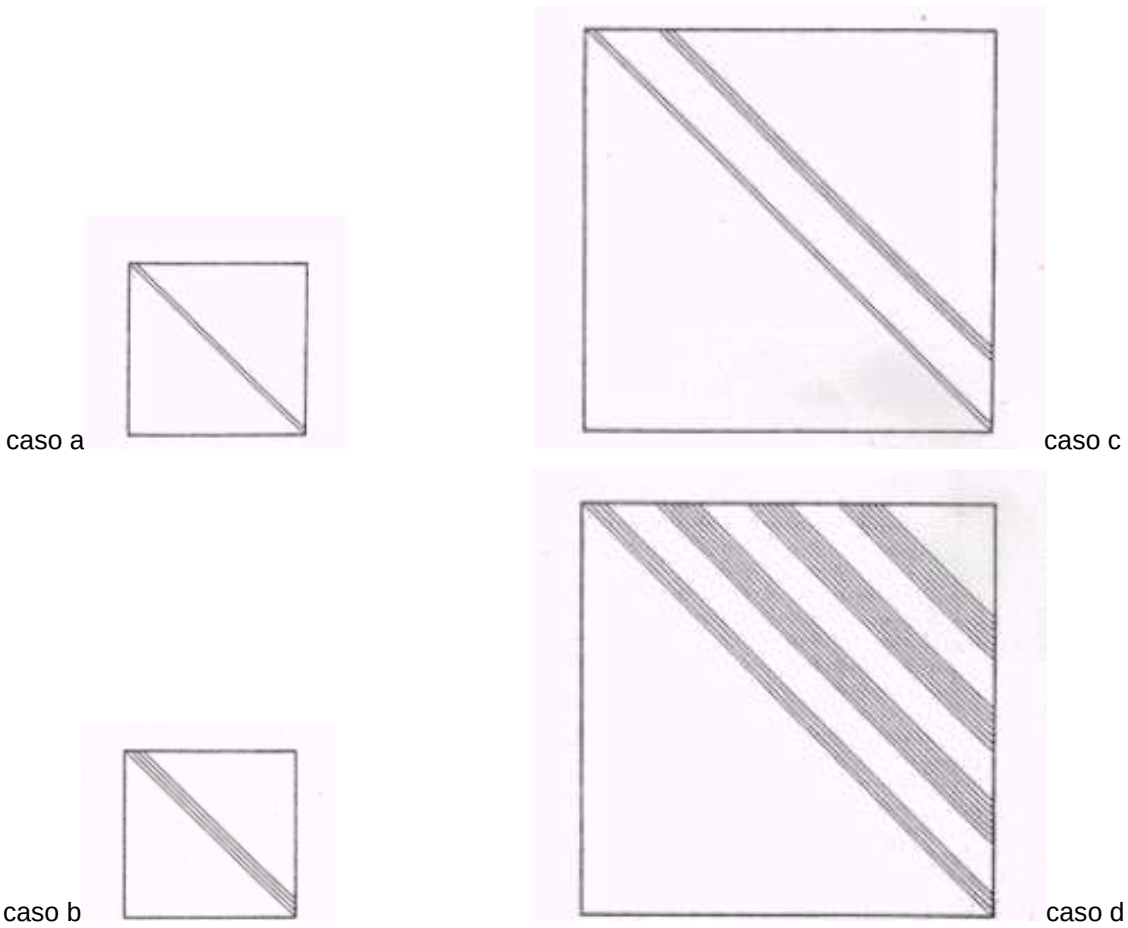


Fig. 3.1 – Matrici normali dei problemi d'interpolazione con funzioni spline

La stima della soluzione e della precisione globale, nonché il calcolo della precisione e dell'affidabilità, si ottengono agevolmente con algoritmi diretti a banda, nei casi a, b, e con algoritmi iterativi rapidamente convergenti, nei casi c, d.

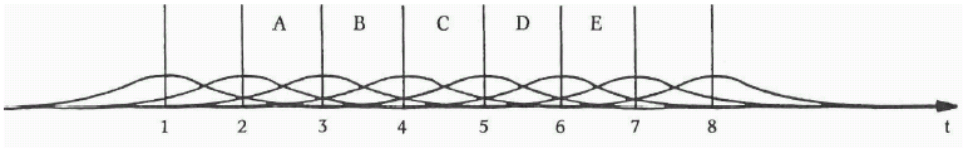
Alla fase di calcolo, segue la fase di output che fornisce i seguenti risultati[4]:

- stima dei coefficienti delle funzioni spline e calcolo dei loro sqm;

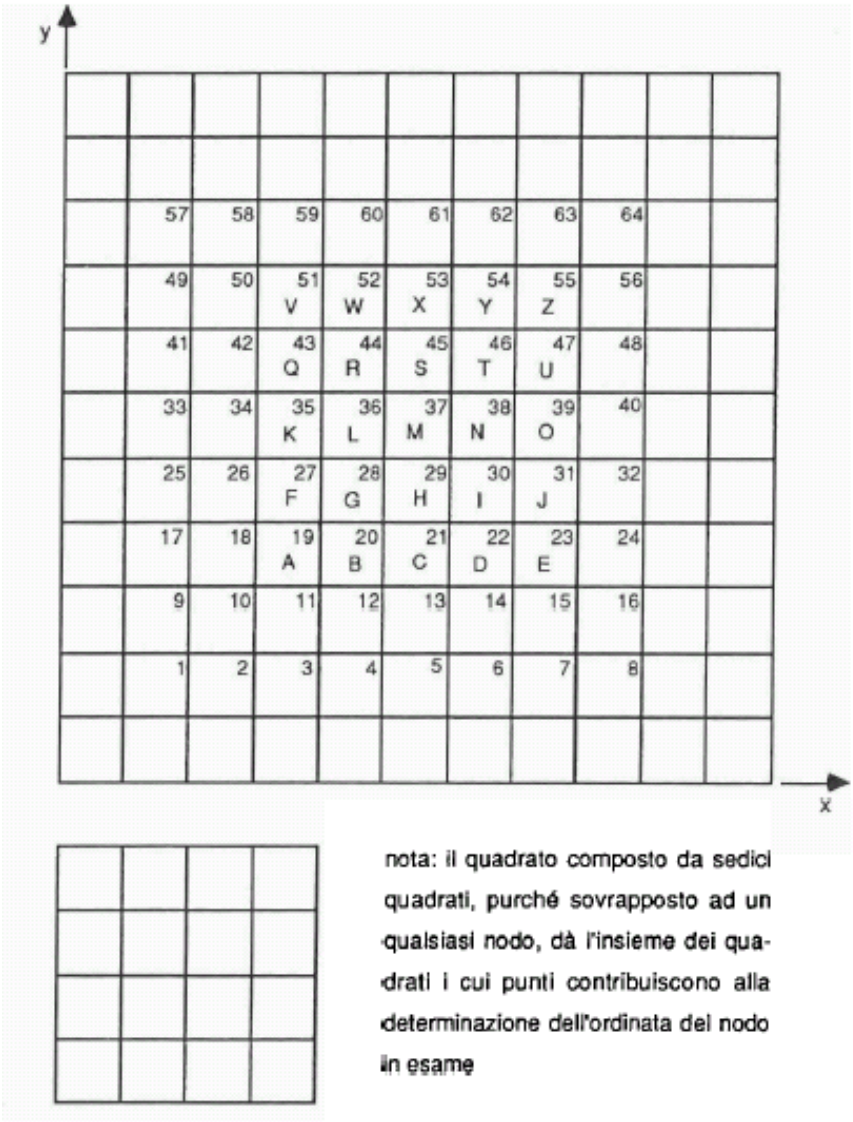
[4] Come noto ed usuale, l'interpolazione avviene a minimi quadrati, nel caso di basi di dati ripuliti, mentre nel caso di basi di dati contaminate da dati anomali è conveniente procedere facendo uso di procedure robuste. In questo caso solitamente, ad una prima elaborazione a minimi quadrati, fa seguito l'applicazione delle procedura robusta propriamente detta, capace di identificare ed eliminare i sospetti dati anomali. Dopodiché la tecnica dei minimi quadrati sequenziali, affiancata da una potente validazione inferenziale dei dati in esame, permette di riammettere le osservazioni compatibili con i modelli funzionale e stocastico, benché inizialmente sospette.

- stima degli scarti – residui delle equazioni di osservazione e calcolo dei loro sqm;
- stima dei valori interpolati dei punti e calcolo dei loro sqm;
- stima di sigma zero e calcolo di altra documentazione statistica,

cui aggiungere la documentazione numerica per il controllo del condizionamento e dell'affidabilità.
Omettendo per ragioni di brevità e data la loro maggiore semplicità i primi due casi (a, b): lineare e bilineare, la Fig. 3.2 illustra l'interpolazione con funzioni spline cubiche (caso c) e bicubiche (caso d).



caso b



caso d

Fig. 3.2 – Interpolazione con funzioni spline cubiche e bicubiche

Si noti come le curve (formate dalle spline cubiche, definite su quattro intervalli) del caso c illustrino, con evidenza grafica, quanto deve essere supposto per il caso d, ovvero tante superfici, formate dalle spline bucubiche (ciascuna definita su sedici quadrati, disposti in forma di quadrato). Ovviamente è poi compito dei

coefficienti delle suddette funzioni spline adattare l'interpolazione ai dati, innalzando od abbassando opportunamente curve e superfici e, come noto, tale compito è assolto dal criterio di stima adottato (ad es., il criterio dei minimi quadrati e/o qualche procedura robusta).

Le funzioni base tri – multidimensionali sono omesse, per motivi di brevità, essendo identiche le modalità della loro costituzione. Resta da sottolineare la loro appartenenza sempre a spazi funzionali (tri – multidimensionali) $C_n(R)$, grazie anche alla continuità delle derivate parziali seconde miste (come noto, date dal prodotto di derivate parziali prime, lungo direzioni ortogonali distinte, a loro volta, come già detto in precedenza, funzioni continue). Allora la scrittura della generica equazione d'osservazione per l'interpolazione di un dato valore argomentale, compreso in un certo intorno cubico (o ipercubico), come combinazione lineare di sessantaquattro (o altra potenza di quattro) funzioni base tridimensionali (o multidimensionali), omessa sempre per motivi di brevità, è formalmente identica a quanto esposto per la scrittura del sistema di equazioni di osservazione e pseudo – osservazione nei casi monodimensinale e bidimensionale.
Un'altra interessante applicazione delle funzioni spline è costituita dai descrittori di forma, comunemente utilizzati per la rappresentazione matematica di contorni di figure piane o gobbe (caso bidimensionale e sua estensione allo spazio 3D) e di superfici (chiuse) di oggetti (caso tridimensionale) comunque conformati. Infatti esistono vari approcci matematici al problema dei descrittori di forma e le funzioni spline rappresentano uno tra i metodi efficaci in questo campo.

- Nel campo della rappresentazione dei contorni di figure piane o gobbe, ogni problema ha già trovato la sua soluzione. In particolare, una struttura di riferimento semplice è data da una qualsiasi spazzata per i punti, approssimazione dell'ascissa curvilinea del contorno di una figura ed il problema consiste nell'individuare l'espressione analitica del contorno, a partire dalle osservazioni costituite da un insieme di coppie (o terne) di coordinate di punti di una figura piana (o gobba). In quest'ambito (ma non solo potendosi, altrettanto bene, applicare ad una classica ricostruzione di linea), sono d'interesse le cosiddette NURBS (Non Uniform Rational B-Splines). Queste curve sono costruite con le stesse identiche funzioni, costituite da polinomi a pezzi e comunemente adottate per le funzioni spline, dove il supporto di definizione varia adattativamente da un intervallo all'altro, spesso operando anche con un campionamento sequenziale: progressivo e selettivo, laddove l'approssimazione è rispettivamente deficitaria od eccessiva.
- Nel campo della rappresentazione tridimensionale delle superfici (chiuse) di oggetti, sono ancora aperti i problemi relativi alla ricerca di strutture semplici di riferimento, adatte anche ad oggetti fortemente concavi che non possono essere approssimati, come quelli convessi e quelli concavi stellati a sfere, cilindri, coni, eccetera. A riguardo, si ricordi che si chiamano stellati gli oggetti concavi contenenti corpi convessi le cui normali raggiungono, senza fuoriuscire, tutti i punti della superficie dei suddetti oggetti concavi in cui sono contenuti. Inoltre anche l'interpolazione di superfici con funzioni spline a supporto triangolare, non è di altrettanto semplice applicazione, quanto quella con funzioni spline a supporto quadrato o rettangolare [5].

[5] Problemi di continuità della superficie (o delle sue derivate prime e seconde) impediscono, in questo caso (come nel caso di una classica ricostruzione di superficie), l'uso di NURBS bidimensionali, obbligando ad operare con tecniche multilivello e multirisoluzione.

PARTE II – TRIANGOLAZIONE E TETRAEDRAZIONE

1. Generalità

Il problema della ricostruzione di superfici d'oggetti tridimensionali, a partire da un insieme di punti assegnati, è formulato Inizialmente con l'ipotesi di trattare oggetti la cui superficie possa essere descritta mediante una funzione:

$$z = f(x,y)$$

dove z è misurata lungo la normale ad un'assegnata superficie di riferimento, trascurando inoltre il problema dell'approssimazione, ovvero considerando la superficie ricostruita come interpolante i punti dati.
Diversi metodi di triangolazione consentono di suddividere l'insieme di punti in terne che definiscono una serie di triangoli capaci della completa descrizione della superficie. Allora data una serie di punti, in un piano, ed un punto arbitrario, giacente anch'esso nello stesso piano, ma non appartenente all'insieme dato, si desidera trovare il punto dell'insieme dato più vicino. La procedura dei diagrammi di Voronoi fornisce la soluzione cercata:

- costruire un diagramma che ripartisce il piano in regioni contenenti ciascuna un punto della serie;
- individuare la regione del diagramma che contiene il punto dato.

In letteratura, esistono diversi algoritmi che consentono di raggiungere la soluzione del problema; la scelta del metodo influisce e determina la velocità della costruzione delle regioni e dell'individuazione del punto nelle medesime.
Nel prosieguo, saranno trattati diversi algoritmi, basati sull'uso del diagramma duale di Delaunay, più semplice del diagramma primario di Voronoi, in quanto ha la capacità di separare, in ciascun problema, gli aspetti geometrici (coordinate, lunghezze, angoli, ecc.) da quelli topologici (lati, facce, ecc.).
Il rilevamento degli oggetti si effettua tradizionalmente per punti, scelti in modo opportuno, con una densità adeguata per descrivere gli oggetti stessi. La descrizione richiede che, oltre alle coordinate dei punti, sia nota anche la loro topologia, definita dalle modalità d'appartenenza dei punti alla superficie dell'oggetto in esame, ovvero dalla connessione esistente tra i punti stessi.
La connessione è solitamente ricostruita manualmente mediante schizzi ed osservazioni preliminari; tuttavia volendo procedere ad automatizzare questo passo, occorre preliminarmente definire i criteri per poterlo fare. L'obiettivo attuale diventa trovare relazioni tra punti di una lista di coordinate di punti rilevati, in modo da descrivere la topologia realmente esistente e ordinarla in una struttura utile per diverse applicazioni nella ricostruzione di superfici.
Qualora sia possibile una rappresentazione della superficie dell'oggetto nella forma: $z = f(x,y)$, dove x e y costituiscono una parametrizzazione di una superficie nota e regolare, la topologia dell'insieme di punti è legata alla connessione dei punti rappresentanti i parametri ordinatori dell'insieme stesso. Dal momento poi che tutti i punti appartengono alla stessa superficie, occorre un criterio aggiuntivo per definire la prossimità. Se la superficie è discretizzata in triangoli, il criterio sarà allora quello d'ottimizzazione, scelto per definire i triangoli stessi.

Uno dei criteri più famosi è quello fornito dalla soluzione di Voronoi, per la ricerca, dato un insieme di n punti ed un punto q esterno ad esso, del punto dell'insieme più prossimo al suddetto punto q.

Il diagramma di Voronoi di n punti in R^2 rappresenta il piano dividendolo in regioni contenenti i punti stessi.

- Ogni regione è associata ad un solo punto;
- tutti i punti, appartenenti alla regione, sono punti la cui distanza dal punto associato alla regione è minore, rispetto alla distanza da qualsiasi altro punto dell'insieme.

Gli algoritmi, basati sul diagramma di Vorornoi, hanno scarsa utilizzazione pratica, in quanto presentano alcuni svantaggi: complessità computazionale, difficoltà d'adattamento a casi degeneri, incapacità di trovare una soluzione sotto particolari condizioni (ad esempio, quando più di tre punti dell'insieme dato appartengono ad uno stesso cerchio). Tuttavia la principale difficoltà nella costruzione del diagramma di Voronoi sta nella determinazione dei suoi caratteri topologici, cioè nella descrizione delle relazioni tra vertici, lati e facce.
Boots fu tra i primi ad osservare che il calcolo del diagramma di Voronoi poteva essere semplificato mediante l'utilizzo del suo duale, il diagramma di Delaunay, che presenta il notevole vantaggio di separare i caratteri topologici di un dato problema da quelli geometrici.

2. Diagrammi di Voronoi e Delaunay

Dato un insieme di punti: $I_P = \{p_1, p_2, ...p_n\}$, la corrispondente regione di Voronoi è definita da:

$$F_1 = \{q \in R^2 : d(q, p_i) \leq d(q, p_j) \ \forall j \neq i\}$$

essendo: $d(q,p_i)$ la distanza euclidea e F_i la regione contenente tutti i punti $q \in R^2$ più vicini a p_i rispetto ad ogni altro punto dell'insieme I_p.
Nel caso in cui siano dati solo due punti, le regioni di Voronoi ad essi associate sono semplicemente i due semipiani aperti, limitati dall'asse del segmento che li congiunge.
In generale, dati più di due punti, la regione associata ad un punto è l'intersezione di tutti i semipiani che contengono il punto stesso e sono limitati dagli assi dei segmenti che congiungono il punto dato a tutti gli altri punti dell'insieme.
Queste regioni sono poligoni convessi i cui lati sono porzioni degli assi ed i vertici (eccetto i punti all'infinito) sono i circocentri dei triangoli definiti da tre punti dell'insieme (Fig. 2.1).
Come già detto in precedenza, si prenderà in considerazione il diagramma di Delaunay, duale del diagramma di Voronoi; questo si basa sulla suddivisione del piano in poligoni, dove:

- i vertici sono i punti dell'insieme dato;
- i lati sono i segmenti che collegano le coppie di punti che hanno ragioni di Voronoi confinanti (Fig. 2.2).

Inoltre si dimostra che:

- esiste sempre almeno un cerchio, passante per due punti collegati, che non contiene al suo interno altri punti dell'insieme dato;

- tre o più punti dati costituiscono i vertici di una faccia interna del diagramma di Delaunay, se e solo se esiste un cerchio passante per essi che non contiene al suo interno altri punti dati.

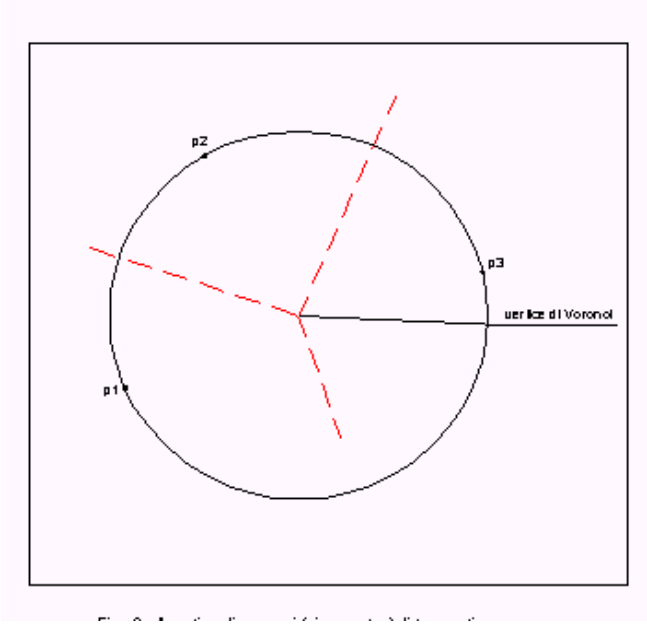


Fig. 2.1 – Il vertice di Voronoi (circocentro di tre punti

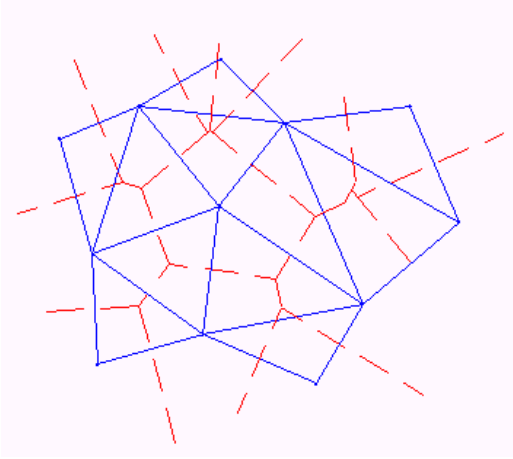


Fig. 2.2 – I diagrammi di Voronoi e Delaunay

Prima di descrivere i criteri e gli algoritmi, impiegati per la costruzione di questi diagrammi, resta da sottolineare come questi abbiamo una naturale estensione anche per problemi a tre o più dimensioni.

3. Triangolazione di Delaunay

L'operazione di divisione del piano in una serie di triangoli (aventi per vertici l'insieme di $n \geq 2$ punti), formando un poligono complementare al perimetro convesso dei suddetti punti, è detta triangolazione ed i punti dell'insieme dato sono detti vertici della triangolazione. Allora una triangolazione si dice di Delaunay, se e solo se per ogni lato ed ogni triangolo, valgono la suddette proprietà. Ogni triangolazione di n punti di cui m sul perimetro convesso, ha $2(n-1)-m$ triangoli e $3(n-1)-m$ lati (Fig. 3.1).

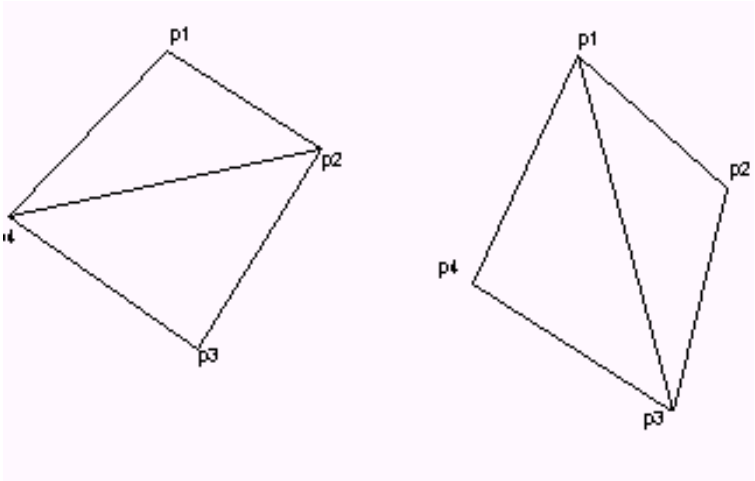


Fig. 3.1 – Le triangolazioni possibili per quattro punti

La triangolazione è uno strumento utile per risolvere problemi che riguardano l'analisi numerica, l'interpolazione e l'approssimazione di superfici, la ricostruzione di modelli 3D, ecc. Per la formazione di una triangolazione, è necessario far uso di due categorie di criteri:

CRITERI LOCALI:

- criterio della diagonale minima: è un criterio di molto semplice, ma poco indicato per una successiva costruzione di funzioni spline a base triangolare; consiste nello scegliere, tra le possibili triangolazioni, quella che forma quadrilateri con la minima diagonale maggiore;
- criterio di min – max angolo: proposto da Barnhill e Little; confronta gli angoli massimi, tra le possibili triangolazioni, e sceglie quella con il minimo angolo;
- criterio di max – min angolo: proposto da Lawson e molto usato, confronta gli angoli minimi, tra le possibili triangolazioni, e sceglie quella con il massimo angolo;
- criterio del cerchio: equivale al criterio del max – min; consiste nello scegliere la triangolazione in cui il cerchio circoscritto non contiene, al suo interno, un quarto vertice.

Una triangolazione si dice localmente ottimale, se ogni quadrilatero è stato triangolato con il criterio scelto.

CRITERI GLOBALI:

Questi criteri considerano l'intera triangolazione, costituita dall'insieme dei punti dati, ed operano un confronto fra le triangolazioni possibili, secondo un certo criterio (solitamente i criteri max – min e min –max). Di conseguenza, una triangolazione si dice globalmente ottimale, se è ottimamente triangolata rispetto alle altre possibili triangolazioni. Una triangolazione globalmente ottimale è anche localmente ottimale

Gli algoritmi per effettuare una triangolazione possono essere suddivisi in tre categorie distinte:

- algoritmi di post–ottimizzazione: si costruisce prima una triangolazione di tentativo che è poi ottimizzata con uno dei criteri di ottimizzazione locali;
- algoritmi iterativi: costruito il primo triangolo, si aggiunge ogni volta un altro punto, secondo il criterio di ottimizzazione locale scelto;
- algoritmi di suddivisione e annessione: l'insieme dei punti dati è diviso in sottoinsiemi, in ciascuno di essi si costruisce una triangolazione con un criterio locale di ottimizzazione; infine si considerano contemporaneamente tutte le triangolazioni ottenute.

Questi tipi di algoritmi, benché differenti fra loro, non differiscono per complessità computazionale.

4. Un algoritmo per la triangolazione di Delaunay con linee di rottura

Esistono diversi tipi di algoritmi per effettuare una triangolazione in 2D, ma quelli che tengono conto anche della presenza di linee di rottura (in inglese: break line o skeleton line) sono in numero minore. Un algoritmo diretto per la triangolazione, valido in entrambi i casi è stato proposto da Fang. Data la presenza di linee di rottura, vengono introdotti due passi ulteriori:

- costruzione di una maglia uniforme;
- selezione di punti appartenenti alle linee di rottura che entrano a far parte dell'insieme dato.

La costruzione di un perimetro convesso è automatica e contemporanea alla ricerca dei punti per la determinazione dei triangoli. L'algoritmo si compone delle seguenti fasi:

- pre – elaborazione dei dati;
- costruzione del primo triangolo;
- assemblaggio dei triangoli.

Di conseguenza, i vantaggi presentati dall'algoritmo sono una riduzione della complessità computazionale. Dato un insieme di n punti, aventi coordinate (x, y), ed una serie di m linee di rottura: $L = [(x,y)_k,(x,y)_l]$, e k, $l \leq n$, si vuole costruire una struttura che permetta una ricerca rapida dei punti e delle linee di rottura[6].

Costruzione del primo triangolo: questo procedimento consiste nell'avvio della triangolazione vera e propria. Si inizia scegliendo un punto iniziale da cui far partire il primo lato (base del primo triangolo). Come punto iniziale è consigliabile scegliere un punto che si trovi approssimativamente al centro della maglia.
Il primo lato è costruito a partire dal primo punto, scegliendo il punto più vicino. Per fare ciò, è necessario effettuare una verifica su tutti i punti delle celle intorno a quella a cui appartiene il punto iniziale, per individuare quello di minima distanza.
Per costruire il triangolo è necessario trovare il terzo punto che sarà scelto (per convenzione) a destra della linea che costruisce il primo lato. Si deve poi costruire il cerchio passante per i tre punti e verificare che, al

---

[6] La struttura scelta è un struttura a maglia regolare e, come primo passo, si devono inglobare i punti nella struttura, adottando un valore numerico arbitrario di tolleranza che permetta anche di risolvere conflitti, qualora i punti siano giacenti sulle linee guida:

$$X_{min} = X_{min} - TOLL \qquad X_{max} = X_{max} + TOLL \qquad Y_{min} = Y_{min} - TOLL \qquad Y_{max} = Y_{max} + TOLL$$

E' poi possibile calcolare le dimensioni della maglia, come numero di celle nelle due direzioni:

$$X_res = int\left(\frac{X_{max} - X_{min}}{size}\right) + 1 \qquad Y_res = int\left(\frac{Y_{max} - Y_{min}}{size}\right) + 1$$

dove: int è un operatore che converte un numero reale in un numero intero, e $size$ è così definito:

$$size = \sqrt{\frac{(X_{max} - X_{min})(Y_{max} - Y_{min})}{n}}$$

Allora le coordinate dei punti, riferite all'origine e standardizzate, diventano:

$$grid_x_i = \frac{x_i - X_{min}}{size} \qquad grid_y_i = \frac{y_i - Y_{min}}{size}$$

e la cella della maglia, in posizione (i_cell, j_cell), è caratterizzata dalle coordinate:

$$i_cell = int\,(grid_x_i) \qquad j_cell = int\,(grid_y_i)$$

Note le coordinate normalizzate e la cella corrispondente, si esegue una verifica sui punti presenti nella stessa cella. Se il punto normalizzato coincide con uno presente nella cella corrispondente, è scartato, altrimenti è inserito nella stessa cella. Alla fine, il processo associa ogni punto dato con una ed una sola cella. Questo è vero anche per punti che si trovano lungo le linee o agli angoli della griglia della struttura.
Le linee di rottura costituiscono invece un problema particolare ed un po' più complesso e sono trattate come segue:

- per ogni linea di rottura, si considerano tutte le celle da essa intersecate;
- per ogni cella coinvolta, si controlla se la linea di rottura passa per i punti in essa presenti; se questo avviene, la linea è spezzata, e la tabella della linea è aggiornata ed associata alla cella.

Le linee di rottura sono mantenute in due diverse strutture di dati: quella della cella e quella delle linee di rottura; ciò è necessario per l'aggiornamento automatico, durante l'inserimento dei punti dell'insieme.

suo interno, non cadano altri punti dell'insieme dato. Se quest'ultima condizione non è verificata, si sceglie quello con angolo massimo nel vertice e si aggiorna il cerchio. In questa fase, non vengono prese in considerazione eventuali linee di rottura.

Assemblaggio dei triangoli: durante questa fase, è necessario fare bene attenzione che non si formino buchi (triangoli mancanti), oppure si creino ponti (lati mancanti). Il procedimento consiste nel costruire ed aggiornare una lista, detta di perimetro, contenente i lati appartenenti al poligono formato fino al quel momento. Essa è utile per controllare la completezza del procedimento.
Si parte dal primo lato del primo triangolo, poi si passa agli altri due procedendo in senso antiorario. A partire dal secondo lato, per ciascuno si controlla se, alla sua destra, è presente un punto dell'insieme dato. In caso affermativo, si costruisce il triangolo di Delaunay e si eliminano gli altri due lati; in caso contrario, il lato appartiene al perimetro del poligono ed è comunque eliminato. Il procedimento prosegue fino ad esaurimento della lista.
Se esistono linee di rottura e se il cerchio ne interseca una, si aggiorna la lista con i lati che collegano i punti di intersezione tra la linea e il cerchio, e si eliminano tutti quei lati che hanno un vertice sulla circonferenza. Per garantire la correttezza della procedura, se un lato si sovrappone ad un altro, si eliminano entrambi. Non è consentito eliminare un lato, a partire dal quale si è costruito un nuovo triangolo, se il terzo vertice appartiene al poligono stesso.
Prima di proseguire alla generalizzazione del problema, estendendo le tecniche introdotte al caso tridimensionale e rimuovendo l'ipotesi che la superficie possa essere descritta nella forma:

$$z = f(x,y)$$

dove una coppia di coordinate funge da parametro ordinatore, si sottolinea ancora come questi metodi di interpolazione di superfici, richiedano di suddividere i punti in cui la superficie è nota in terne di punti. Dato il numero notevole di possibili combinazioni, è necessario scegliere un criterio per effettuare la triangolazione. La triangolazione di Delaunay sceglie quei triangoli che hanno come vertici i punti dell'insieme dato per i quali il cerchio circoscritto non contiene altri punti. Esistono diversi algoritmi per la costruzione di questo diagramma, la scelta è effettuata in base a criteri di ottimizzazione. Nell'interpolazione delle superfici, un posto notevole è occupato dal trattamento delle linee di rottura. L'algoritmo proposto da Fang è capace di tenere conto anche della loro presenza; si noti tuttavia, come l'assenza di linee di rottura garantisca una ottimizzazione locale, non sempre assicurata altrimenti[7].

[7] Una verifica possibile consiste nel controllare che l'algoritmo realizzi effettivamente una triangolazione di Delaunay, ovvero che il cerchio circoscritto, ad ogni triangolo costruito, sia tale che non sia presente alcun vertice, al suo interno dalla parte sinistra di ogni lato.

- Durante la costruzione del primo triangolo, l'algoritmo sceglie il punto B del primo lato, come il più vicino al punto A. Il terzo vertice, il punto C, è scelto tra quelli che stanno a destra del primo lato, in modo da avere l'angolo ACB massimo, rispetto a quello di tutti gli altri punti. Per questo motivo, il lato AC è maggiore o al più uguale al lato AB, l'angolo ACB è minore di 90 gradi e l'arco (che non contiene il punto C) è più piccolo di un semicerchio. Se accadesse che un punto D dell'insieme dato si trovi all'interno del cerchio, passante per i tre vertici e giacente a sinistra del lato AB, ne conseguirebbe che il lato AD è più piccolo del lato AB, il che non può accadere, poiché è in contraddizione con il fatto che B è, per costruzione, il punto più vicino ad A (Fig. 4.1).
- Dopo la costruzione del primo triangolo, l'algoritmo continua la triangolazione con la costruzione del secondo triangolo. In questa fase, viene scelto un punto D, come era stato scelto il punto C del primo triangolo. Per dimostrare che il punto B, appartenente al primo triangolo, non si trovi all'interno del cerchio passante per A,C,D, si traccia l'asse del lato AC e si trova il punto Q di intersezione con la circonferenza. L'arco complementare a A,Q,C, si trova all'interno del cerchio che passa per i vertici del triangolo ABC. Pertanto il punto B non può essere all'interno del cerchio passante per i vertici del triangolo ACD.

5. Generalizzazione del problema

Una prima generalizzazione fa passare dalla connessione di terne di punti (che generano triangoli) alla connessione di quaterne di punti (che generano tetraedri), dove le condizioni per la scelta dei punti devono rispettare il criterio della sfera.

Una tetraedrazione di Delaunay riempie l'intero involucro convesso di punti con tetraedri ed un algoritmo di tetraedrazione è stato proposto da Joe. Talvolta tutti i punti dell'insieme dato non compaiono sull'involucro, poiché alcuni sono nascosti all'interno; in questo caso, è più opportuno l'utilizzo di un algoritmo di tetraedrazione con regioni. Per eliminare poi i tetraedri, esterni alla superficie, che nascondono i punti della superficie stessa, si utilizza un algoritmo basato su tecniche euristiche.

Dato un insieme di punti significativi, appartenenti alla superficie di un oggetto nello spazio 3D, si desidera costruire un modello geometrico della superficie stessa che le sia il più prossimo possibile. Al fine di costruire detto modello, è necessario conoscere le relazioni fra i punti dell'insieme dato.

- Se i punti sono preordinati, si possono facilmente dedurre connessioni tra i punti e si può poi effettuare la ricostruzione della superficie.
- Se invece non si hanno informazioni a priori, è necessario cercare dei criteri che possano fornire regole per individuare le possibili relazioni tra punti.

Diversi sono i criteri esistenti e, non tutti, danno come risultato una tetraedrazione di Delaunay. Un criterio adatto allo scopo prevede di collegare i punti con tetraedri e richiede che i punti della serie non siano tutti sullo stesso piano. Un vantaggio, presentato da questo tipo di approccio, consiste nel fatto che l'involucro convesso formato dai tetraedri è facilmente modificabile e permette di determinare il volume totale.

Un modo semplice per costruire la tetraedrazione è costituito da un algoritmo iterativo, partendo da un tetraedro ed aggiungendo, ogni volta, un punto fino a costruire la tetraedrazione completa. Si segue allora un criterio locale (convergente poi al globale): il criterio della sfera, per verificare che la circosfera di quattro vertici di ogni tetraedro non contenga, al suo interno, punti dell'insieme dato.

L'algoritmo proposto da Joe utilizza trasformazioni locali per costruire la tetraedrazione di Delaunay ed una sua estensione utilizza una trasformazione locale, capace di rispettare automaticamente il criterio della sfera, restituedo sempre una tetraedrazione di Delaunay. Come nel caso di una triangolazione, anche nel caso di una tetraedrazione il tempo empirico impiegato per il calcolo, per un insieme di n punti, casualmente disposti, è non superiore a $O(n^2)$.

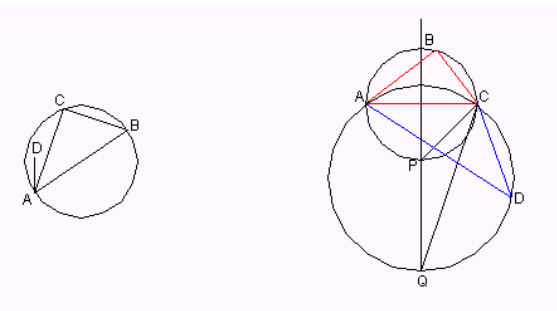


Fig. 4.1 – Verifica del criterio di Delauny

Infine l'algoritmo garantisce la realizzazione della triangolazione di Delaunay, poiché la procedura per la costruzione dei successivi triangoli è proprio la stessa utilizzata per la costruzione del secondo triangolo.

6. Tetraedrazione di Delaunay

Si definisce tetraedro l'involucro convesso di quattro punti non complanari in R^3; i quattro punti sono detti vertici e, poiché non complanari, determinano un'unica sfera: la circosfera del tetraedro. La tetraedrazione di un insieme dato di punti consiste nella costruzione di una serie di tetraedri non intersecatisi che hanno come vertici i punti dell'insieme dato.

Una tetraedrazione è detta di Delaunay, se soddisfa il criterio della sfera, ovvero se la sfera passante per i vertici di ogni tetraedro non contiene, al suo interno, punti dell'insieme dato. Nel caso in cui non più di quattro punti dell'insieme dato si trovino sulla sfera che soddisfa il criterio, la tetraedrazione di Delaunay è univocamente definita; in caso contrario occorrerà controllare che la tetraedrazione della serie non sia tale da generare intersezioni di tetraedri. In questo caso, si verifica una delle due condizioni:

- una sfera, passante per i tre vertici di un triangolo, può passare per un quarto punto ad esso esterno;
- una sfera, di raggio infinito, degenera in un semispazio vuoto: il triangolo giace sull'involucro convesso.

Gli algoritmi, già esposti per la triangolazione 2D, possono essere estesi anche al caso tridimensionale. La maggior parte degli algoritmi di tetraedrazione, finora sviluppati, sono di tipo iterativo. In generale, essi partono dalla definizione di un primo tetraedro e, ogni volta, aggiungono un punto, fino a formare la tetraedrazione completa.

L'algoritmo proposto da Joe è basato su un criterio locale (simile a quello di scambio proposto da Lawson) e consiste in scambi di facce che permettano di avere sempre una tetraedrazione Delaunay. La procedura di trasformazione locale è basata sulle possibili configurazioni di cinque punti e, in particolare, sul numero di questi giacenti su uno stesso piano:

- nessuna della possibili quaterne (ricavabili dai cinque punti) giace sullo stesso piano;
- almeno quattro dei cinque punti giacciono sullo stesso piano.

Nel primo caso, le configurazioni possibili sono due e la tetraedrazione può comprendere due, tre o quattro tetraedri (Fig. 6.1):

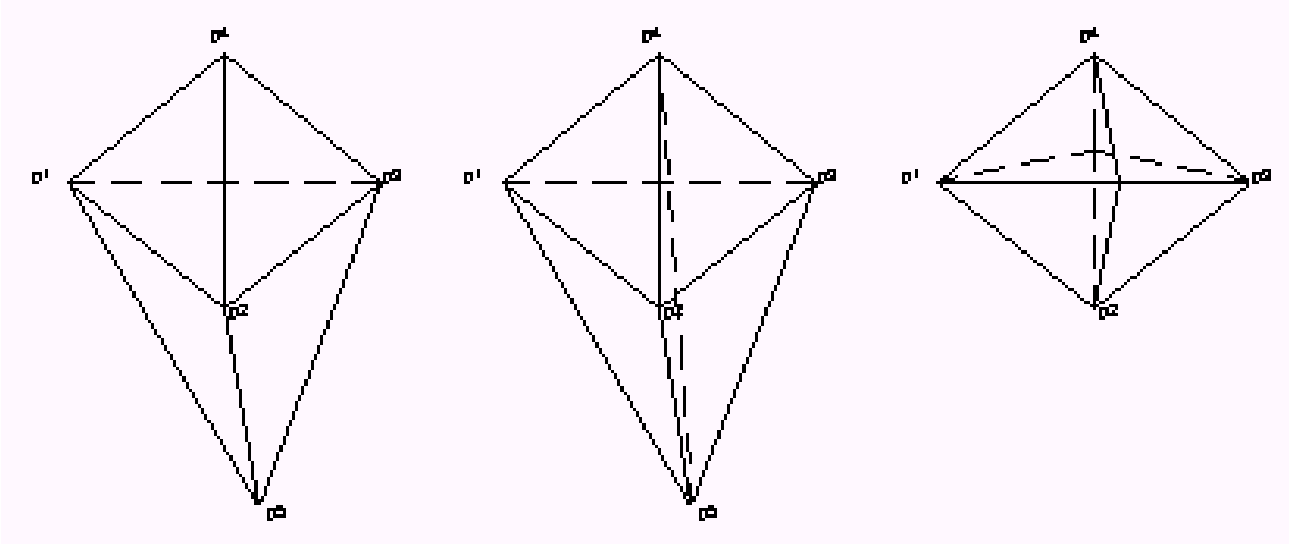


Fig. 6.1 – Le configurazioni del 1° caso (configura zioni 1a, 1b e 2)

- nella prima configurazione, tutti i punti giacciono sull'involucro convesso;
- nella seconda, uno dei cinque punti non giace sull'involucro.

In accordo con la numerazione della Fig. 6.1, se il lato $p_4\ p_5$ interseca la faccia $p_1\ p_2\ p_3$, allora i due tetraedri della configurazione 1a hanno i vertici: ($p_1\ p_2\ p_3\ p_4$) e ($p_1\ p_2\ p_3\ p_5$), ed i tre tetraedri della configurazione 1b hanno i vertici: ($p_1\ p_2\ p_4\ p_5$), ($p_1\ p_3\ p_4\ p_5$) *e* ($p_2\ p_3\ p_4\ p_5$). Invece la tetraedrazione della configurazione 2 comprende quattro tetraedri, aventi vertici: ($p_1\ p_2\ p_3\ p_4$), ($p_1\ p_2\ p_3\ p_5$), ($p_1\ p_2\ p_4\ p_5$) *e* ($p_1\ p_3\ p_4\ p_5$).
Nel secondo caso, dove i quattro dei cinque punti giacciono sullo stesso piano (Fig. 6.2), si possono avere tre configurazioni, a seconda della forma del quadrilatero ($p_1\ p_4\ p_2\ p_5$):

- la configurazione 3 accade, quando il quadrilatero è convesso;
- la configurazione 4, quando il quadrilatero degenera in un triangolo;
- la configurazione 5, quando il quadrilatero è concavo.

La tetraedrazione della configurazione 3 comprende due tetraedri, aventi vertici: ($p_1\ p_2\ p_3\ p_5$) , ($p_1\ p_2\ p_3\ p_4$) (configurazione 3a) o due tetraedri, aventi vertici: ($p_1\ p_3\ p_4\ p_5$) e ($p_2\ p_3\ p_4\ p_5$) (configurazione 3b). La tetraedrazione della configurazione 4 comprende due tetraedri, aventi vertici: ($p_1\ p_2\ p_3\ p_4$) e ($p_1\ p_2\ p_3\ p_5$), mentre la tetraedrazione della configurazione 5 comprende tre tetraedri di cui uno cavo, aventi vertici: ($p_1\ p_2\ p_3\ p_4$), ($p_1\ p_2\ p_3\ p_5$) e ($p_1\ p_3\ p_4\ p_5$).

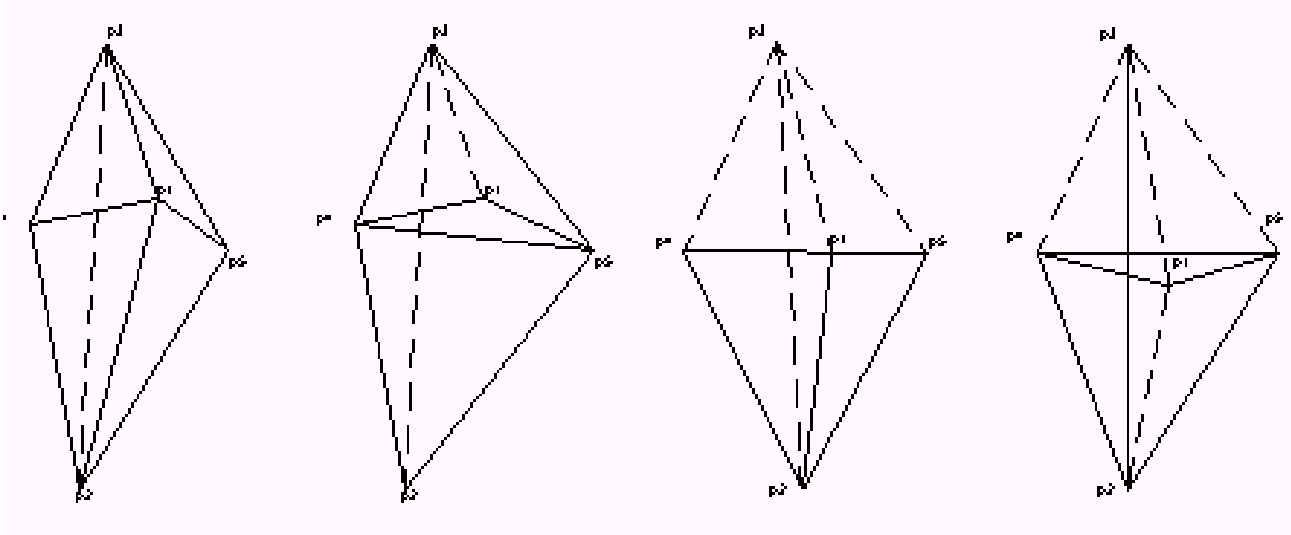


Fig. 6.2 – Le configurazioni del 2°caso (configura zioni 3a, 3b, 4 e 5)

La trasformazione locale consiste nella trasformazione delle tetraedrazioni delle configurazioni 1 e 3, ovvero nello scambio della faccia ($p_1\ p_2\ p_3$) con le facce ($p_1\ p_4\ p_5$) e ($p_2\ p_4\ p_5$) o viceversa nella configurazione 1, e della faccia ($p_1\ p_2\ p_3$) con la faccia ($p_3\ p_4\ p_5$) o viceversa nella configurazione 3.

7. Un algoritmo per la tetraedrazione Delaunay

Un algoritmo che rispetta il criterio della sfera ed fa uso della trasformazione locale è stato proposto da Joe e si compone delle seguenti fasi:

- pre – elaborazione dei dati: riordino dei punti dell'insieme dato, secondo l'ordine lessicografico delle coordinate;
- costruzione del primo tetraedro: scelta di quattro punti non complanari e costruzione del primo tetraedro;
- assemblaggio dei tetraedri: costruisce la tetraedrazione vera e propria, riguardando i restanti $n - 4$ punti.

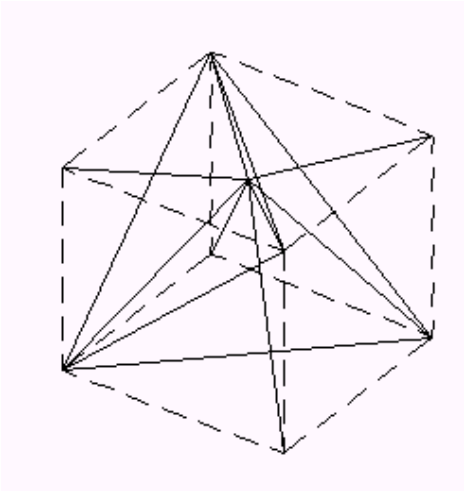


Fig. 7.1 –Tetraedrazione di Delauny per 9 punti dati

8. L'algoritmo in pseudo – codice scritto da Floriani e Bruzzone

Si definiscono inizialmente due tetraedrazioni di Delaunay successive: T_{i-1}^{D}, T_{i}^{D}, formate rispettivamente da i–1 ed i punti dell'insieme dato e soddisfacenti entrambe il criterio della sfera; inoltre sono date due liste vuote Q e L, per memorizzare le facce da analizzare. L'algoritmo si avvia con le seguenti operazioni[8]:

- riordino dei punti $v_1, v_2, \ldots, v_n$, secondo un ordine lessicografico; controllo della complanarità dei primi quattro punti e, in caso di esito positivo, ricerca di un punto adatto a formare il primo tetraedro;
- costruzione del primo tetraedro e calcolo del suo centroide: $w = \frac{(v_1 + v_2 + v_3 + v_4)}{4}$
- inizializzazione delle liste Q e L.

for $i := 5$ **to** n **do**

sia T_{i-1}^{D} una triangolazione corrente dei primi i–1 vertici

for ogni faccia sul contorno $v_a v_b v_c$ di T_{i-1}^{D} **do**

if v_i è sul lato opposto di $v_a v_b v_c$ rispetto a w **then**

8 La base di dati in ingresso contiene le coordinate dei vertici e l'algoritmo presentato produce una lista dei tetraedri costruiti, caratterizzata:

- dalle coordinate del centro della sfera passante per i vertici di ciascun tetraedro;
- dagli indici: dei vertici di ciascun tetraedro;
- dagli indici dei tetraedri adiacenti ad ogni il tetraedro, identificati tramite il vertice opposto alla faccia comune.

aggiungere la faccia $v_a v_b v_c$ alla coda della lista Q

endif

endfor

while Q non è vuoto **do**

rimuovere la faccia v_a, v_b, v_c da Q

if v_a, v_b, v_c è nella triangolazione **then**

trovare il tetraedro $v_a v_b v_c v_d$ incidente sulla faccia $v_a v_b v_c$

if la circosfera di $v_a v_b v_c v_d$ non contiene v_i al suo interno **then**

inserire nella lista L

else cancellare sia $v_a v_b v_c$ che $v_a v_b v_c v_d$ dalla triangolazione

for ogni faccia $v_a v_b v_d$, $v_a v_c v_d$, $v_b v_c v_d$ **do**

if la faccia è una faccia esterna per T_{i-1}^D e T_i^D **then**

inserire nella lista Q

else if faccia è già in Q **then**

cancellare dalla triangolazione

else aggiungere in coda alla Q

endif

endfor

endif

endif

while L non è vuota **do**

prendere $v_a v_b v_c$ dalla L

if $v_a v_b v_c$ è una faccia sul contorno T_{i-1}^D e con v_c sullo stesso piano **then**

cancellare la $v_a v_b v_c$ dalla triangolazione

else aggiungere il tetraedro $v_a v_b v_c v_i$ (ed i tre lati coinvolti con v_i) alla triangolazione

endif

endwhile

endfor

9. Costruzione della superficie esterna

La tetraedrazione Delaunay forma un poliedro convesso, ovvero riempie l'interno dell'involucro convesso dell'insieme di punti dati. A seconda della distribuzione dei punti e della forma dell'oggetto:

- in alcuni casi, tutti i tetraedri hanno almeno una faccia sul contorno;
- in altri casi, solo una parte di essi verifica tale condizione,

ed anche i punti dell'insieme dato, essenziali per la definizione della superficie esterna, possono tutti trovarsi sull'involucro convesso, oppure soltanto alcuni.

Nel primo caso, dove tutti i punti dell'insieme dato si trovano sull'involucro convesso, la triangolazione Delaunay è una rappresentazione volumetrica dell'oggetto. Infatti l'oggetto è costituito da una serie di tetraedri ed il contorno di questa serie, cioè l'involucro convesso, è il poliedro che approssima, al meglio, la superficie dell'oggetto, individuando la minima distanza fra il poliedro e l'oggetto, rispetto ad ogni altra possibile tetraedrazione.

Il secondo caso è invece più complesso, in quanto solo alcuni punti dell'insieme dato giacciono sull'involucro convesso e la superficie ottenuta non rappresenta la superficie dell'oggetto, se non con una approssimazione, in generale, non accettabile. Pertanto occorre trovare un poliedro concavo, sempre con facce triangolari, che abbia per vertici tutti i punti dell'insieme dato. Purtroppo non esiste una soluzione generale, per qualsiasi problema di questo tipo; allora occorre adottare tecniche euristiche.

10. Algoritmo di scultura

L'algoritmo di scultura si avvale di tecniche euristiche per costruire una superficie esterna, passante per tutti i punti dell'insieme dato, e richiede informazioni a priori sulle relazioni tra i punti, ovvero una prima tetraedrazione di Delaunay.

Esso classifica i tetraedri in tetraedri eliminabili e non eliminabili, in relazione al numero delle facce del tetraedro appartenenti alla superficie del poliedro chiuso. I tetraedri che si trovano sul poliedro possono avere una, due o tre facce: quelli con tre facce non sono eliminabili.

L'algoritmo è un sequenziale, cioè procede all'eliminazione di un tetraedro dopo l'altro, a condizione che ad ogni passo il contorno individuato soddisfi la definizione di poliedro chiuso. Questo può essere facilmente verificato, creando tutte le possibili configurazioni (Fig. 10.1); successivamente si procede secondo la regola in base alla quale gli unici tetraedri eliminabili rispondono ad una delle seguenti condizioni:

- i tetraedri hanno una faccia, tre lati e tre punti sul poliedro chiuso che, ad un dato passo, costituisce la superficie esterna di miglior approssimazione;
- i tetraedri hanno due facce, cinque lati e quattro punti sullo stesso poliedro.

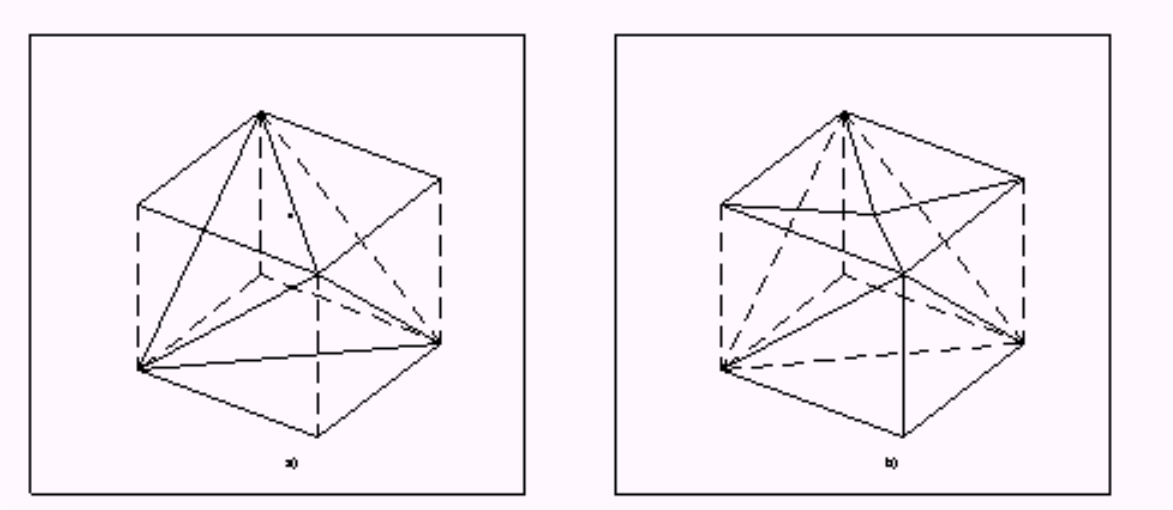


Fig. 10.1 – La superficie esterna di un poliedro: involucro convesso ed algoritmo di scultura

Secondo la regola sopraccitata, i tetraedri eliminabili sono solamente quelli dei casi d ed i della Fig. 10.2.

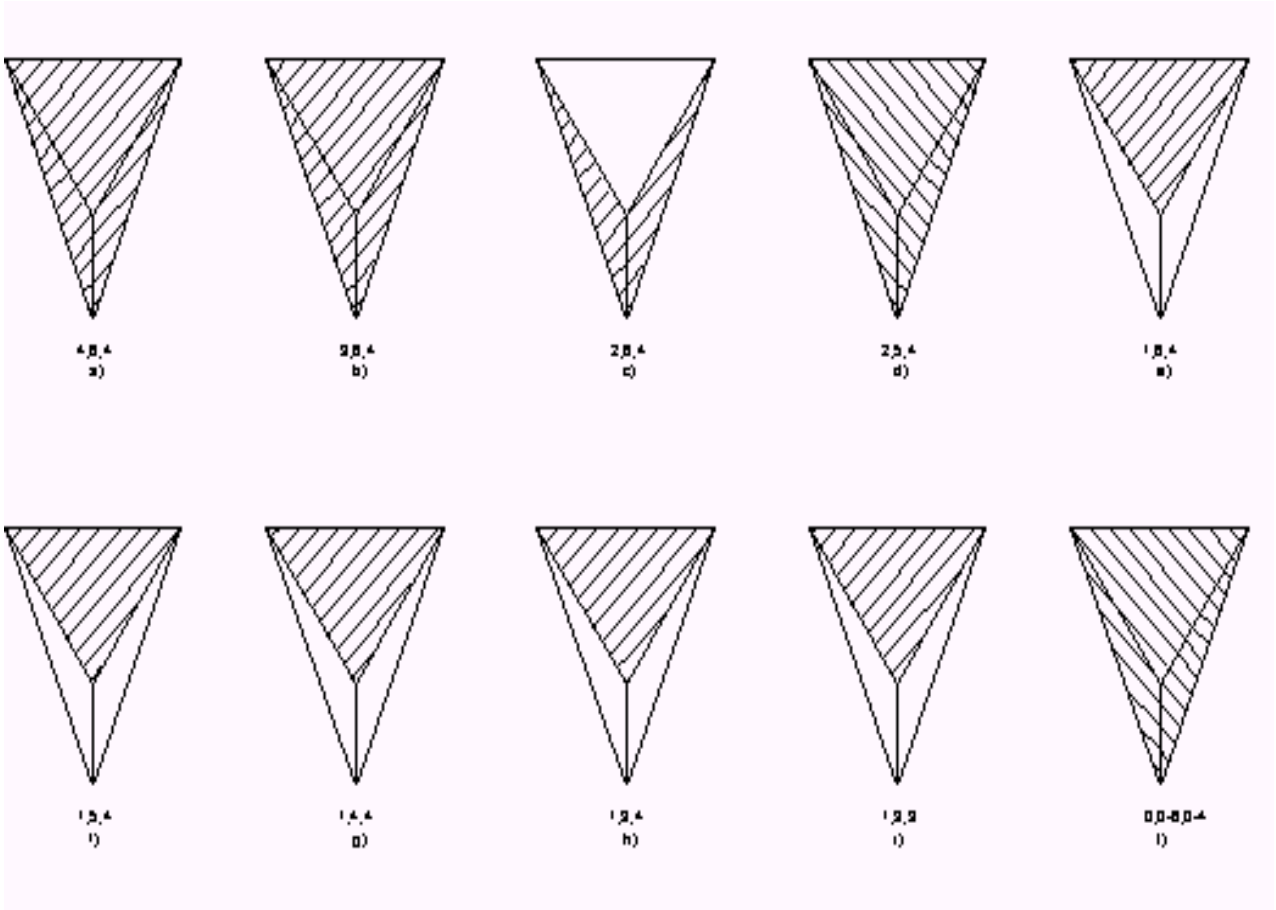


Fig. 10.2 – Le possibili configurazioni di tetraedri (i tetraedri eliminabili sono solamente quelli dei casi d ed i)

La classificazione dei tetraedri, candidati all'eliminazione, avviene associando, ad ogni tetraedro T_i , un opportuno valore $V(T_i)$, ad esempio:

- la distanza massima tra ciascuna faccia di T_i del poliedro e la porzione della circosfera associata a T_i esterna al poliedro;
- l'angolo solido A/R^2 (cioè l'area della faccia di T_i sul poliedro rapportata al quadrato del raggio della circosfera associata a T_{ii});
- il rapporto fra la somma dei raggi dei cerchi circoscritti ai tre vertici di ciascuna faccia ed il raggio della circosfera associata a $T_{i)}$.

Il procedimento termina, quando tutti i punti diventano i vertici del poliedro, e nessun tetraedro ha un valore associato maggiore degli altri, ovvero $V(T_i) = V(T_{i-1})$.

Ulteriori approfondimenti si possono trovare nei lavori di Veltkamp. Infatti a partire da una tetraedrazione Delaunay prodotta con l'algoritmo proposto da Joe, esso costruisce la superficie esterna, con tutti i punti dell'insieme dato, e consente all'utente di scegliere una fra quattro possibili soluzioni, derivate da tecniche euristiche, per l'eliminazione dei tetraedri ricoprenti i punti dati.

11. Tetraedrazione Delaunay con presenza delle regioni

La generalizzazione di una triangolazione di Delaunay, in presenza di linee di rottura, è una tetraedrazione Delaunay, in presenza di regioni (quantunque piane e convesse) di discontinuità, benché la sua costruzione non sia sempre possibile, come evidenziato da Schönhard, O'Rourke e Hazelwood.

Una via per eseguire la tetraedrazione, in presenza di regioni, aventi vertici estremi coincidenti con alcuni punti dell'insieme dato, costruisce tetraedrazioni, regione per regione, con l'algoritmo di scultura. Per questo, bisogna innanzitutto determinare, per ogni regione, i tetraedri intersecati e poi suddividere ogni regione in sottoregioni, ciascuna delimitata da questi tetraedri.

L'introduzione dei nuovi tetraedri avviene in modo da conservare la forma della regione. Nella Fig. 11.1, è presentato un esempio di tetraedrazione, in un insieme di 5 punti $I_p = \{p_1, p_2, p_3, p_4, p_5\}$ con due vincoli costituiti da due regioni regioni $R_1 = \{p_2, p_3, p_4\}$ e $R_2 = \{p_1, p_2, p_3\}$. Al fine di mantenere la forma di queste regioni, è allora necessario introdurre un nuovo punto (p_6).

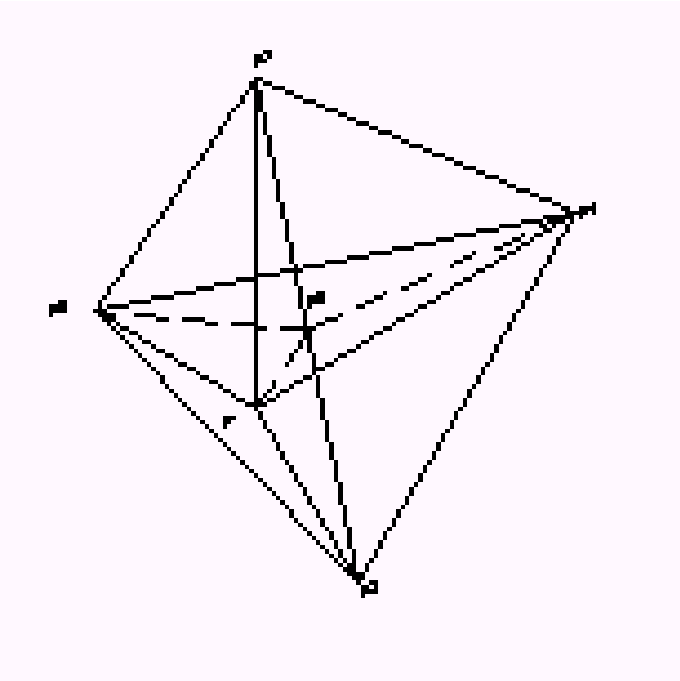


Fig. 11.1 – Introduzione di nuovi punti

12. Ottimizzazione della triangolazione della superficie esterna

L'ottimizzazione della triangolazione della superficie esterna, rilevata con un insieme di punti, consiste nella sua semplificazione, ovvero nella costruzione di un'altra triangolazione della esterna, con un numero inferiore di punti, comunque sufficienti a rappresentarla. Utile in molti casi (ad esempio, per compattare i dati ed elaborali più facilmente), consiste nell'eliminazione di alcuni triangoli vicini che formano una superficie pressoché piana (rispetto ad una certa tolleranza prefissata).

Due diversi approcci portano a minimizzare l'energia, supponendo la superficie esterna una membrana deformabile, oppure a stimare la curvatura dei triangoli attorno ad un dato vertice. Nel primo approccio (Hoppe), la triangolazione adottata è definita dalla posizione dei vertici V e dalla sua topologia K, definita dalle loro connessioni. L'obiettivo è minimizzare una funzione composta da tre addendi:

$$E(K,V) = E_{dist}(K,V) + E_{rep}(K) + E_{spring}(K,V)$$

dove:

- $E_{dist}(K,V) = \sum_i d^2(x_i, \phi_v(|K|))$, misura l'approssimazione realizzata dalla triangolazione semplificata ed è dato dalla somma dei quadrati delle distanze dei punti dell'insieme dato dalla superficie esterna ottenuta (nella triangolazione iniziale, questo termine è nullo, poiché essa passa per tutti i punti dell'insieme dato);
- $E_{rep}(K) = c_{rep} * m$, serve a penalizzare le configurazioni con un numero eccessivo di vertici costituenti una triangolazione (esso è un massimo nella triangolazione iniziale e diminuisce via, via con la sua semplificazione), dove il parametro c_{rep} è un indice di bilanciamento tra la fedeltà geometrica della triangolazione e la parsimonia della sua rappresentazione;
- $E_{spring}(K,V) = \sum_{\{j,k\} \in K} k \left\| v_j - v_k \right\|^2$, è un termine aggiuntivo necessario per garantire l'esistenza di un minimo ragionevole (infatti il funzionale, con i soli primi due addendi, non porta sempre a risultati soddisfacenti).

Il secondo approccio (Hannan), stima la curvatura dei triangoli attorno a ciascun vertice, per determinare pesi opportuni dei triangoli e procedere poi all'eliminazione di quelli con basse curvature. Essi sono sostituiti con nuovi punti, giacenti sul piano locale approssimante, e le regioni influenzate da questi triangoli sono triangolate nuovamente. In modo iterativo:

- si inizia con l'eliminazione del triangolo di peso minimo;
- si procede poi, dopo una nuova triangolazione, ad una successiva eliminazione;
- si continua, allo stesso modo, prendendo sempre in considerazione un triangolo alla volta.

Alla fine, avendo rimosso tutti i triangoli con bassa curvatura, si ottiene una superficie con grandi triangoli che, inglobando i triangoli rimossi, è comunque prossima alla superficie esterna.
La triangolazione di un insieme di punti, non tutti complanari ed appartenenti alla superficie esterna di un oggetto tridimensionale, può essere effettuata con due approcci differenti, basati rispettivamente sull'appartenenza ad una superficie o sul riempimento del volume. Il secondo approccio è più generale, in quanto applicabile comunque siano dati i punti rilevati e l'algoritmo proposto da Joe permette di costruire una tetraedrazione Delaunay. D'altra parte, poiché l'involucro convesso può talvolta contenere solo alcuni punti dell'insieme dato, occorre rimuovere i tetraedri che ricoprono alcuni punti. In questi casi, pur non esistendo una soluzione unica, l'algoritmo di scultura fornisce spesso una soluzione accettabile. Infine per una riduzione mirata del numero dei triangoli si può procedere ad una ottimizzazione della triangolazione.

APPENDICE A – RICOSTRUZIONE DIRETTA DELLA SUPERFICIE ESTERNA

Dato un insieme di punti nello spazio 3D, appartenente alla superficie esterna di un oggetto solido, e supponendo la base di dati organizzata in una struttura tale che i punti appartengano a linee di contorno di

sezioni piane (ad esempio, curve di livello), anche se non necessariamente tutte parallele tra loro, ma sempre tali che le sezioni non si intersechino, si vuole ricostruire direttamente la DSM dell'oggetto rilevato. Un metodo, operante con una tecnica di tipo euristico, esatto per sezioni convesse o concave – stellate (altrimenti il metodo introduce approssimazioni locali), si compone delle seguenti fasi:

- elaborazione dei dati;
- predizione dei nuovi punti (ottimizzazione dei dati);
- ricerca delle corrispondenze fra i punti di due piani limitrofi;
- scelta di un criterio di triangolazione.

La fase di ricostruzione della DSM può essere eseguita con metodi deterministici, integrati eventualmente da metodi stocastici, e si compone dei seguenti passi:

- costruzione della topologia dei punti mediante triangolazione;
- ricostruzione della superficie che, per caratteristiche di levigatezza, meglio si avvicina a quella dell'oggetto rilevato.

I metodi, già descritti in precedenza, sono utili per costruire la triangolazione e la tetraedrazione dei punti. Nel caso tutta la serie dei punti giaccia sull'involucro convesso ottenuto, si ottiene un modello prossimo all'oggetto; al contrario, quando solo alcuni punti giacciono sull'involucro convesso, la superficie ottenuta rappresenta la superficie dell'oggetto con una approssimazione, in generale, non accettabile. Allora in mancanza di una soluzione unica, occorre cercare un poliedro (chiuso) con facce triangolari, avente per vertici tutto l'insieme di punti, mediante tecniche euristiche.
Tuttavia se i punti sono preordinati in una struttura dati per linee di contorno, può essere convienente usare tecniche dirette, sensibili alla particolare struttura dei dati. Da ultimo, poiché gli algoritmi di triangolazione e tetraedrazione sono poco adatti alla rappresentazione di elementi con tangente e curvatura continue, è opportuno raffinare l'interpolazione ottenuta con superfici che possiedono tali caratteristiche, quali le funzioni spline di Bézier a supporto triangolare, capaci di generare una superficie finale ben levigata e, nel caso di triangolazione abbastanza equilatera con elementi sufficientemente uniformi, dotata di una continuità C_2. Sia allora $I_p = \{P_1, P_2, ..., P_n\}$ un insieme di *n* punti di una struttura dati in R^3 la cui analisi topologica e riorganizzazione interna permette di individuare la successione ottimale dei punti complanari. In letteratura, esistono diversi criteri per ricavare le relazioni topologiche fra punti di uno stesso piano. Il criterio qui usato è quello di minima distanza: due punti complanari P_i e P_k sono successivi, se e solo se hanno distanza minima rispetto ad ogni altro punto dell'insieme dato: $d(P_i, P_K) \leq d(P_i, P_j) \quad \forall j \neq i \neq k$

La Fig. A.1a mostra dati originali, con una qualsiasi locazione di punti, mentre la Fig. A.1b presenta la topologia ottenuta con il criterio della minima distanza e la nuova numerazione dei punti.
La fase successiva ha lo scopo di riorganizzare la struttura dati, calcolando nell'ordine:

- il centroide C dei punti dello stesso piano (x_c, y_c, z_c),
- una traslazione dell'origine delle coordinate sul centroide,

- la trasformazione delle coordinate di ogni punto da cartesiane a cilindriche, avendo assunto come primo punto quello in cui l'angolo di direzione assume valore minimo: $\theta_1 = \theta_{min}$,
- la rinumerazione della successione partendo dal primo punto individuato.

Fig. A.1 – Topologia con il criterio della minima distanza

La Fig. A.2a mostra la traslazione del sistema di riferimento e la Fig. A.2b la rinumerazione dei punti.

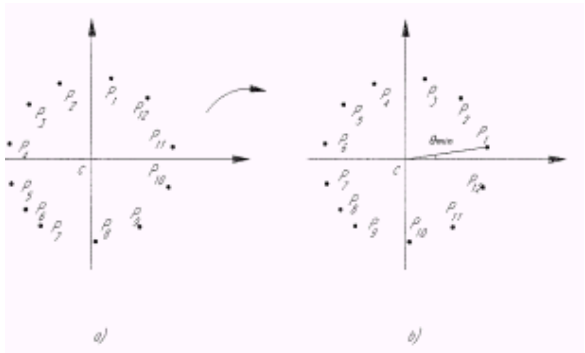


Fig. A.2 – Rinumerazione dei punti

Nella fase di predizione dei dati, si vuole poi ottenere una modellazione continua delle linee di contorno ottenute, cosicché sia possibile eseguire predizioni di nuovi dati (ad esempio, regolarmente cadenzati). La curva di Catmull – Rom bene si presta all'interpolazione richiesta, in quanto risponde positivamente a tre esigenze fondamentali:

- necessità di trovare una curva interpolante con una continuità almeno C_1;
- necessità di trovare una curva che risenta, solo localmente, di una eventuale perturbazione;
- possibilità di cambiare un punto appartenente alla linea, oppure di determinare di un nuovo punto, senza operare sull'intera curva, ma solo nella zona attorno a questo punto.

La funzione spline cubica di Catmull – Rom è una curva che interpola tutti i punti di una qualsiasi linea. Il vettore tangente in ciascun punto è parallelo alla congiungente il punto precedente ed il punto seguente e ha continuità parametrica C_1 nei punti stessi. L'equazione di un segmento della curva di Catmull – Rom, interpolante quattro punti successivi, in forma compatta, è data dalla seguente espressione matriciale: $p_t = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t^T \cdot H \cdot g$, ovvero in forma estesa:

$$p_t = t^T \cdot H \cdot g = \frac{1}{2}\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}\begin{bmatrix} P_{i-1} \\ P_i \\ P_{i+1} \\ P_{i+2} \end{bmatrix} \qquad 0 \le t \le 1$$

dove *t* è un vettore parametrico, H la matrice base (o di trasformazione) e g un vettore geometrico. La scelta del parametro è arbitraria e può dipendere dalla forma finale che si vuole dare, poiché esso influenza tale forma. In ogni caso, se i nodi della linea sono equispaziati, la parametrizzazione a distanze uguali offre comunque buoni risultati.

Per predire un punto, sono presi in considerazione i quattro nodi più vicini, introducendo il valore parametrico *t* del punto predetto nell'equazione di un segmento della curva di Catmull – Rom (la stessa procedura può essere estesa anche al caso di una curva di approssimante, ad esempio, la funzione B – spline uniforme).

La tecnica proposta consiste nello scegliere arbitrariamente un numero l di punti da predire, al fine di ottenere un raffinamento dell'interpolazione ed un miglioramento della successiva triangolazione. Tale numero è moltiplicato per un coefficiente *k* che può assumere valore 1 o 2 rispettivamente per curve di ordine dispari o pari.

La ricerca della esatta corrispondenza fra i punti di curve limitrofe, a partire da una serie di curve di contorno piane, è una parte particolarmente complessa per la ricostruzione della superficie esterna. Infatti una corrispondenza sfalsata comporta una falsa ricostruzione dell'oggetto, con la conseguente formazione di un doppio cono invece di un cilindro (Fig. A.3). Questo argomento attiene ad un insieme particolarmente complesso di problemi, facenti capo alla tematica del matching relazionale, detto esatto, nel caso esista una perfetta corrispondenza biunivoca, ed inesatto altrimenti. In questo ultimo caso, data:

- una matrice di incidenza le cui righe corrispondono ai lati (od archi) di un grafo e le cui colonne ai nodi dello stesso grafo, ovvero
- una matrice di connessione i cui elementi diagonali principali corrispondono ai suddetti nodi ed i cui elementi extra – diagonali ai suddetti lati (od archi),

si vuole trovare, per tentativi mirati, la migliore corrispondenza con un'altra matrice corrispondente ad un grafo simile, ma non uguale.

Sono stati sperimentati diversi modi, per definire i punti corrispondenti, appartenenti a curve di contorno limitrofe.

- Uno di questi è il criterio della minima distanza e consiste nel calcolare la misura dei segmenti congiungenti i punti di due curve limitrofe, individuando e scegliendo i più piccoli.
- Un altro criterio è quello della verticale e consiste nell'individuare i punti di una curva più vicini alle verticali (asse *z*) calate dai punti della curva successiva.
- Un terzo criterio, estensibile anche a figure complesse, è quello della direzione: i punti corrispondenti tra due curve successive sono quelli aventi pressoché la stessa direzione rispetto ad un unico sistema di riferimento (ovvero l'angolo di direzione delle coordinate polari è pressoché uguale).

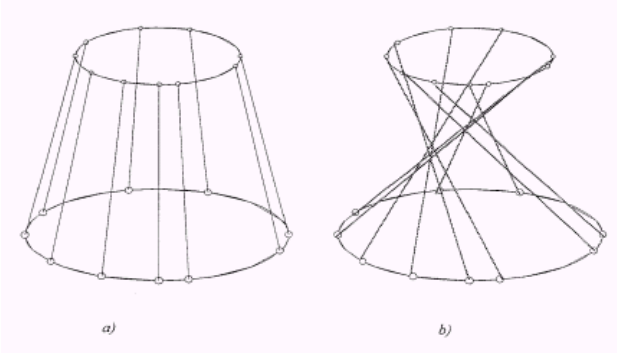


Fig. A.3 – Effetto del doppio cono

Avendo definito, in base al criterio appena introdotto, gli accoppiamenti tra i punti di livelli limitrofi, è possibile effettuare la ricostruzione della superficie esterna:

- su ogni livello pari, si predicono N punti;
- su ogni livello dispari, se ne predicono $2N$, sempre ad intervalli equidistanti in θ.

Di conseguenza, è assai naturale istituire i collegamenti di una rete di triangoli fra curve successive (Fig. A.4).
Si noti inoltre come questo criterio di triangolazione, ad eccezione di casi particolari con una notevole rugosità nelle curve di Catmull – Rom, soddisfi il criterio del cerchio di Delaunay.

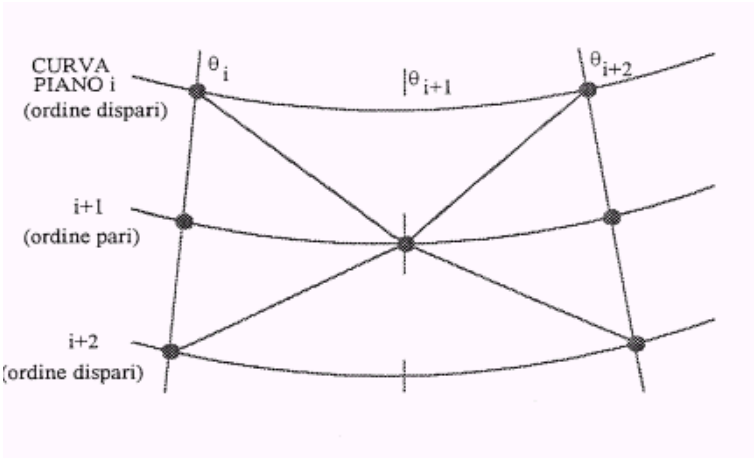


Fig. A.4 – Criterio di triangolazione tra curve successive di diverso ordine:
si collegano i punti che hanno la stessa coordinata angolare ϑ e quelli prossimi ad essi

Infine poiché la superficie costruita con la triangolazione è ovviamente poco adatta alla rappresentazione di elementi con tangente e curvatura continue, come già detto in precedenza, è opportuno raffinare l'interpolazione con superfici che possiedono caratteristiche più appropriate.
Pertanto l'uso di superfici parametriche, affiancate ed integrate da linee parametriche per seguire le linee direttrici e le linee di rottura della superficie, permettono di dare alla superficie (ed alle suddette linee) caratteristiche di continuità e graduale variabilità, perché costruite con particolari elementi finiti, dotati delle proprietà geometriche richieste.

PARTE III – RAPPRESENTAZIONE PARAMETRICA DI CURVE E SUPERFICI

1. Generalità

Dato un insieme di punti nel piano o nello spazio 3D, appartenenti ad una curva o superficie data, si desidera rappresentare una curva o superficie che interpola od approssima (al meglio) tale insieme. A tal fine, si utilizzano funzioni parametriche, poiché queste offrono molti vantaggi computazionali rispetto ad altre funzioni non – parametriche e, prendendo in considerazione polinomi di basso ordine, permettono di costruire curve o superfici con levigatezza desiderata ed oscillazioni indesiderate minime.

L'immagine di un intervallo continuo (aperto, chiuso, semiaperto, finito od infinito), ottenuta tramite una funzione continua, localmente iniettiva nello spazio in R^2 o R^3, è detta curva. Una curva può essere considerata un insieme ordinato di punti, riferita ad un dato punto origine. Spesso delle curve, sono presi in esame solo alcuni punti significativi, pertanto per rappresentare una curva, interpolante od approssimante l'insieme dato, bisogna innanzitutto scegliere una certa equazione. Una scelta vantaggiosa porta ad equazioni parametriche, poiché esse offrono vantaggi quali:

- offrono più gradi di libertà nel controllo della forma, poiché ciascuna componente è espressa da una particolare funzione del parametro;
- sostituiscono le pendenze geometriche (cioè le funzioni derivate che talvolta possono essere infinite) con vettori tangenti in forma parametrica che, non essendo mai infiniti, evitano problemi computazionali;
- trattano tutte le curve in una stessa forma, quando le loro funzioni parametriche sono impostate sotto le medesime condizioni.

Le funzioni parametriche più scelte sono quelle che offrono maggiori facilità nei calcoli, come la derivazione e l'integrazione: una di queste è la funzione polinomiale. Selezionata una serie di parametri t_i, associati all'insieme dei punti dati: P_i ($i = 0, n$), appartenenti alla curva data, esistono diversi modi per costruire una curva di interpolazione con polinomi.

Tuttavia se la curva interpolante attraversa tutti i punti dell'insieme dato ed è unicamente determinata, l'ordine del polinomio corrispondente tende ad essere elevato quando il numero dei punti diventa elevato e questo fatto può dare alla curva oscillazioni indesiderate. Questo è contraddittorio con la aspettative; infatti dato un maggiore numero di punti dell'insieme selezionato, maggiore dovrebbe essere la convergenza della curva trovata alla curva originale.

Anche per tale problema, si cerca di ridurre l'ordine del polinomio e, di conseguenza, le oscillazioni indesiderate. A tal fine, si cerca di costruire piccoli segmenti di curve (attraversanti piccoli gruppi di punti successivi) e poi si cerca di unirli far loro, con la continuità desiderata.

- Una curva composta da due archi ha la continuità geometrica G_0 nel punto di unione.
- Se entrambi gli archi, in questo punto, hanno uguale direzione del vettore tangente (anche se non necessariamente lo stesso modulo), allora la curva possiede anche continuità G_1.
- Se invece hanno in comune tanto la direzione, quanto il modulo, allora la curva ha continuità parametrica C_1 in quel punto.

- Infine se hanno uguale direzione e modulo fino alla n – esima derivata, la curva ha la continuità parametrica C_n.

I polinomi più convenienti sono quelli con grado: $3 \leq m \leq 5$.
Fra queste curve, la curva di Catmull – Rom e la funzione B – spline bene si prestano a risolvere problemi di interpolazione ed approssimazione.

2. Funzioni di miscelamento

La forma algebrica di un segmento di curva parametrica cubica in R^3 è data dalle equazioni:

$$x(t) = a_{3x}t^3 + a_{2x}t^2 + a_{1x}t + a_{0x}$$
$$y(t) = a_{3y}t^3 + a_{2y}t^2 + a_{1y}t + a_{0y}$$
$$z(t) = a_{3z}t^3 + a_{2z}t^2 + a_{1z}t + a_{0z}$$

Per rendere limitato il segmento sopraccitato, spesso si sceglie un intervallo limitato del parametro *t*, ovvero $0 \leq t \leq 1$. Se si definisce *t* il vettore parametrico ed A la matrice dei coefficienti:

$$t^T = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} a_{3x} & a_{3y} & a_{3z} \\ a_{2x} & a_{2y} & a_{2z} \\ a_{1x} & a_{1y} & a_{1z} \\ a_{0x} & a_{0y} & a_{0z} \end{bmatrix}$$

si ottiene una forma più compatta dell'equazione: $p(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t^T A$, dove $p(t)$ è il vettore delle posizioni di un punto qualsiasi appartenente alla curva.
La derivata di tale vettore è il vettore parametrico tangente alla curva nel punto stesso; esso si ottiene derivando la funzione $p(t)$ rispetto al parametro *t*, dove l'apice indica la derivazione rispetto al parametro *t*.

$$\frac{d}{dt}p(t) = p'(t) = \begin{bmatrix} \frac{d}{dt}x(t) & \frac{d}{dt}y(t) & \frac{d}{dt}z(t) \end{bmatrix} = \frac{d}{dt}T \cdot A = \begin{bmatrix} 3t^2 & 2t & 1 & 0 \end{bmatrix} \cdot C$$
$$= \begin{bmatrix} 3a_{3x}t^2 + 2a_{2x}t + a_{1x} & 3a_{3y}t^2 + 2a_{2y}t + a_{1y} & 3a_{3z}t^2 + 2a_{2z}t + a_{1z} \end{bmatrix}$$

I coefficienti della matrice A si ottengono imponendo all'equazione di soddisfare le condizioni geometriche, ovvero i valori delle coordinate e/o delle tangenti agli estremi, oltreché le condizioni dipendenti dalla curvatura, dalla torsione e dalle derivate di ordine più elevato.

Tali condizioni sono selezionate a seconda del modello desiderato e, trattandosi di sola geometria, conviene rappresentare la curva in forma geometrica anziché in quella algebrica. La forma geometrica è data dalle seguenti equazioni:

$$p(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t^T \cdot M \cdot g \qquad b = M \cdot g$$

dove b è il vettore dei coefficienti geometrici, g il vettore dei vincoli geometrici, detto anche vettore geometrico, e M la matrice base o di trasformazione:

$$M \cdot g = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}$$

Gli elementi della matrice base M sono costanti; pertanto il loro prodotto con gli elementi del vettore parametrico *t* dà ancora un polinomio parametrico cubico, chiamato vettore delle funzioni di miscelamento *f* (in inglese: blending functions).
Le funzioni di miscelamento sono utili per ottenere le coordinate e le due derivate (prima e seconda) in un punto qualsiasi appartenente al segmento di curva $0 \leq t \leq 1$. Esse mescolano i vari contributi portati dalle condizioni geometriche e sono identiche per tutte e tre le coordinate. Riscrivendo la formula della forma geometrica si ha:

$$p(t) = f^T \cdot g = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}$$

dove f_1, f_2, f_3, f_4 sono le funzioni di miscelamento. Differenziando queste funzioni rispetto al parametro *t*, si ottiene la derivata prima di un punto qualsiasi appartenente al segmento della curva:

$$p'(t) = f'^T g = \begin{bmatrix} f_1' & f_2' & f_3' & f_4' \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}$$

$$p(t)'' = f'' \cdot g = \begin{bmatrix} f_1'' & f''_2 & f_3'' & f_4'' \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}$$

ed analogamente si ha, derivando un'altra volta, per ottenere la derivata seconda.

3. Segmenti di curva polinomiale di Hermite

Il segmento di curva parametrica cubica di Hermite (indicato con il pedice H) è una curva interpolante due punti successivi, conformemente alle condizioni imposte dalle loro coordinate e dalle loro derivate prime. Il vettore di posizione di un punto qualsiasi, appartenente a tale segmento, è dato dalla seguente equazione:

$$p(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t^T \cdot M_H \cdot g_H = f_H^T \cdot g_H$$

dove f_H è il vettore delle funzioni di miscelamento (Fig. 3.1) e g_H quello geometrico, formato da quattro componenti (le coordinate dei due punti estremi P_1 e P_2, e le loro derivate prime P'_1 e P'_2): $g_H^T = \begin{bmatrix} p_1 & p_2 & p_1' & p_2' \end{bmatrix}$.

Per costruire il segmento di curva, in modo unico, bisogna che i coefficienti delle funzioni di miscelamento siano determinati dalle condizioni al contorno (le coordinate e le derivate prime di due punti successivi). Si analizza ora la sola componente x, poiché le funzioni di miscelamento non cambiano per le altre componenti:

$$x(t) = t^T \cdot M_H \cdot p_X = f^T \cdot p_X = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_1' \\ p_2' \end{bmatrix}_x$$

Prendendo in considerazione le condizioni al contorno (per $t = 0$ e $t = 1$), si hanno i seguenti valori della funzione e della sua derivata:

$$x(0) = p_{1x} = t^T \cdot M_H \cdot p_x = f^T \cdot p_x = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \cdot M_H \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_1' \\ p_2' \end{bmatrix}_x$$

$$x(1) = p_{2x} = t^T \cdot M_H \cdot p_x = f^T \cdot p_x = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \cdot M_H \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_1' \\ p_2' \end{bmatrix}_x$$

$$x^t(0) = p'_{1x} = t^{T^t} \cdot M_H \cdot p_x = f^{T^t} \cdot p_x = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \cdot M_H \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_1' \\ p_2' \end{bmatrix}_x$$

$$x^t(1) = p'_{2x} = t^{T^t} \cdot M_H \cdot p_x = f^{T^t} \cdot p_x = \begin{bmatrix} 3 & 2 & 1 & 0 \end{bmatrix} \cdot M_H \cdot \begin{bmatrix} p_1 \\ p_2 \\ p'_1 \\ p'_2 \end{bmatrix}_x$$

da queste quattro relazioni, si ottiene la matrice di base M_H imponendo l'identità:

$$\begin{bmatrix} p_1 \\ p_2 \\ p'_1 \\ p'_2 \end{bmatrix}_x = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix} \cdot M_H \cdot \begin{bmatrix} p_1 \\ p_2 \\ p'_1 \\ p'_2 \end{bmatrix}_x$$

$$M_H = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

essa è unica e, come già detto in precedenza, vale anche per le altre componenti *y* e *z*. Per trovare poi le derivate prima e seconda in un punto qualsiasi, appartenente al segmento di curva, si differenziano le funzioni di miscelamento:

$$f^T = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 \end{bmatrix}_H = t^T \cdot M_H = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

rispetto al parametro *t* e riordinando si ha:

$$f^{T'} = \begin{bmatrix} f'_1 & f'_2 & f'_3 & f'_4 \end{bmatrix}_H = t^T \cdot M_H^{(1)} = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 6 & -6 & 3 & 3 \\ -6 & 6 & -4 & -2 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

ed analogamente per la derivata seconda:

$$f^{T''} = \begin{bmatrix} f''_1 & f''_2 & f''_3 & f''_4 \end{bmatrix}_H = t^T \cdot M_H^{(2)} = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 12 & 12 & 6 & 6 \\ -6 & 6 & -4 & -2 \end{bmatrix}$$

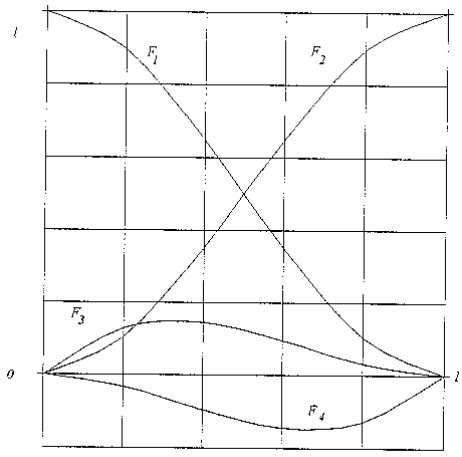


Fig. 3.1 – Funzioni di miscelamento di Hermite

Tramite le curve di Hermite si possono descrivere infinite curve, variando i moduli dei vettori tangenti k_1 e k_2 nei due punti estremi:

$p'(0) = k_1 r_1$ e $p'(1) = k_2 r_2$ dove: $r_1 = \frac{p'(0)}{|p'(0)|}$ e $r_2 = \frac{p'(1)}{|p'(1)|}$

Inoltre questa possibilità permette di controllare la curvatura in ogni estremo, oppure di fissare la posizione di alcuni punti intermedi. Per due segmenti di curve di Hermite contigue, aventi in comune un punto estremo si dice che la curva composta ha continuità G_1 (ed oltre alla continuità geometrica, si ha anche continuità parametrica C_1, se $k = 1$), se i vettori tangenti ai segmenti di curve hanno la stessa direzione, ovvero quando i vettori geometrici hanno la seguente forma:

$\begin{bmatrix} P_i & P_{i-1} & P_i' & P_{i-1}' \end{bmatrix}$ e $\begin{bmatrix} P_i & P_{i-1} & kP_i' & P_{i-1}' \end{bmatrix}$ con $k \geq 0$

4. Segmenti di curva polinomiale di Bèzier.

Il segmento di curva parametrica cubica di Bèzier (indicato con il pedice B), definito da quattro punti, detti punti di controllo, ha le seguenti caratteristiche:

- esso interpola il primo e l'ultimo punto, mentre approssima i due punti interni;
- il primo e l'ultimo lato del poligono di Bèzier sono tangenti al segmento di curva nei punti estremi;
- è interamente contenuto nell'involucro convesso dei punti di controllo, detto poligono di Bèzier.

Il vettore posizione di un punto qualsiasi, appartenente a tale segmento, è dato dalla seguente equazione:

$$p(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t^T \cdot M_B \cdot g_B = f_B^T \cdot g_B$$

dove f_B è il vettore delle funzioni di miscelamento (Fig. 4.1) e g_B il vettore geometrico, formato da quattro componenti (i quattro punti di controllo P_0, P_1, P_2, P_3): $g_B^T = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 \end{bmatrix}$. Le funzioni di miscelamento sono le funzioni polinomiali di Bernstein:

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

e, per il caso cubico ($n = 3$), hanno espressione:

$$B_{i,3}(t) = \frac{3!}{i!(3-i)!} t^i (1-t)^{3-i} \qquad i = 0, 1, 2, 3$$

Sviluppando tali funzioni, si ottengono le seguenti quattro funzioni di miscelamento:

$$f_{B0} = B_{0,3} = (1-t)^3$$
$$f_{B1} = B_{1,3} = 3t(1-t)^2$$
$$f_{B2} = B_{2,3} = 3t^2(1-t)$$
$$f_{B3} = B_{3,3} = t^3$$

Riordinando si ha:

$$f_B = t^T \cdot M_B = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

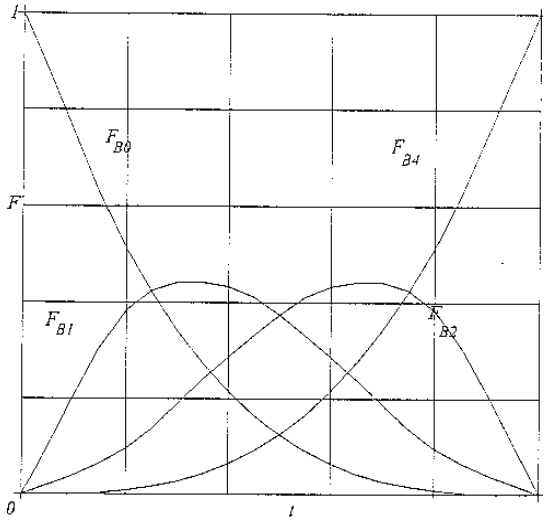


Fig. 4.1 – Funzioni di miscelamento di Bézier

Le derivate prima e seconda del vettore delle funzioni di miscelamento F_B sono rispettivamente:

$$f_B{}' = t^T \cdot M_B^{(1)} = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ -3 & 9 & -9 & 3 \\ 6 & -12 & 6 & 0 \\ -3 & 3 & 0 & 0 \end{bmatrix}$$

$$f_B{}'' = t^T \cdot M_B^{(2)} = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -6 & 18 & -18 & 6 \\ 1 & -12 & 6 & 0 \end{bmatrix}.$$

Nota la derivata prima del vettore delle funzioni di miscelamento, è immediata la determinazione del vettore tangente ad un punto qualsiasi appartenente alla curva:

$$p'(t) = f^{T'} \cdot g_B$$

e per i punti estremi ($t = 0$ e $t = 1$), si ha: $p'(0) = 3(P_1 - P_0)$ e $p'(1) = 3(P_3 - P_2)$. Questo dimostra come il segmento di curva abbia tangenti comuni nei punti estremi al primo ed all'ultimo lato del poligono di Bèzier.

La curva composta, ottenuta tramite unione di due segmenti di curve successive a partire da un insieme di punti P_1, P_2,...,P_6, P_7, ha continuità geometrica G_1, se nel punto di unione si ha: $P_3 - P_4 = k(P_4 - P_5)$, con $k > 0$, cioè quando i tre punti P_3, P_4, P_5 sono distinti ed allineati.

Infine si noti come l'estensione ai casi $n > 3$ sia possibile, ma dia luogo a relazioni di maggiore complessità, spesso senza notevoli vantaggi.

5. Parametrizzazione delle curve

Dato un insieme di punti, appartenenti ad una curva, la costruzione di una curva (interpolante od approssimante), in forma parametrica, richiede la conoscenza di un opportuno parametro. Esistono diversi modi per scegliere un parametro e dipendono dalla forma della curva desiderata, dalle operazioni di calcolo e dalla precisione che si vuole ottenere:

- parametrizzazione ad intervalli uguali: l'intero intervallo $t[a,b]$, dove solitamente si assume: $a = 0$, è diviso in n intervalli: $t_i = ib/n$ con $i = 0, n$.
- parametrizzazione di un arco: si calcolano le lunghezze degli archi fra punti successivi Δ_i e la lunghezza totale: $s = \sum_{i=0}^{n} \Delta_i$ è divisa in n intervalli: $t_i = \sum_{j=0}^{i} b\Delta_j / s$ con $i = 0, n$.

- parametrizzazione di una corda: è una scelta alternativa a quella precedente; infatti spesso non è facile determinare la lunghezza di un arco fra due punti successivi e pertanto si sceglie, in alternativa, la lunghezza della corda: $\Delta_i = \left|P_i - P_{i+1}\right|$.
- parametrizzazione centripeta: $\Delta_i = \left|P_i - P_{i+1}\right|^k$, dove solamente si assume: k = ½.
- parametrizzazione invariante affine, dove la distanza fra due punti Δ_i è calcolata secondo la norma definita dalla matrice Q (matrice inversa della matrice di covarianza delle coordinate dell'insieme dato di punti): $\Delta_i = E_2[P] = (P_i - P_{i+1})\,Q\,(P_i - P_{i+1})^T$.

Inoltre esistono criteri che tengono conto, non solo delle distanze, ma anche degli angoli fra punti successivi.

6. Funzioni spline cubiche

Per costruire una curva di interpolazione di un insieme dato di $n+1$ punti: P_i ($i = 0,...,n$), ed una serie scelta di intervalli dei parametri $t_i \in [a,b]$, soddisfacente la condizione: $a = t_0 < t_1 < ... < t_{n-1} < t_n = b$, si può utilizzare la funzione spline (smooth path line) cubica, cioè una curva composta da una serie di segmenti di curve successive che sono raccordate tra loro, in modo gradualmente variato, ovvero hanno continuità C_2 in tutti i nodi. Allora se si considera il segmento fra due punti successivi P_i, P_{i+1}:

$$p_i(t) = a_{3i}(t - t_i)^3 + a_{2i}(t - t_i)^2 + a_{1i}(t - t_i) + a_{0i} \qquad t \in [t_i, t_{i+1}] \qquad (i = 0,...,n)$$

e si impone la condizione di continuità fra i due segmenti di curve successivi:

$$p_i(t_i) = p_{i-1}(t_i) \qquad p_i'(t_i) = p_{i-1}'(t_i) \qquad p_i''(t_i) = p_{i-1}''(t_i)$$

si ottiene l'equazione:

$$p_i(t) = \left(2\frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - 3\frac{(t-t_i)^2}{(t_{i+1}-t_i)^2} + 1\right)P_i + \left(-2\frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} + 3\frac{(t-t_i)^2}{(t_{i+1}-t_i)^2} + 1\right)P_{i+1} +$$
$$+ \left(\frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - 2\frac{(t-t_i)^2}{(t_{i+1}-t_i)^2} + (t-t_i)\right)P_i^t + \left(\frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - \frac{(t-t_i)^2}{(t_{i+1}-t_i)^2}\right)P_{i+1}^t$$

Nel caso di parametrizzazione ad intervalli uguali $(t - t_i) = 1$, si ottiene la curva di interpolazione di Hermite. Inoltre poiché spesso non è nota la derivata prima dei nodi, occorre stimarla o calcolarla; di conseguenza, imponendo le condizioni di continuità C_2 al nodo t_i, si ottiene la formula ricorsiva:

$$(t_{i+1} - t_i)P_{i-1}^t + 2(t_i - t_{i-1} + t_{i+1} - t_i)P_i^t + (t_i - t_{i-1})P_{i-1}^t = 3\frac{t_i - t_{i-1}}{t_{i+1} - t_i}P_{i+1} - P_i + 3\frac{t_{i+1} - t_i}{t_i - t_{i-1}}P_i - P_{i-1}$$

Per avere l'unicità della curva spline, occorre introdurre due ulteriori condizioni, poiché le equazioni sono $n - 1$ e le incognite (cioè le derivate prime) $n + 1$. Queste condizioni (solitamente condizioni al contorno) possono essere scelte per soddisfare alcune esigenze; ad esempio, si ha:

- spline naturale: $p''(a) = 0$ e $p'(b) = 0$
- spline ciclica (utile per generare curve chiuse $P_0 = P_n$): $p'(a) = p'(b)$ e $p''(a) = p''(b)$
- spline vincolata (campled spline): $p'(a) = P_0'$ e $p'(b) = P_n'$, dove P_0' e P_n' sono valori assegnati dei vettori tangenti.

La funzione B – spline cubica uniforme (con parametrizzazione ad intervalli uguali) nasce dall'unione di segmenti di curve che hanno una continuità C_2 nel loro punto di unione. I coefficienti del polinomio di tale curva dipendono da pochi punti di controllo, al contrario delle precedenti curve dove i coefficienti dipendono da tutti gli $n + 1$ punti. Questa caratteristica, detta anche caratteristica locale, comporta due vantaggi:

- uno legato alla variazione della curva, a seguito del cambiamento di un punto, che dà luogo solo ad una variazione locale (ovvero variando la posizione di un punto, varieranno solo quattro segmenti di curve attorno a quello stesso punto);
- l'altro al ridotto tempo di calcolo dei coefficienti.

Inoltre la curva di approssimazione di un insieme di $n + 1$ punti, consiste in $n - 2$ segmenti di curve, definite da quattro punti successivi dell'insieme dato ed unite con continuità C_2. Nella determinazione del segmento di curva, non è richiesta la conoscenza del vettore tangente. Il primo segmento di curva $p_3(t)$ è formato dai primi quattro punti: P_0, P_1, P_2, P_3, e dall'intervallo parametrico compreso tra $t_3 = 0$ e $t_4 = 1$, mentre l'ultimo è formato dai punti: P_{n-3}, P_{n-2}, P_{n-1}, P_n, e dall'intervallo parametrico compreso fra $t_n = n - 3$ e $t_{n+1} = n - 2$. In generale, il segmento di curva $p_i(t)$, dove: $t_i \leq t \leq t_{i+1}$, è dato dall'equazione:

$$p_i(t) = t_i \cdot M_{Bs} \cdot g_{Bsi}$$

dove t_i è il vettore parametrico, g_{BSi} il vettore geometrico: $g_{Bsi}^T = \begin{bmatrix} P_{i-3} & P_{i-2} & P_{i-1} & P_i \end{bmatrix}$ e M_{Bs} la matrice base:

$$M_{Bs} = \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}$$

Sviluppando l'equazione e sostituendo l'intervallo $[t_i, t_{i+1}]$ con $[0,1]$, si ottiene:

$$p_i(t-t_i) = t_i \cdot M_{Bs} \cdot g_{Bsi} = t \cdot M_{Bs} \cdot g_{Bsi} = F_{Bs} \cdot g_{Bsi} = F_{Bs-3} \cdot P_{i-3} + F_{Bs-2} \cdot P_{i-2} + F_{Bs-1} \cdot P_{i-1} + F_{Bs0} \cdot P_i =$$
$$= \frac{(1-t)^3}{6} P_{i-3} + \frac{3t^3 - 6t^2 + 4}{6} P_{i-2} + \frac{-3t^3 + 3t^2 + 3t + 1}{6} P_{i-1} + \frac{t^3}{6} P_i$$

Essendo le quattro funzioni di miscelamento (Fig. 6.1) non negative e la loro somma unitaria in tutto l'intervallo $[0,1]$, ogni segmento giace all'interno dell'involucro convesso dei quattro punti base.

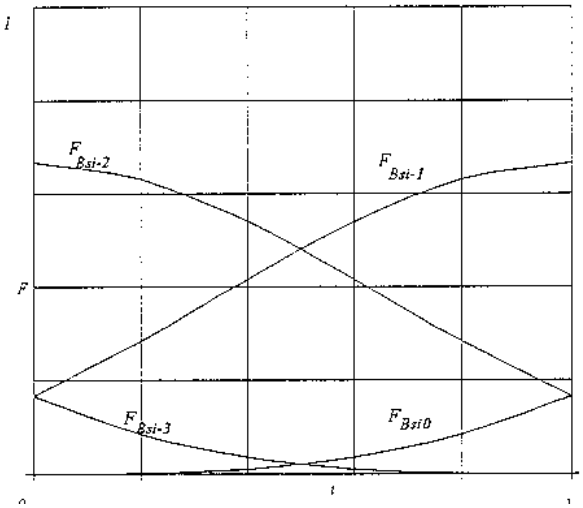


Fig. 6.1 – Funzioni di miscelamento della funzione B – spline cubica uniforme

La <u>funzione B – spline cubica non – uniforme</u> è una generalizzazione della precedente. In questo caso, l'intervallo del parametro, fra i nodi successivi, può essere non costante, facendo sì che le funzioni di miscelamento (definite ricorsivamente più oltre) fra le curve successive possano variare:

$$f_{Bi,1}(t) = \begin{cases} 1 & t_i \le t \le t_{i-1} \\ 0 & altrimenti \end{cases}$$

$$f_{Bi,2}(t) = \frac{t - t_i}{t_{i+1} - t_i} f_{Bi,1}(t) + \frac{t_{i+2} - t}{t_{i+2} - t_{i+1}} f_{Bi+1,1}(t),$$

$$f_{Bi,3}(t) = \frac{t - t_i}{t_{i+2} - t_i} f_{Bi,2}(t) + \frac{t_{i+3} - t}{t_{i+3} - t_{i+1}} f_{Bi+1,2}(t)$$

$$f_{Bi,4}(t) = \frac{t - t_i}{t_{i+3} - t_i} f_{Bi,3}(t) + \frac{t_{i+4} - t}{t_{i+4} - t_{i+1}} f_{Bi+1,3}(t)$$

Le funzioni di miscelamento sono sempre non negative e la loro somma unitaria; pertanto il segmento di curva di B – spline cubica non – uniforme giace ancora all'interno dell'involucro convesso dei suoi quattro punti. La curva approssima un insieme di $n + 1$ punti.

La sequenza del valore nodale non deve essere decrescente, in tutto l'intervallo $[a,b]$ e, se necessario, possono essere introdotti valori nodali multipli. In presenza di nodi multipli, il denominatore della funzione di

miscelamento può diventare nullo. In tal caso, convenzionalmente si considera zero tale funzione. Il segmento di curva B – spline non uniforme è definita dai quattro punti di controllo: P_{i-3}, P_{i-2}, P_{i-1}, P_i , e dai loro pesi: $f_{Bi-3,4}$, $f_{Bi-2,4}$, $f_{Bi-1,4}$, $f_{Bi,4}$, espressi nella funzione di miscelamento:

$$p_i(t) = P_{i-3} \cdot f_{Bi-3,4}(t) + P_{i-2} \cdot f_{Bi-2,4}(t) + P_i \cdot f_{Bi-1,4}(t) + P_i \cdot f_{Bi,4}(t)$$

L'intera curva è definita all'interno dell'intervallo compreso fra t_3 e t_{n+1}.

7. Curve di Catmull – Rom

La curva di Catmull – Rom è composta da un insieme di piccoli segmenti di curve (che attraversano un insieme di $n \geq 4$ punti successivi), uniti tra loro con continuità C_1. L'equazione, in forma compatta, del segmento di curva è data dalla seguente espressione:

$$Q(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix} = t \cdot M \cdot g$$

dove t è il vettore parametrico, g il vettore dei vincoli geometrici e M la matrice base o di trasformazione. La matrice base è determinata imponendo le condizioni di contorno e le loro derivate prime, definite considerando la parallela alla congiungente dei due punti precedente e seguente il punto dato:

$$M = \frac{1}{2}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}$$

ed il prodotto del vettore parametrico t per la matrice base M, ovvero il vettore delle funzioni di miscelamento f, risulta:

$$f = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}$$

La parametrizzazione di una curva di interpolazione od approssimazione offre notevoli possibilità di controllo dei punti, oltreché vantaggi computazionali. Spesso conviene contenere il grado del polinomio, poiché un grado maggiore produce oscillazioni indesiderate. Le curve di Hermite sono applicabili, quando sono note, oltre all'insieme dei punti, anche le derivate delle funzioni in quegli stessi punti; al contrario, le curve di Bèzier non richiedono la conoscenza delle derivate. Le funzioni spline dipendono dall'intero insieme di punti; invece le curve B – spline e di Catmull – Rom dipendono solo dai punti circostanti (in particolare, quattro nel caso cubico) e pertanto muovere un punto ha solo un effetto locale sulla curva cercata.

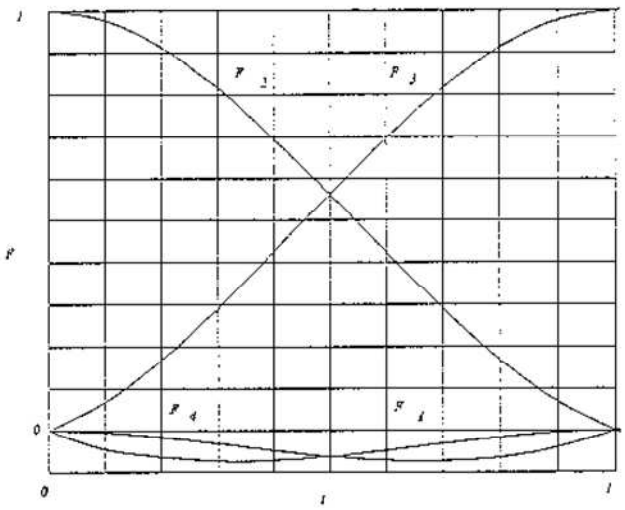


Fig. 7.1 – F unzioni di miscelamento di Catmull – Rom

8. Generalizzazione del problema

La modellazione degli oggetti può essere ottenuta tramite elementi (patch) di dimensione limitata, geometricamente semplici e facilmente descrivibili con funzioni matematiche elementari. Ogni elemento è formato da un insieme di punti, racchiusi all'interno del poligono descritto dall'intersezione di curve le cui coordinate sono date da funzioni parametriche continue a due variabili (*t* e *s*), definite nell'intervallo $[0,1]$ e delimitato da due archi di curva nello spazio geometrico. Un'ottima modalità di rappresentazione di superfici, con le proprietà della continuità e della graduale variazione (almeno sotto certe condizioni), è offerta dalle funzioni spline triangolari che, a partire da una triangolazione di Delaunay (od anche altra), si basano sulla costruzione dei punti di Bézier e, da questi ultimi, di superfici interpolanti od approssimanti.

In ogni caso, la scelta del tipo di elementi (triangolare, quadrangolare, ecc.) e la forma dei suoi lati dipendono dal metodo di interpolazione scelto; ad esempio, se si parte da una triangolazione di Delaunay, ogni elemento è uno dei suoi triangoli. Invece se si usa una interpolazione a maglia regolare (con funzioni polinomiali), i lati saranno curvilinei. Per descrivere tale superficie, occorre estendere al caso bidimensionale l'equazione parametrica delle curve.

In questo caso, se si impone che il vettore geometrico sia non costante, come nel caso delle curve, ma variabile in funzione del parametro, si ottiene la superficie parametrica dell'elemento, dove: $G_i(t) = G_i^T \cdot M^T \cdot T^T$, già un vettore geometrico diventa una matrice e, se i suoi elementi $G_i(t)$ sono polinomi cubici, allora si ottiene una superficie parametrica bicubica:

$$p(s,t) = S \cdot M \cdot \begin{bmatrix} G_1(t) \\ G_2(t) \\ G_3(t) \\ G_4(t) \end{bmatrix} \qquad 0 \le s,t \le 1$$

Riscrivendo l'equazione, in modo da evidenziare la dipendenza da *s* e *t*, si può nuovamente isolare un vettore geometrico G costante:

$$p(s,t) = S \cdot M \cdot G \cdot M^T \cdot T^T$$

La superficie bicubica di Bézier, a forma regolare, è una superficie costituita da una serie di elementi rettangolari, dove la matrice geometrica G è formata da 16 punti di controllo (detti punti di Bézier):

$$G = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

Essi definiscono il poliedro caratteristico:

- i quattro punti p_{11}, p_{41}, p_{14}, p_{44} sono gli unici quattro punti giacenti sull'elemento;
- i punti p_{21}, p_{31}, p_{12}, p_{13}, p_{42}, p_{43}, p_{24}, p_{34} controllano la pendenza delle curve di contorno;
- i rimanenti quattro punti p_{22}, p_{32}, p_{23}, p_{33} controllano le torsioni lungo le curve di contorno.

Tuttavia quando i dati non sono acquisiti su una maglia regolare rettangolare, bensì sono composti da un insieme di punti sparsi, la scelta delle superfici con elementi quadrangolari non è conveniente, poiché per la costruzione della topologia si adoperano usualmente metodi di triangolazione. In questo caso, una superficie con elementi triangolari è la scelta più consona e, per la sua costruzione, occorre procedere all'analisi della superficie triangolare di Bézier (con polinomi di sesto grado), delle condizioni di continuità tra i elementi, del calcolo delle coordinate dei punti e dei vettori tangenti di Loop, nonché dei punti di Bézier.

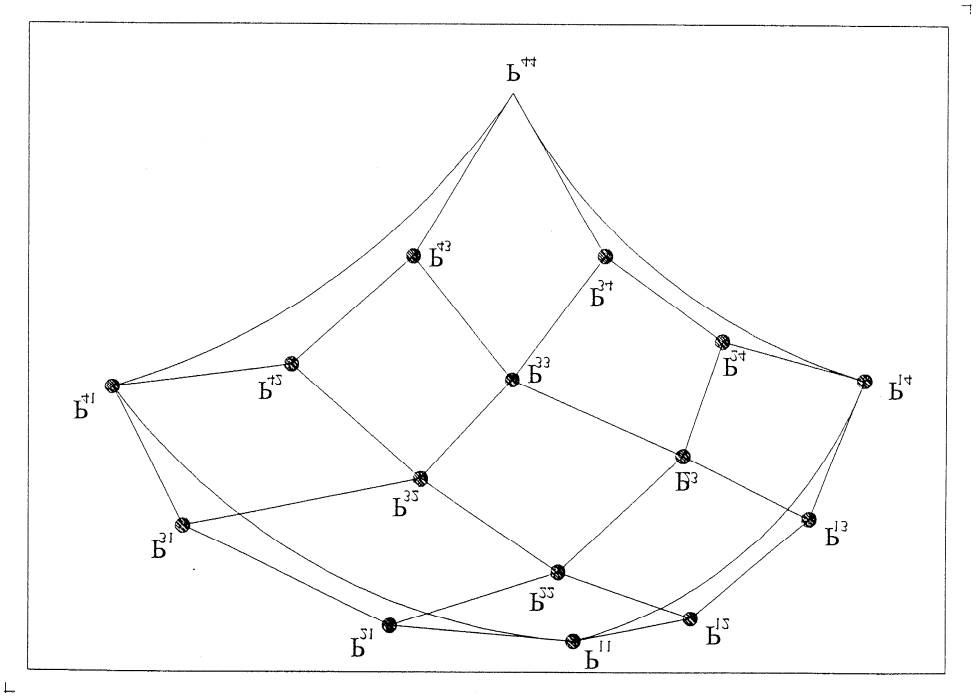


Fig. 8.1 – Punti di controllo e superficie bicubica di Bézier

9. Superficie spline triangolare di Bézier

L'equazione parametrica della spline triangolare di Bézier (di grado $m = 6$) può essere espressa nella forma:

$$p(u) = \sum_{|I|=n} p_I B_I^m(u)$$

dove p_I sono i punti di Bézier e $B_I^m(u)$ i polinomi di Bernstein, definiti come:

$$B_{|I|}^6(u) = \frac{6!}{i!\ j!\ k!} r^i s^j t^k \qquad i+j+k=6 \quad i,j,k \geq 0 \qquad r+s+t=1 \quad r,s,t \geq 0$$

dove: $|I| = i+j+k$ ed $u = r+s+t$.

I punti p_i non sono tutti noti e si determinano a partire dai tre vertici di ogni elemento e dalle condizioni di continuità scelte tra elementi adiacenti. La determinazione di tali punti è fondamentale nella costruzione della superficie, poiché essi definiscono il poliedro caratteristico.
Un metodo efficace, per determinare i punti di Bézier, è stato proposto da Loop e considera facce triangolari, aventi vertici nei 28 punti di Bézier. Essi sono calcolati, a partire da ciascun lato, tenendo conto della continuità G_1 fra facce adiacenti, nonché determinando quattro punti (su ciascun lato) e, per ogni segmento da essi formato, la direzione del piano e della faccia triangolare cui appartiene.

10. Condizioni di continuità

Due elementi limitrofi hanno continuità G_0, lungo un lato (Fig.10.1), se hanno in comune tutti i punti del lato stesso: $p_{i-1}(0,t_i) = p_i(t_i,0)$. Inoltre se esistono tre funzioni: $\phi_i(t_i)$, $\mu_i(t_i)$, $\nu_i(t_i)$, tali che:

$$\phi_i(t_i)\frac{\partial p_i}{\partial t_i}(t_i,0) = \mu_i(t_i)\frac{\partial p_{i-1}}{\partial t_{i-1}}(0,t_i) + \nu_i(t_i)\frac{\partial p_i}{\partial t_{i+1}}(t_i,0)$$

allora gli elementi p_i e p_{i-1} hanno in comune il piano tangente lungo tutto il lato comune. Infine se le stesse funzioni sono costanti, allora i due elementi hanno continuità parametrica C_1. Tale condizione impone che i punti di controllo di Bézier, lungo il lato comune, ed i punti più vicini siano complanari (Fig. 10.2).
Per garantire la continuità dell'intera superficie, oltre alla condizione di continuità tra due elementi adiacenti, occorre garantire anche la condizione di continuità di tutti gli elementi che si congiungono in un vertice:

$$\frac{\partial \phi_i}{dt_i}(0)\frac{\partial p_i}{\partial t_i}(0,0) + \phi_i(0)\frac{\partial^2 p_i}{\partial t_i \partial t_i}(0,0) =$$
$$= \frac{\partial \mu_i}{dt_i}(0)\frac{\partial p_{i-1}}{\partial t_{i-1}}(0,0) + \mu_i(0)\frac{\partial^2 p_{i-1}}{\partial t_i \partial t_{i-1}}(0,0) + \frac{\partial \nu_i}{dt_i}(0)\frac{\partial p_i}{\partial t_{i+1}}(0,0) + \nu_i(0)\frac{\partial^2 p_i}{\partial t_i \partial t_{i+1}}(0,0)$$

Infine risolte le condizioni di continuità fra elementi, è ancora necessario imporre le condizioni per impedire eventuali sovrapposizioni fra elementi e definire il vettore normale, in ogni punto del lato in comune, ovvero:

$$\mu_i(t_i)\nu_i(t_i) > 0 \qquad \text{e} \qquad \frac{\partial p_i}{\partial t_i}(t_i,0) \times \frac{\partial p_i}{\partial t_{i+1}}(t_i,0) \neq 0$$

L'individuazione dei punti di Bézier, con il metodo proposto da Loop, consiste nel determinare quattro punti su ogni lato degli elementi triangolari ed i valori delle pendenze delle facce triangolari (Fig. 10.3).

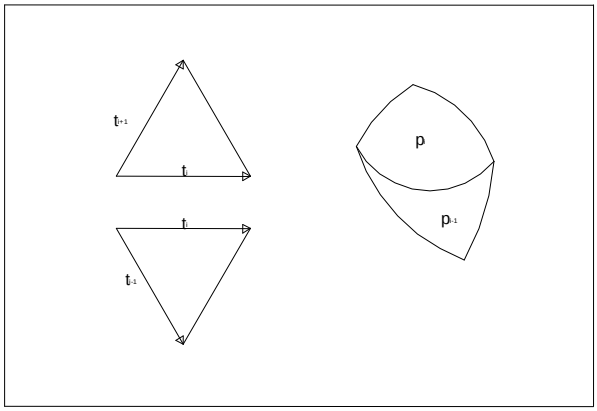


Fig. 10.1 – Unione di due superfici con elementi triangolari

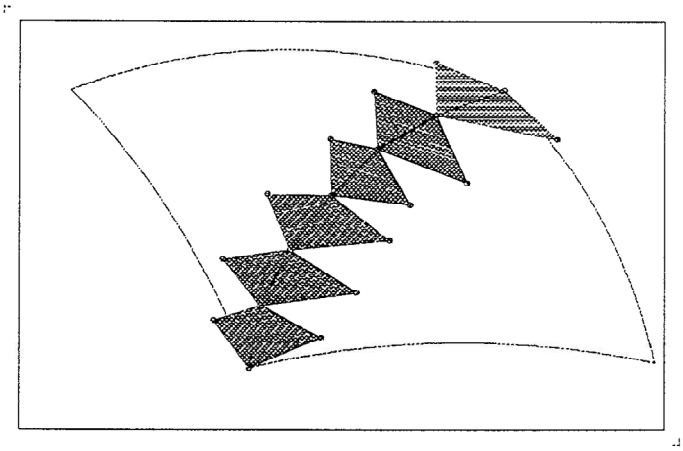


Fig. 10.2 – Condizione di continuità di due elementi triangolari

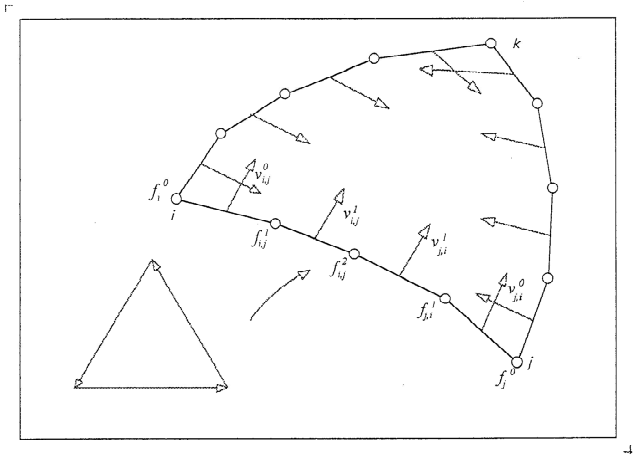


Fig. 10.3 – I punti dell'elemento ed i vettori tangenti per la determinazione dei punti di Bézier

Pertanto al fine di procedere al calcolo delle coordinate dei punti e dei vettori tangenti di Loop, è necessario semplificare le condizioni di continuità G_1, assumendo

$$\mu_i(t_i) = \nu_i(t_i) = \frac{1}{2} \qquad \phi_i(0) = \phi_{i-1}(0) = \phi^0 \qquad \frac{\partial \phi_i}{dt_i}(0) = \frac{\partial \phi_{i-1}}{\partial t_{i-1}}(0) = \phi^1$$

ed introducendo tali valori nella condizione di non sovrapposizione, in modo da ottenere:

$$\phi_i(t)\frac{\partial p_i}{\partial t_i}(t_i,0) = \frac{1}{2}\frac{\partial p_{i-1}}{\partial t_{i-1}}(0,t_i) + \frac{1}{2}\frac{\partial p_i}{\partial t_{i-1}}(t_i,0)$$

dove $\phi_i(t_i)$ è data in funzione dei polinomi quadratici di Bernstein:

$$\phi_i(t_i) = \left(\cos\frac{2\pi}{n}\right)B_0^2(t_i) + \frac{1}{2}B_1^2(t_i) + \left(1-\cos\frac{2\pi}{n_i}\right)B_2^2(t_i)$$

11. Costruzione dell'interpolazione

Per la determinazione dei punti di controllo f^0, f^1, f^2, f^3, si considera un punto q, vertice delle triangolazione, ed i punti attorno ad esso, appartenenti ai triangoli con vertice in q (Fig. 11.1).

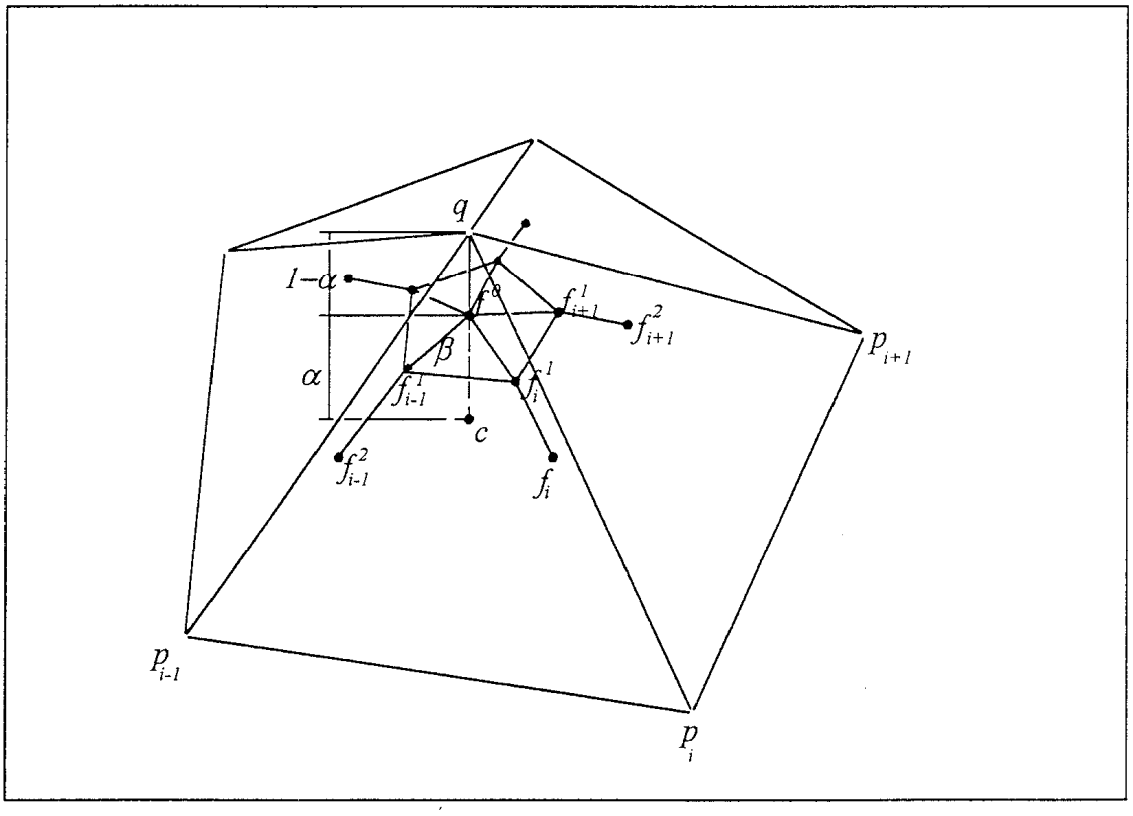


Fig. 11.1 – Costruzione dei punti di controllo f^0, f^1, f^2

- Costruzione dei punti f^0: i punti f^0 si ottengono considerando i vertici q ed i centroidi degli n punti circostanti: $f^0 = \alpha q + F^0 p$, ove α è un parametro di forma, ottenuto empiricamente: $\alpha = (4 + cos(2\pi / n))/9$, F^0 una matrice $n \times n$, avente elementi: $F^0_{i,j} = (1-\alpha)/n$, $\forall i, j = 0,...,n$ e p il vettore dei punti attorno al vertice q preso in considerazione.
- Costruzione dei punti f_i^1: i punti f_i^1, giacenti sullo stesso piano del punto f^0, sono determinati dall'equazione: $f^1 = \alpha q + F^1 p$, dove: $F^1_{i,j} = (1 - \alpha + \beta cos(2\pi(j-i)/n))/n$, $\forall i, j = 1,...,n$. Il parametro di forma β serve per il controllo del modulo del vettore tangente e ha valore: $\beta = (1 + cos(2\pi / n))/3$.
- Costruzione dei punti f_i^2: i punti f_i^2 sono influenzati solamente dai quattro vertici dei triangoli aventi il lato in comune: $f_i^3 = \frac{1}{3}q + \frac{1}{6}p_{i-1} + \frac{1}{3}p_i + \frac{1}{6}p_{1+1}$, ovvero in forma compatta: $f^2 = \frac{1}{3}q + F^2 p$, dove:

$$F^{21}_{i,j} = \begin{cases} \frac{1}{3} & se\ j = i \\ \frac{1}{6} & se\ j = i-1, i+1 \quad \forall i, j = 1,...,n \\ 0 & altrimenti \end{cases}$$

  Tali punti contribuiscono alla definizione della curvatura nei vertici delle curve di contorno degli elementi..
- Costruzione dei punti f_i^3: i punti f_i^3, sono determinati nello stesso modo dei punti f_i^1, poiché non sono altro che punti di controllo f_i^1, riferiti tuttavia all'altro vertice del lato.
- Determinazione dei vettori tangenti v_i^0, v_i^1: i vettori tangenti v_i^0, v_i^1 sono determinati imponendo le condizioni di continuità G_1 e si scrive l'equazione in funzione di $V_i(t_i)$ e $\psi_i(t_i)$:

$$\frac{\partial p_i}{\partial t_{i+1}}(t_i, 0) = \phi_i(t_i)\frac{\partial p_i}{\partial t_i}(t_i, 0) + \psi_i(t_i)V_i(t)$$
$$\frac{\partial p_{i-1}}{\partial t_{i-1}}(t_i, 0) = \phi_i(t_i)\frac{\partial p_i}{\partial t_i}(t_i, 0) - \psi_i(t_i)V_i(t)$$

  dove $V_i(t_i)$ è una funzione che caratterizza il comportamento del piano tangente alla superficie, lungo il contorno $p_i(t_i, 0)$. Da queste equazioni, si ottengono i valori dei vettori tangenti: $v^0 = V^0 p$, dove: $V^0_{i,j} = 4\beta sin(2\pi(j-i)/n)/n$, $\forall i, j = 1, n$, ed inoltre: $v^1 = V^1 p$, essendo:

$$V_{i,j}^1 = \kappa_j V_{i,j}^0 \begin{cases} \dfrac{2\phi^0}{3\psi^0} & se\ \ j = i \\ \dfrac{-2\phi^0}{3\psi^0} & se\ \ j = i-1, i+1 \quad \forall\, i, j = 1,\ldots,n \\ 0 & altrimenti \end{cases}$$

avendo posto: $\kappa_j = 1 - \dfrac{1}{3\psi^0} tan\dfrac{\pi}{n}(6\phi^0 - \phi^1) + \psi^1$, ed i parametri contenuti rispettivamente:

$$\phi^0 = cos\left(\frac{2\pi k}{n}\right),\ \phi^1 = 1 - cos\left(\frac{2\pi k}{n}\right),\ \psi^0 = sin\left(\frac{2\pi}{n}\right) \text{ e } \psi^1 = sin\left(\frac{2\pi}{n_1}\right) - sin\left(\frac{2\pi}{n}\right)$$

Infine si noti come questo metodo comporti il notevole vantaggio di costruire superfici, solo localmente sensibili ad eventuali perturbazioni. Note le coordinate dei punti f^0, f^1, f^2 ed i vettori v_i^0, v_i^1, il problema del calcolo dei 28 punti di Bézier, costituenti il poliedro caratteristico (Fig. 10.5), è risolto seguendo un procedimento di innalzamento del grado del polinomio da $m = 4$ a $m = 6$.

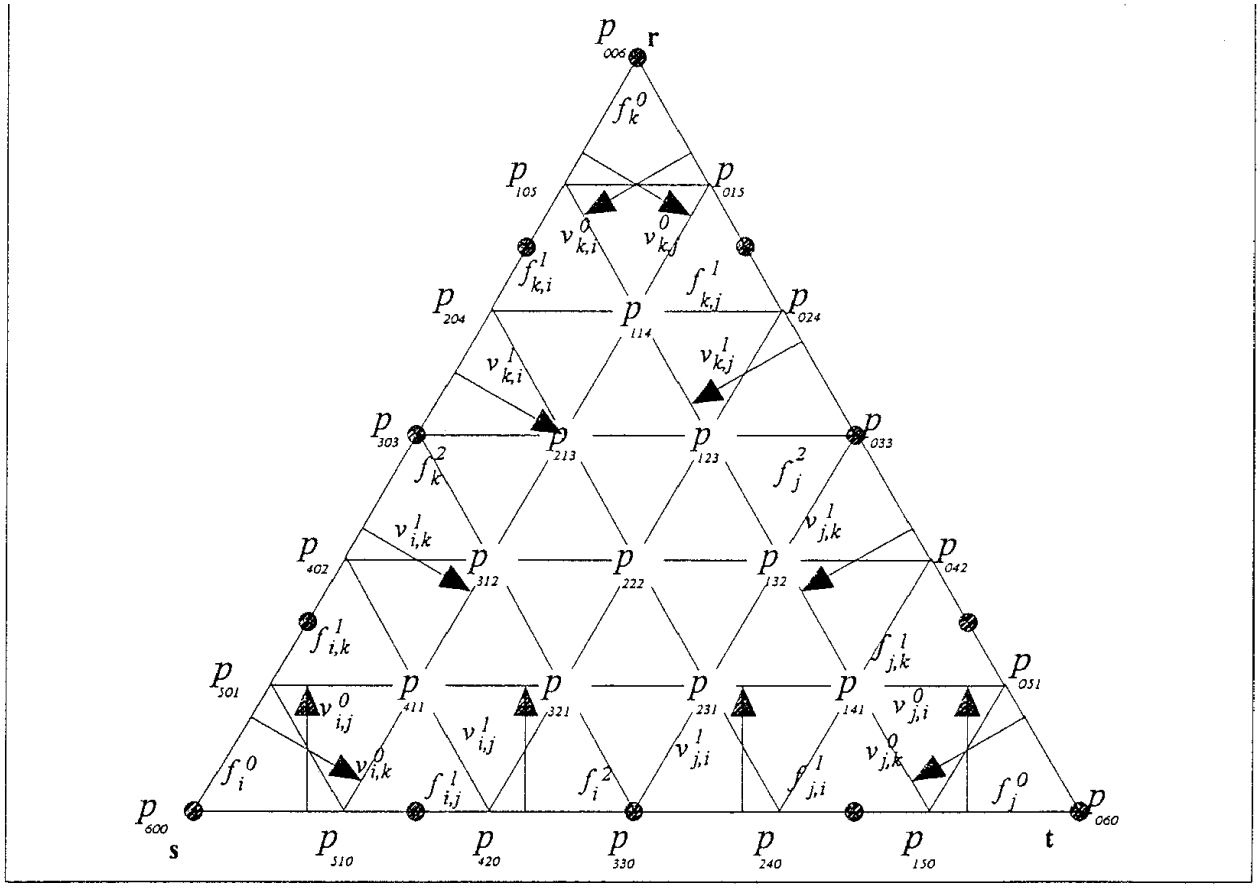


Fig. 10.5 – Coordinate parametriche dei punti di Bèzier, punti di controllo e vettori tangenti di un elemento

I valori delle coordinate dei punti di Bézier sul lato *s*, *t* sono dati:

$$p_{6,0,0} = f_i^0 \qquad p_{5,1,0} = \frac{1}{3} f_{i,j}^0 + \frac{2}{3} f_{i,j}^1 \qquad p_{4,2,0} = \frac{1}{15} f_{i,j}^0 + \frac{8}{15} f_{i,j}^1 + \frac{6}{15} f_{i,j}^2$$

$$p_{3,3,0} = \frac{2}{10} f_{i,j}^1 + \frac{6}{10} f_i^1 + \frac{2}{10} f_{i,j}^1 \qquad p_{2,4,0} = \frac{6}{15} f_i^2 + \frac{8}{15} f_{i,j}^1 + \frac{1}{15} f_j^2 \qquad p_{1,5,0} = \frac{2}{3} f_{j,i}^1 + \frac{1}{3} f_j^0$$

I punti di Bézier, successivi ai primi, si ottengono allo stesso modo:

$$p_{4,1,1} = \frac{1}{5} f_i^0 + \frac{4-2c_i}{5} f_{i,j}^1 + \frac{2c_i}{5} f_{i,j}^2 + \frac{s_i + s_j}{30} v_{i,j}^0 + \frac{s_i}{10} v_{i,j}^1$$

$$p_{3,2,1} = \frac{c_i + 1}{5} f_{i,j}^0 + \frac{6-c_j}{15} f_{i,j}^1 + \frac{9-3c_i}{15} f_{i,j}^2 + \frac{3c_i}{15} f_{j,i}^1 + \frac{s_i}{60} v_{i,j}^0 + \frac{s_i + s_j}{20} v_{i,j}^1 \frac{s_i}{10} v_{i+j}^1$$

$$p_{2,3,1} = \frac{c_i + 1}{15} f_{i,j}^1 + \frac{9-c_j}{15} f_{i,j}^2 + \frac{6-c_i}{15} f_{j,i}^1 + \frac{3c_i}{15} f_{j,i}^1 + \frac{c_i}{15} f_{j,i}^0 + \frac{s_j}{60} v_{i,j}^1 + \frac{s_i + s_j}{20} v_{j,i}^1 + \frac{s_i}{60} v_{j,i}^0$$

dove si ha (essendo n_i e n_j rispettivamente il numero dei vertici attorno ai vertici i e j):

$$c_i = cos\left(\frac{2\pi}{n_i}\right) \qquad c_j = cos\left(\frac{2\pi}{n_j}\right) \qquad s_i = sin\left(\frac{2\pi}{n_i}\right) \qquad s_j = sin\left(\frac{2\pi}{n_j}\right)$$

Infine determinati tutti i punti di Bézier, il calcolo del punto centrale è dato dalla seguente espressione:

$$\begin{aligned} p_{2,2,2} = {} & \frac{1}{3}\left(p_{3,2,1} + p_{3,1,2} + p_{2,1,3} + p_{2,3,1} + p_{1,3,2} + p_{1,2,3}\right) - \frac{1}{3}\left(p_{4,1,1} + p_{1,4,1} + p_{1,1,4}\right) + \\ & - \frac{1}{12}\left(p_{4,2,0} + p_{4,0,2} + p_{2,4,0} + p_{2,0,4} + p_{0,4,2} + p_{0,2,4}\right) + \\ & + \frac{1}{10}\left(p_{5,1,0} + p_{5,0,1} + p_{1,5,0} + p_{1,0,5} + p_{0,5,1} + p_{0,1,5}\right) - \frac{1}{30}\left(p_{6,0,0} + p_{0,6,0} + p_{0,0,6}\right) \end{aligned}$$

Benché la procedura possa apparire laboriosa, giova rilevare, come trattasi esclusivamente di una lunga, ma non difficile, sequenza di operazioni algebriche. I risultati sono abbastanza soddisfacenti, specialmente nel caso di triangolazioni regolari, dove si ottiene approssimativamente una superficie gradualmente variata nella unione degli elementi. Infatti una continuità C_2 è ottenuta solo approssimativamente grazie ad una triangolazione abbastanza equilatera con elementi sufficientemente uniformi.

12. Un'applicazione interessante

Diverse possono essere le applicazioni di quanto finora esposto, relativamente tanto modellazione solida, quanto alla ricostruzione di superfici. A tale proposito, tre esempi significativi sono presentati nella prossima quarta parte del presente lavoro. Tuttavia a conclusione di questa terza parte, una prima applicazione diretta è certamente d'interesse e serve a mostrare i due passi di un problema di ricostruzione di una superficie complessa ed implicitamente, almeno per questo caso specifico, anche di modellazione solida, perché la

superficie ricostruita racchiude, in sé, un corpo 3D di cui si è pure ricostruito il modello digitale. Infatti una superficie che avvolge un corpo, a sua volta complesso, richiede per la sua ricostruzione la triangolazione della superficie esterna (o la tetraedrazione completa dell'intero corpo, se sono disponibili anche punti interni al corpo stesso) cui far seguire un'interpolazione della suddetta superficie esterna, più raffinata, mediante l'uso di curve e superfici parametriche, gradualmente variate.

L'esempio proposto riguarda la riproduzione di una delle statue più antiche, sita su una guglia del Duomo di Milano. A tale riguardo, corre il dovere di segnalare come le statue più antiche, del periodo gotico, a differenza di quelle successive (a partire già dal più tardo periodo barocco), siano finite in tutte le loro caratteristiche morfologiche, a prescindere dal posto in esse siano poi collocate e da cosa effettivamente si possa vedere di loro da lontano. Allora conservare le statue dall'inevitabile degrado significa collocarle e proteggerle in un museo, ponendo copie al loro posto originale. Tuttavia non alterare il monumento nella sua originale complessità significa collocare copie davvero fedeli degli originali rimossi.

Una strada possibile per produrre copie fedeli e far uso di strumenti a tastatore con fresa rigidamente accoppiata. Tuttavia questa strada, frequente per la riproduzione di minuteria meccanica, è complessa per quanto riguarda una statua di consistenti dimensioni, dovendo passare da pochi centimetri a parecchi metri, nel contempo, garantendo precisioni pressoché comparabili. Allora un'altra via praticabile è quella d'adire ad una ricostruzione di superficie ed ad un modello digitale di un corpo 3D, a valle dell'acquisizione e della restituzione di immagini fotogrammetriche (come mostrato nella mosaicatura di immagini di fig. 12.1) e/o di nuvole di punti tracciati mediante scansione laser.

Infatti entrambi i mezzi tecnologici e maggiormente se usati insieme, in quanto complementari tra loro, hanno ottime capacità a riguardo. Dopodiché la ricostruzione di superficie si effettua dapprima mediante la triangolazione della superficie esterna del corpo 3D e successivamente mediante un'interpolazione della suddetta superficie esterna, più raffinata, attraverso l'uso di curve e superfici parametriche, gradualmente variate. A mo' d'esempio, l'applicazione è ristretta alla sola ricostruzione della superficie esterna di una testa, (come mostrato nell'immagine di fig. 12.2 e dalle curve di livello di fig. 12.3) certamente il particolare più complesso ed interessante dell'intera statua, come mostrato in due diverse viste nelle figure 12.4 e 12.5.

Fig. 12.1 – Mosaicatura di immagini della statua rilevata del Duomo di Milano

Di seguito, è sommariamente descritta l'intera procedura seguita, per giungere al risultato voluto. La statua è stata segnalizzata con un certo numero di target retro – riflettenti (alcuni dei quali dotati di mira) per la presa e l'appoggio delle immagini. Le operazioni successive hanno portato, mediante correlazione di immagini (o matching), all'orientamento delle suddette immagini ed alla restituzione della statua per punti di spezzoni di curve di livello.

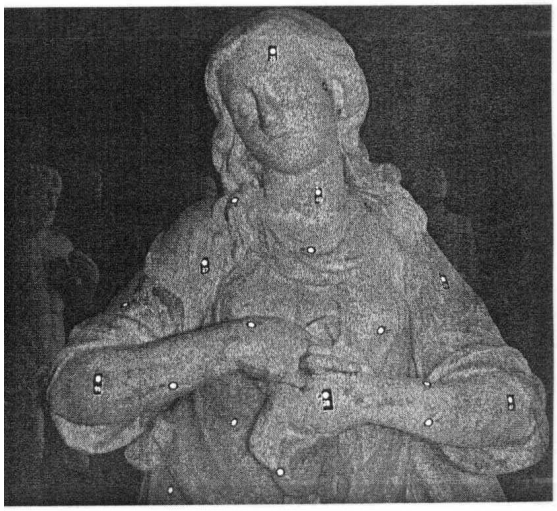


Fig. 12.2 – Un'immagine digitale di un particolare della statua

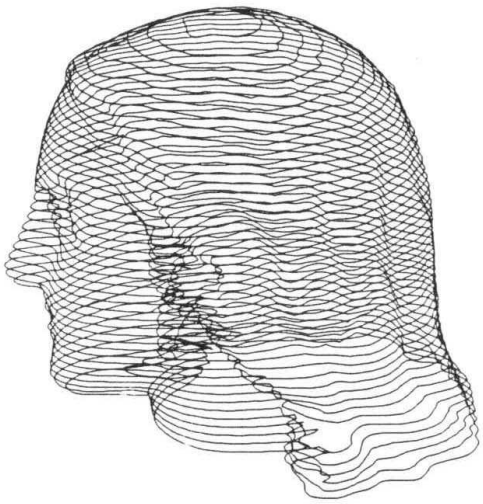


Fig. 12.3 – Rappresentazione a curve di livello della testa della statua

Prima di procedere alla triangolazione dei dati, su ciascuna curva è stata estrapolata una serie di punti, in modo tale da ottenere una triangolazione sufficientemente regolare (e comunque non troppo dissimile da una triangolazione di Delauny). Infine a partire dai vertici della triangolazione (la cui approssimazione risulta insoddisfacente), si è costruita un'interpolazione con funzioni spline a supporto triangolare, migliorando insieme rappresentazione / visualizzazione e riproduzione dei dati.

Un giudizio sulla bontà del risultato ottenuto è dato dai seguenti numeri. A fronte di una precisione d'appoggio di circa $0.2\ mm$, le operazioni di presa, orientamento, restituzione, interpolazione delle linee,

triangolazione ed interpolazione della superficie hanno comportato un errore complessivo non superiore a 3 mm, come richiesto dalle macchine a controllo numerico.

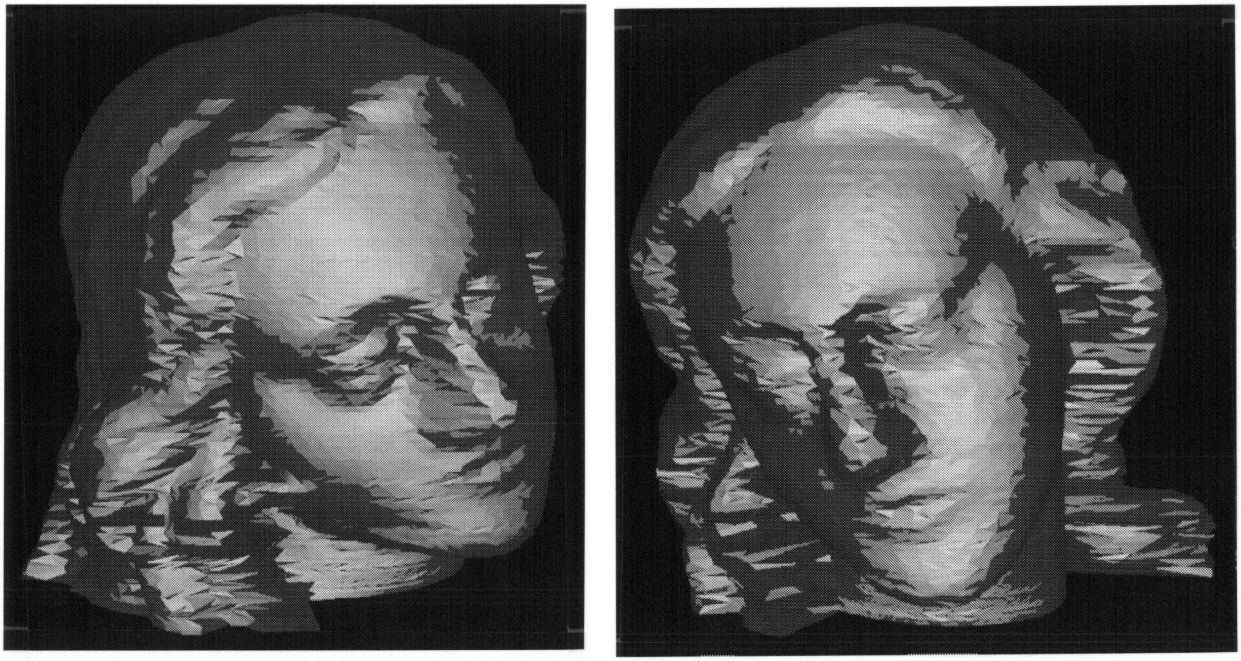


Fig. 12.4 – Due diverse viste della triangolazione della testa della statua

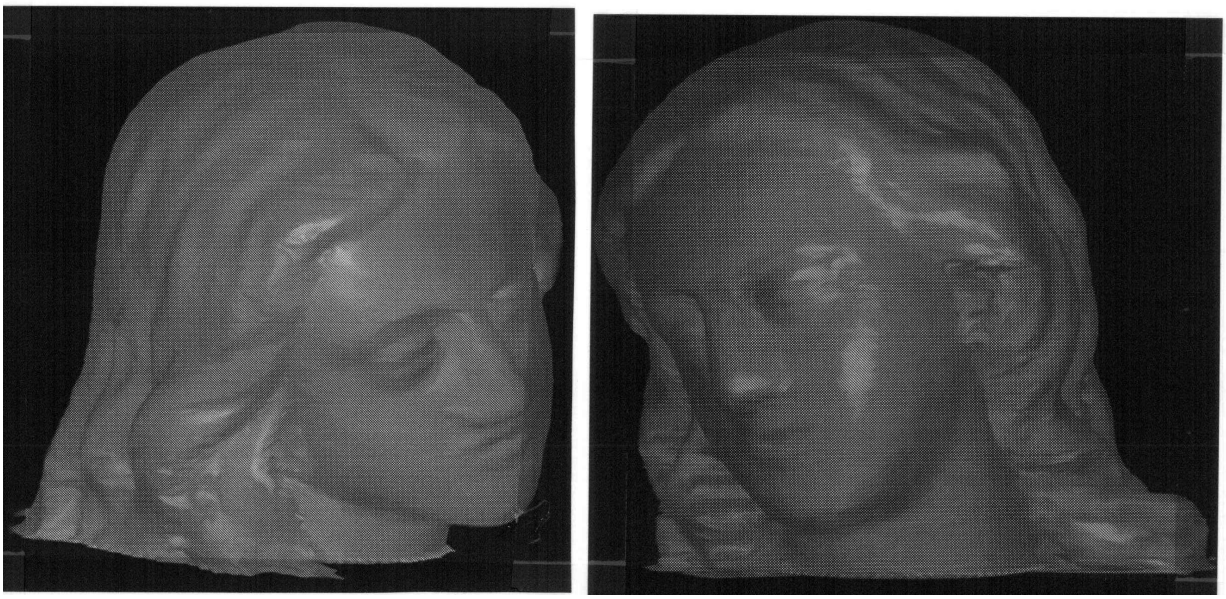


Fig. 12.5 – Due diverse viste dell'interpolazione con funzioni spline a supporto triangolare

PARTE IV – QUATTRO ESEMPI DI MODELLAZIONE SOLIDA E RICOSTRUZIONE DI SUPERFICI

La modellazione di oggetti non è una delle parti più semplici e scontate del trattamento delle osservazioni, per quanto riguarda sia l'analisi dei dati che la statistica computazionale. Per queste ragioni, si ritiene utile mettere in luce le sue potenzialità, ricorrendo alla presentazione di quattro esempi, con l'intento di mostrare:

- la modellazione solida di un corpo 3D da riprese di fotogrammetria terrestre;

- la ricostruzione di superficie (2D e ½) da immagini digitali telerilevate:
  - ottiche – pancromatiche,
  - complesse da telemetria radar ed inteferometria;
- la ricostruzione di superficie (2D e ½) da digitalizzazione di cartografia tecnica esistente.

L'obiettivo degli esempi è mostrare chiaramente le potenzialità dei metodi presentati che fanno da supporto matematico – statistico – informatico ai principali programmi di informatica grafica, comunemente in uso.

A. MODELLAZIONE SOLIDA DA RIPRESE DI FOTOGRAMMETRIA TERRESTRE

Un modello geometrico solido è una rappresentazione matematica non – ambigua e completa della forma di un oggetto fisico. La necessità di elaborare modelli sempre più complessi e la mancanza di un processo di unificazione e formalizzazione per la modellazione geometrica di corpi 3D lascia ampie opportunità per definire nuovi approcci. La teoria di modellazione rigorosa permette di sfruttare completamente la potenza dei matematismi per analizzare i processi di definizione e modifica.
Solidi complessi possono essere così generati dalla combinazione di solidi semplici, purché sia rispettata la rispondenza ad alcuni criteri formali e ad alcune restrizioni topologiche che ne assicurino la validità. In prima istanza, è necessario estrarre dagli oggetti solidi un insieme di caratteristiche geometriche: superficie e forma dell'oggetto.

- La forma di un solido è determinata dall'insieme totale di punti definiti nello spazio.
- La superficie di un oggetto è un sottoinsieme di questi punti e costituisce la frontiera tra l'oggetto e lo spazio altro (in generale, esterno, ma non solo, per la possibile presenza di cavità interne).

Esistono metodi differenti per creare i solidi semplici e possono essere classificati in:

- rappresentazioni wireframe;
- forme parametrizzate;
- rappresentazioni mediante sweep;
- solidi generati da trasformazioni non – lineari.

Per la creazione di modelli più complessi invece, grazie ad operatori booleani o ad altri sistemi di assemblaggio si distinguono:

- la geometria solida costruttiva (CSG);
- le rappresentazioni al contorno;
- la scomposizione in celle;
- la più recente geometria relazionale.

Alla base di questi metodi, esistono sempre i concetti di modello fondati sui grafi e booleani.

1. Modelli basati sui grafi

Si fa riferimento ai modelli geometrici che utilizzano la struttura topologica e, attraverso puntatori, collegano tra loro i dati relativi a facce, spigoli e vertici di un oggetto. Un oggetto solido può essere descritto mediante una lista di facce (con le rispettive equazioni delle superfici), composte dalla sequenza di spigoli che le determinano (con le rispettive equazioni di curve), con puntatori agli estremi, individuati dai vertici dell'oggetto e condivisi dalle facce stesse. I vertici a loro volta sono rappresentati da liste di coordinate con puntatori agli spigoli incidenti in ciascun vertice. Pertanto esistono due tipi di informazioni:

- i puntatori che definiscono la topologia o il tipo di connessione tra facce, spigoli e vertici;
- i dati numerici che rappresentano le equazioni di superfici e curve e le coordinate dei vertici.

Le trasformazioni di scala e i movimenti rigidi (traslazioni e rotazioni) alterano solo i dati numerici, lasciando invariati i puntatori. Interessanti differenze tra i vari sistemi basati su grafi mostrano, oltre alla scelta del tipo di ridondanze, la definizione di diversi approcci anche nella disposizione degli spigoli di una faccia. Poiché tutti gli spigoli sono costituiti da segmenti di retta, possono essere memorizzati per mezzo di una coppia di vertici, mentre le facce tramite una lista ordinata di vertici. In questo caso, il poliedro può essere trattato come un grafo semplice e rappresentato da una lista di vertici, identificata da una matrice relativa alle coordinate, e da una matrice, associata a ciascuna connessione, chiamata matrice di connessione (fig. 4.1).

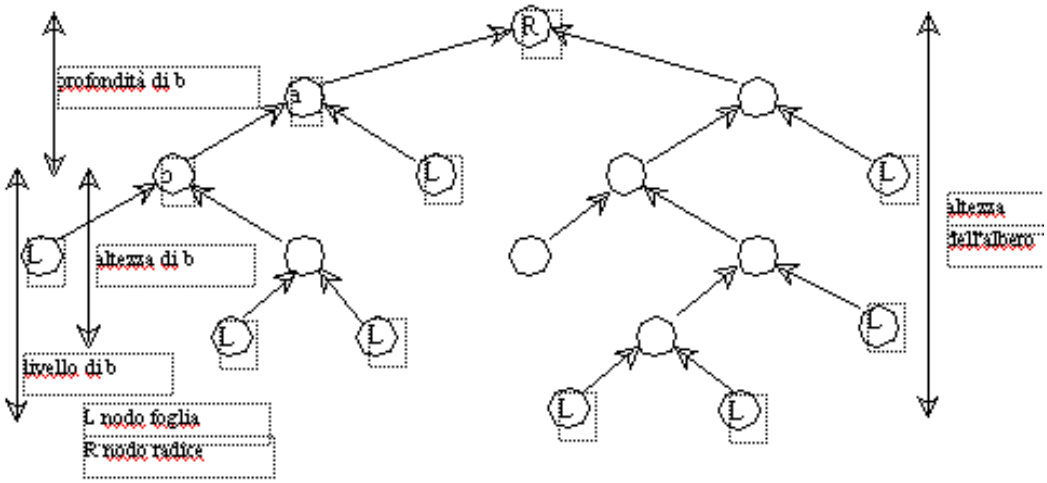


Fig. 4.1 – Esempio di albero binario

La matrice di connessione contiene solo valori binari: gli elementi con un valore corrispondente a zero indicano la mancanza di connessione mentre quelli con valore pari a uno denotano l'esistenza di una connessione tra la coppia di vertici a essa associata. Questa matrice può essere ottenuta per descrivere relazioni intercorrenti anche tra altre coppie di elementi (ad es., le facce del poliedro), definendo così un albero un grafo diretto, privo di circuiti.

2. Modelli booleani

Si parla di modelli booleani quando i solidi più o meno complessi sono definiti dalla combinazione booleana di solidi semplici. Tali relazioni di unione e intersezione non forniscono informazioni sul nuovo solido creato, ma specificano le combinazioni delle primitive solidi costituenti. In questo caso non vi è alcun riferimento alle

coordinate dei vertici, agli spigoli e alle facce del solido creato, ma solo a quelli di partenza, motivo per cui si dice che il modello booleano è una rappresentazione procedurale o anche un modello non valutato. Ulteriori informazioni saranno possibili solo dopo una valutazione del modello attraverso cui, ad esempio, si calcolano le intersezioni per determinare gli spigoli nuovi ed i vertici nuovi, e si analizzano le nuove connessioni. Si comprendono così le caratteristiche topologiche del nuovo elemento e si definisce la rappresentazione del nuovo contorno. In molti sistemi, le primitive sono memorizzate in un modello basato su grafo e, una volta definite, costituiscono elementi primitivi unitari (o parametrizzati).

I calcoli delle intersezioni sono importanti nella valutazione del contorno di un modello booleano. Se sono definiti i contorni completi di due primitive, il contorno completo dell'unione dei due solidi è dato dalla somma dei segmenti di contorno. Tali segmenti rappresentano delle regioni attive su entrambi e sono essi stessi limitati dall'intersezione dei contorni originari delle primitive,

Il solido (comprese le primitive), in questo contesto, è inteso come una collezione di punti, alcuni dei quali costituenti l'interno, altri l'esterno del contorno. In un modello booleano è molto importante l'ordine con cui sono effettuate le operazioni. Infatti esistono situazioni difficili e potenzialmente patologiche che un sistema di modellazione geometrica deve saper superare durante la fase di valutazione del contorno di una forma complessa. Alcune generalizzazioni consentono, attraverso algoritmi, di velocizzare calcoli e verifiche della validità. Lo schema di figura 4.2 rappresenta il punto d'arrivo di una ragionevole interpretazione delle relazioni che intercorrono all'interno di una certa struttura formale di dati spaziali.

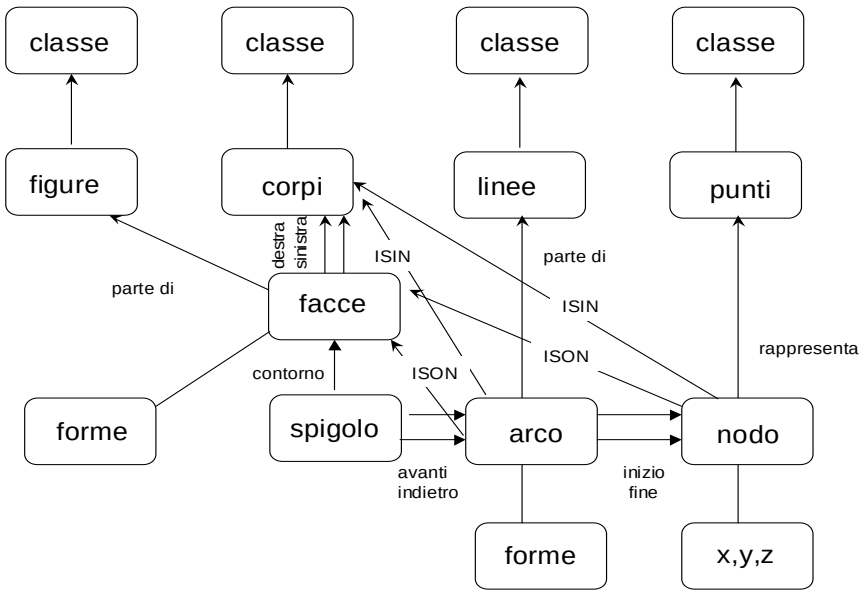


Fig. 4.2 – Esempio di struttura formale dei dati

3. Sistemi wireframe

I sistemi *wireframe* rappresentano l'oggetto attraverso i suoi spigoli componenti; pertanto sono costituiti solamente da punti, rette e curve, entità tipiche dei sistemi 2D. Quando sono nati, negli anni settanta del '900, non consentivano una rappresentazione analitica dei solidi. Con l'attribuzione della profondità è stato possibile effettuare traslazioni e rotazioni tridimensionali, pur mancando dell'analisi volumetrica automatica. Successivamente sono stati introdotti piani e altre superfici semplici (come sfere, cilindri e coni), diventando

potenti modellatori, con ampie possibilità d'applicazione ad altre aree tematiche (come l'analisi ad elementi finiti ed il controllo numerico nella produzione). Tuttavia la pura rappresentazione wireframe presenta alcuni problemi, perché può creare modelli ambigui e privi di senso, per mancanza di coerenza grafica o visiva. Inoltre tali modelli risultano deboli nel calcolo delle proprietà di massa, nel sezionamento, nell'eliminazione delle linee nascoste e nell'associazione (e memorizzazione) degli attributi relativi ad una superficie.

4. Forme parametrizzate

E' un metodo che consente di definire una nuova forma, applicando una semplice trasformazione lineare ad una già esistente. Ad esempio, se si parte da un cubo, si può ottenere un parallelepipedo, utilizzando un operatore di scala differenziato sulle tre componenti tridimensionali. Tali trasformazioni influiscono sulla geometria, ma non sulla topologia della forma. Possono essere creati anche solidi più complessi (ad es., una sezione ad S), purché siano rispettate alcune restrizioni. Un piccolo insieme di dimensioni chiave è normalmente sufficiente per definire una classe di oggetti relativamente semplici. Se ogni dimensione è una variabile indipendente, è possibile generare una forma particolare, definita all'interno di una classe.

5. Rappresentazioni sweep

Le rappresentazioni *sweep* si basano sul concetto di spostamento di un punto, una curva od una superficie lungo un percorso. E' possibile così modellare, con relativa semplicità, oggetti in una, due o tre dimensioni (come mostrato dall'esempio di figura 4.3). Sono richiesti due elementi: un oggetto da spostare (curva, superficie, solido), detta generatrice, ed una traiettoria attraverso la quale muovere l'oggetto (un cammino definibile analiticamente), detta direzione. Esistono due movimenti principali di sweep: la traslazione e la rotazione. Nel primo caso (detto di estrusione), si ottiene una superficie ed un solido muovendo una curva o una forma piana lungo una linea normale al piano stesso; nel secondo caso (detto di rotazione) si ottengono superfici di rotazione, facendo ruotare una curva generatrice intorno ad un asse.

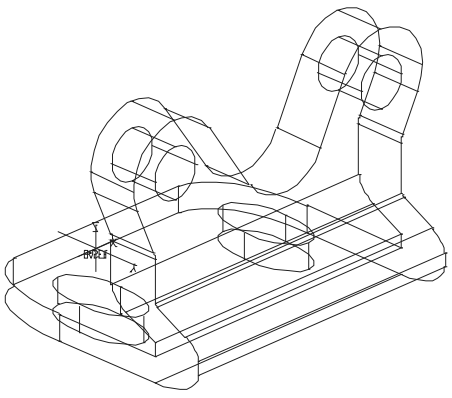


Fig. 4.3 – Rappresentazione sweep per la modellazione di un pezzo meccanico

6. Solidi generati da trasformazioni non lineari

Si possono usare trasformazioni non lineari per ottenere nuovi oggetti dalla deformazione di forme di origine. Le trasformazioni *assiali* controllano scala, curvatura, torsione e stiramento riferendosi all'asse lungo il quale si introduce la deformazione. Le trasformazioni bivariate permettono di ottenere forme nuove dalla

deformazione introdotta nel piano parametrico in cui tali superfici sono incluse; analogamente nello spazio è possibile ottenere un solido trivariato a partire da una curva inclusa in uno spazio parametrico normalizzato.

7. Geometria solida costruttiva (CSG)

E' una tecnica, detta anche di costruzione per blocchi, che consente di definire solidi complessi come composizione di oggetti più semplici, attraverso operatori booleani di unione, differenza ed intersezione. Le rappresentazioni CSG degli oggetti sono alberi binari ordinati i cui nodi terminali individuano primitive o valori di trasformazione per movimenti rigidi, mentre i nodi non terminali rappresentano operatori booleani o movimenti rigidi che agiscono sui loro due sotto – nodi. La radice rappresenta l'oggetto finale. Se gli elementi primitivi di un sistema di modellazione sono oggetti validi limitati e definiti dal sistema e se gli operatori di composizione sono regolari, i modelli solidi risultanti sono anch'essi validi e limitati. Si possono usare anche primitive definite dall'utente, purché ne sia stata verificata la validità. Solitamente i sistemi CSG utilizzano i tre operatori booleani tradizionali di unione, differenza e intersezione, ma a questi operatori, possono aggiungersi altri operatori come l'incollaggio ed il complemento. I sistemi di modellazione più potenti generano due rappresentazioni per ogni solido:

- la prima è una rappresentazione di GSC, con una struttura dati ad albero binario come già detto;
- la seconda è costituita da una rappresentazione del contorno, descritta tramite facce, spigoli e vertici.

Quest'ultima possiede informazioni topologiche sulla connettività degli elementi di contorno e dei dati numerici che descrivono la forma geometrica e la posizione di tali elementi. La rappresentazione del contorno è calcolata a partire da quella costruttiva, per mezzo di algoritmi di valutazione del contorno che determinano dove le facce sono limitate e dove sono creati o eliminati spigoli e vertici.
Spesso sono necessari molti calcoli per generare la rappresentazione del contorno a partire da una rappresentazione costruttiva, ma quest'ultima contiene una considerevole quantità di informazioni utili per il calcolo delle proprietà globali, la visualizzazione ecc. La forma costruttiva è anche chiamata modello procedurale non – valutato (un esempio in figura 4.4), mentre la forma al contorno è detta modello valutato.

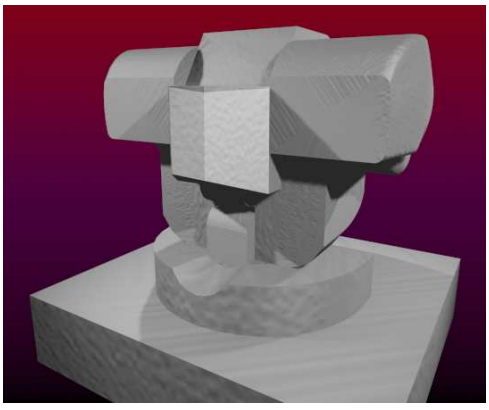


Fig. 4.4 – Rappresentazione mediante la Geometria solida costruttiva (CSG)

8. Rappresentazione del contorno

Dal punto di vista computazionale, è conveniente suddividere la superficie di contorno del solido in facce (o patch) ciascuna delle quali limitata da spigoli o vertici. Le facce sono una forma di rappresentazione, non

sempre corrispondente alla nostra intuizione. Pertanto si pongono alcune necessarie condizioni; infatti mentre il contorno di un cubo è suddiviso in sei facce piane, quello di un cilindro è segmentato in modo arbitrario. Le facce piane inoltre possono essere rappresentate dai loro spigoli, mentre le facce curve richiedono un maggior numero di informazioni (in ogni caso, esse possono essere regioni appartenenti a superfici di Bézier, dove la superficie e le curve su di essa sono definite da coefficienti geometrici). Per indicare su quale lato di uno spigolo (od una curva di delimitazione) giace la faccia si adottano alcune convenzioni, ad es., la curva di delimitazione di una faccia è parametrizzata relativamente ad una direzione coerente, in modo che il vettore indichi il lato della faccia associato alla curva.
Come già detto in precedenza, il contorno di un oggetto può essere segmentato in facce, spigoli e vertici in un numero infinito di modi, non esistendo una rappresentazione una del contorno stesso. Pertanto talvolta per evitare confusione, durante la segmentazione del contorno, si introducono i termini di struttura combinatoria e informazione metrica, in generale tra loro dipendenti, dove le condizioni di validità devono essere stabilite per entrambi.

9. Scomposizione in celle

Questo procedimento rappresenta i solidi come la somma (o l'unione) di un insieme di celle fino a raggiungere una predeterminata approssimazione di descrivibilità. Tale metodo di rappresentazione è utile in tutti i casi in cui la difficoltà nel descrivere l'oggetto intero può essere ovviata da una scomposizione dello stesso in oggetti più semplici. Infatti esistono molti modi, tutti non ambigui, per scomporre un solido in celle componenti. La scomposizione in celle, regolarmente utilizzata per l'analisi strutturale, costituisce la base della modellazione che utilizza gli elementi finiti.
L'enumerazione, per occupazione spaziale, rappresenta un caso particolare della scomposizione in celle in cui queste ultime (*voxel*) sono di forma cubica e collocate in una griglia spaziale fissata. Un metodo semplice ed efficace di rappresentazione dell'insieme delle celle (la cui dimensione determina la massima risoluzione del modello, altrimenti detta granularità) consiste nell'elencarne le coordinate dei centri. Se da una parte è assicurata l'unicità spaziale ed è facile accedere a un dato punto, dall'altra non esistono esplicite relazioni topologiche e/o geometriche tra le parti dell'oggetto.
Inoltre tali schemi inoltre richiedono una grande quantità di dati da memorizzare, perchè tutte le celle di un oggetto vengono riempite con uno dei due valori binari. Per ridurre la ridondanza, manipolare velocemente e visualizzare gli oggetti solidi con la computer graphics, è stato sviluppato un metodo di modellazione solida (basato sulla codifica *octree*, come mostrato in figura 4.5).

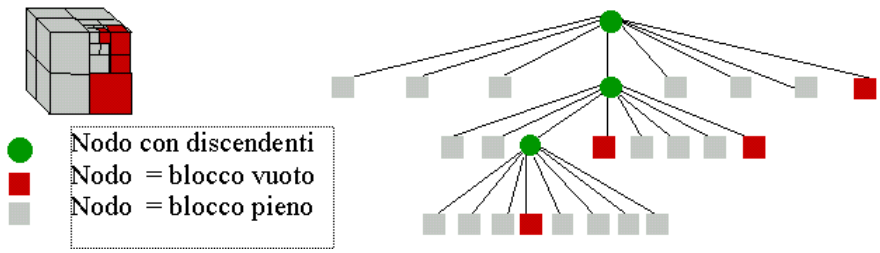


Fig. 4.5 – Rappresentazione Octree

Tale approccio utilizza una struttura gerarchica ad albero ottagonario, per rappresentare i solidi, e si serve di algoritmi che crescono solo linearmente con la complessità dell'oggetto. La codifica di un modello è analoga

a quella quatree di una figura (piana): una regione cubica è suddivisa ricorsivamente in ottanti, ovvero in otto regioni cubiche. Ogni nodo, purché non sia un nodo terminale, detto foglia, possiede otto discendenti.

10. Strutture ibride

La pressante richiesta di una maggiore rispondenza della modellazione alla complessità dei fenomeni reali sta velocizzando le risposte del mondo della ricerca. Pertanto alcune sperimentazioni sono maggiormente dirette verso l'integrazione dei due tipi di struttura precedentemente analizzati, al fine di incrementare l'efficienza delle operazioni sui dati tridimensionali.

In sintesi, i due gruppi di modelli possono essere distinti in base al tipo di rappresentazione basata sulla superficie o sul volume:

- nel primo caso, le caratteristiche geometriche degli oggetti sono descritte utilizzando piccole celle di superficie o primitive grafiche;
- nel secondo caso, la rappresentazione geometrica è basata sul volume.

Ognuno dei sistemi possiede vantaggi e svantaggi nella gestione di dati 3D. Una integrazione dei due modelli potrebbe mantenere i vantaggi di entrambi. Coloro che lavorano nell'ambito dei sistemi informativi a referenza spaziale utilizzano spesso ricerche sviluppate nell'ambito dell'informatica grafica, ma le esigenze si rivelano molto differenti. Infatti le dimensioni e caratteristiche degli oggetti descritti, la topologia e gli attributi, e le analisi spaziali rendono indispensabile una struttura dati, specificamente creata per i suddetti sistemi informativi, e la necessità di derivare i contorni da un oggetto spaziale (ad es., un metodo per identificare i contorni più esterni da un oggetto strutturato in octree). Il modello ibrido è stato costruito come una integrazione dei due modelli precedentemente descritti (come mostrato dalla figura 4.6):

- uno basato sulla superficie, adatto all'analisi e visualizzazione di superfici in uno spazio 3D;
- un secondo basato sul volume, per le operazioni e analisi di dati 3D, nella loro totalità, e costituente un modello digitale dei volumi (DVM, cioè Digital Volume Model).

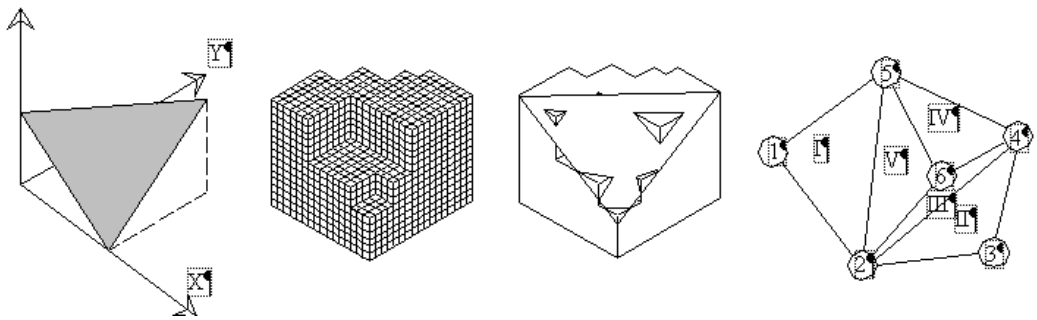


Fig. 4.6 – Da sinistra a destra: faccia triangolare in un TIN;
modello octree;
modello ibrido;
organizzazione dei dati di un modello ibrido

Nel modello ibrido, ogni faccia del TIN è designata per contenere il proprio ottante. Il TIN è principalmente usato per la visualizzazione e la ricerca di relazioni topologiche; pertanto individuata una o più sfaccettature, è decodificato il dato octree per le operazioni spaziali.

In questo modello i triangoli sono oggetti spaziali di base ed ogni triangolo ha i suoi attributi topologici che descrivono i nodi, i triangoli adiacenti ed i puntatori agli elementi octree più prossimi. L'interfaccia interattiva consente all'utente di controllare il sistema e ottenere risultati. Il modulo di modellazione tridimensionale è provvisto di metodi diversi:

- octree;
- reti di tetraedri (TEN);
- *Nurbs*: funzioni parametriche con struttura di vettori per le osservazioni disposte nello spazio interno.

E' altresì possibile una conversione dei dati tra i differenti tipi di struttura ed anche la riduzione ad un unico tipo di struttura di dati provenienti da diverse sorgenti, in quanto un database relazionale ne consente vantaggiosamente la gestione.

11. Relational Geometric Synthesis

Una struttura per la modellazione solida orientata ad oggetti memorizza le relazioni (contenute nel grafo diretto) tra gli oggetti e le preserva, in fase di editing, facilitandone modifiche ed gli aggiornamenti. I nodi del grafo sono gli oggetti geometrici, mentre gli archi rappresentano le relazioni tra gli stessi. Tutte le parti del modello possono essere costruite in modo sequenziale, dalla più semplice alla più complessa, a partire da oggetti costituiti da tre entità: punti, curve e superfici. Ad ogni oggetto, sono associate le relazioni (dimensione, posizione ed orientamento) che lo legano, in termini di dipendenza o supporto, alle altre parti del modello, in modo da conservarle, in fase di editing e trasformazione. Il grado di ingresso dell'oggetto indica il numero dei suoi supporti, il grado di uscita, il numero degli oggetti che da esso dipendono.
Questo metodo di modellazione consente di costruire gli oggetti in maniera più intuitiva e di rendere più semplici le intersezioni tra superfici (si può partire dalle curve che descrivono le intersezioni, per poi definirne le superfici che si intersecano), e permette di creare modelli relazionali che conservano le proprietà topologiche e geometriche, con molti gradi di libertà, attraverso variazioni parametriche. Inoltre esso estende il campo di applicazione della modellazione a tutte le forme libere (dalle superfici scultoree, al design di imbarcazioni – fig. 4.7 – ed aerei, ecc.) la cui acquisizione sta diventando sempre più diffusa, grazie ai sistemi di scansione laser, aprendo nuove prospettive per applicazioni realmente tridimensionali

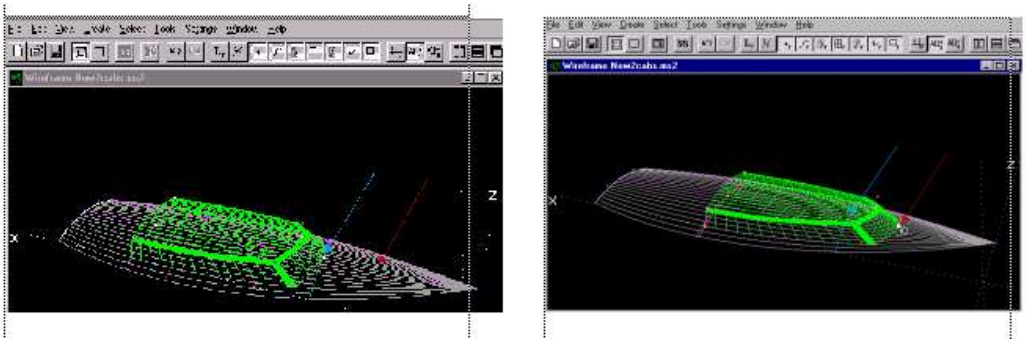


Fig. 4.7 – Relational Geometric Synthesis

12. Un'applicazione ad una struttura scultorea complessa

Dal rilevamento e la restituzione della fontana sita in Piazza Leonardo da Vinci di Milano, opera dello scultore Giuseppe Cascella. Gli output cartografici, frutto della restituzione fotogrammetrica, consistono in

una mappa 3D di punti, in un modello digitale di una parte (scomposta per simmetria) della parte visibile della fontana ed una ortoimmagine completata dalla superimposizione dell'altimetria. A partire dalla mappa 3D di punti, si sono applicate alcune delle metodologie e procedure, descritte ed illustrate precedentemente, che hanno alla definizione di tutti gli elementi essenziali per una sua modellazione solida.

La figura 4.8 mostra la scomposizione della fontana in solidi elementari e, per ogni solido in cui è stata scomposta la suddetta fontana, l'approssimazione lineare di ciascun solido da cui possono derivare le matrici di incidenza, adiacenza e cross - connessione e le tavole di corrispondenza della numerazione dei vertici comuni a solidi adiacenti, per l'eventuale inserimento la modellazione solida prodotta in un sistema informativo a referenza spaziale, come opportuno per provvedere al catasto dei beni culturali ed ambientali.

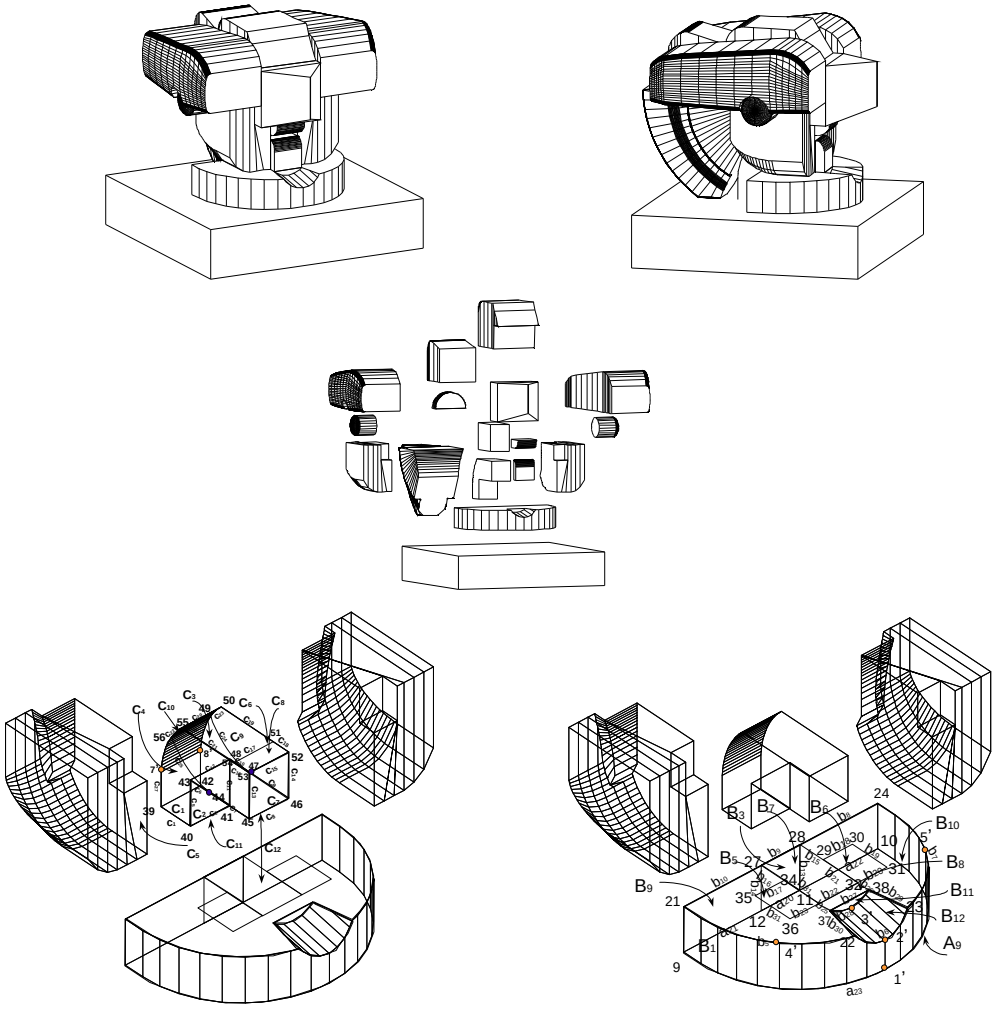


Figura 4.8 – Un oggetto complesso, la sua scomposizione in solidi più semplici e la successiva approssimazione lineare di ciascun solido

In conclusione resta da osservare come i modelli 3D, più complessi e strutturati, superino i limiti intrinseci dei modelli 2D, giocoforza più semplici, dove l'altimetria è trattata separatamente come un attributo, seppur metrico. Infatti questo approccio, tipico della cartografia tradizionale, mostra i suoi limiti più evidenti, quando si vogliono rappresentare oggetti complessi, quali i corpi concavi, non – stellati ed eventualmente pluriconnessi. Pertanto i diagrammi di Voronoi, la triangolazione / tetraedrazione di Delaunay e gli elementi finiti (ad es., facendo uso delle curve parametriche di Catmull – Rom e delle funzioni spline a supporto triangolare di Bézier) sono strumenti ben rispondenti alla modellizzazione di manufatti od oggetti.

B. RICOSTRUZIONE DI SUPERFICI CON DATI DI IMMAGINI OTTICHE TELERILEVATI

Il satellite francese SPOT 2 fornisce immagini telerilevate, nella banda ottica, coprendo l'intero globo, in circa nove giorni e mantenendo un assetto ed un piano orbitale, quasi polare (come si ha avuto modo di mostrare in altri lavori cui si rimanda). La precisione ottenibile nel posizionamento geodetico:

- è dell'ordine di una decina di metri;
- ma, con adeguato trattamento delle osservazioni, tale precisione può scendere fino al metro.

Da tutto ciò, risulta la possibilità pratica di rappresentare l'intero pianeta, quanto meno alla scala 1:250.000, dotandolo così di una carta di base a piccola scala (di cui il mondo, a tutt'oggi, è sprovvisto per intere regioni e di cui la comunità tutta ha grande ed urgente bisogno). Alla misura automatica dei punti sulle immagini digitali, è seguito il trattamento delle osservazioni raccolte, eseguito in due tappe distinte:

- innanzitutto la compensazione della triangolazione aerea con metodo delle stelle proiettive, avendo cura di introdurre condizioni di regolarizzazione dei parametri d'assetto, nella compensazione stessa, in quanto ogni linea di pixel deve considerarsi a sé stante, poiché proveniente da uno scanner lineare e non da una CCD [9] (in questo caso, la stima di tanti insiemi di parametri d'assetto, quante sono le linee prese in considerazione, darebbe luogo ad un sistema normale mal – condizionato, mentre condizioni di regolazzazione – che esprimono continuità e regolare variabilità dei parametri d'assetto – determinano un sistema ben – condizionato);
- successivamente per l'elaborazione delle immagini digitali, adottando il metodo della piramide di immagini a risoluzione multipla, cioè passando da una grana molto grossa, dove è possibile definire solo una forma approssimativa della rappresentazione, a grane sempre più raffinate, fino ad ottenere non solo le informazioni planimetriche, contenute nelle registrazioni di partenza, ma anche i dati altimetrici.

La geometria di presa del test eseguito consiste in:

- quattro strisciate di immagini SPOT;
- punti di appoggio, forniti dal Servizio Geodetico Francese
- griglie di interpolazione, sovrapposte al blocco di immagini dallo spazio,

ed il prodotto finale è una carta, in forma di ortoimmagine digitale, con la ricostruzione dell'altimetria del terreno ottenuta, nell'ordine, mediante (come si ha avuto modo di mostrare in altri lavori cui si rimanda):

- una triangolazione di Delauny;
- un'approssimazione ad elementi finiti con funzioni spline bicubiche.

Anche in questo esempio, risalta il legame intrinseco fra alta tecnologia, statistica, calcolo numerico, rilevamento e rappresentazione del territorio. Infatti sarebbe stato vano disporre di dati ed elaborazioni, se non si fosse applicato un metodo di calcolo, originato dalle tecniche multilivello, e se non si fossero fatti stime di parametri di modelli e test statistici, necessari alla compensazione delle strisciate ed alla ricostruzione della superficie del terreno.

[9] Si osservi come la presenza condizioni di regolarizzazione sia necessaria anche per camere con scanner tri – lineari e, maggiormente necessaria, nel caso di immagini raccolte, pixel a pixel, tramite uno specchietto rotante.

C. RICOSTRUZIONE DI SUPERFICI CON DATI DI TELEMETRIA RADAR ED INTERFEROMETRIA

La produzione di un DEM, con dati telerilevati da sensori attivi (SAR), mette in luce le tecniche applicate all'elaborazione di immagini complessse, rilevate durante missioni spaziali (ERS-1): telemetria radar, metodi interferometrici (coerenti) e ricostruzione di superfici.

1. Metodologie

Il radar ad apertura sintetica (SAR) è un sistema di acquisizione dei dati che utilizza segnali nel campo delle microonde, con frequenze comprese tra 1 e 10 Ghz. L'impiego di tale gamma di frequenze offre particolari vantaggi, come la capacità di operare con ogni condizione di tempo (dato che le microonde possono oltrepassare le nubi senza subire forti attenuazioni), di operare anche di notte e di penetrare parzialmente nella vegetazione e nel terreno secco.

Il SAR si differenzia in modo sostanziale dai sistemi radar convenzionali, avvicinandosi per le sue caratteristiche alle prestazioni offerte dai sistemi ottici. Infatti in un radar convenzionale la rilevazione dei bersagli è eseguita, trasmettendo un segnale impulsivo in una direzione definita dal lobo dell'antenna e misurando il ritardo dell'eco ricevuto.

La risoluzione in senso radiale necessaria è ottenuta trasmettendo impulsi molto stretti. La risoluzione nella direzione dell'azimut dipende invece dall'angolo di apertura dell'antenna, in generale, non molto piccolo. Considerando un angolo di apertura costante, la risoluzione in azimut decresce ovviamente con l'aumentare della distanza del punto riflettente. Pertanto quanto maggiore è la distanza, tanto più si rende necessario l'utilizzo di antenne fortemente direttive.

Di conseguenza, un sistema radar convenzionale offre buone prestazioni, solo se la distanza non è troppo elevata, se la lunghezza d'onda non è eccessiva e se l'antenna è sufficientemente grande. Se il sistema di rilevazione è montato su satellite e si vuole ottenere una mappa di una determinata zona della superficie terrestre con una buona risoluzione, la distanza assume un valore così elevato da rendere necessaria un'antenna di dimensioni spropositate.

La risoluzione richiesta può essere ottenuta con antenne di piccole dimensioni, grazie al concetto di apertura sintetica. Si sintetizza un'antenna molto grande utilizzando un vettore di piccole antenne affiancate. In realtà, nel caso del SAR non sono presenti tante antenne distinte, ma un unico sensore che cambia posizione ad ogni impulso trasmesso. Se nella zona osservata non esistono parti in movimento e se è possibile trascurare la velocità del sensore rispetto alla velocità della luce è ovvio che i due metodi sono identici.

Le immagini sono sintetizzate lungo la direzione radiale, considerando la forma ed il tempo di occorrenza degli impulsi rilevati; per comporre i dati lungo la direzione azimut sono invece utilizzate opportune tecniche di localizzazione.

Le tecniche di rilevazione altimetrica più diffuse ed anche le uniche realizzabili, utilizzando come postazione di rilevamento un satellite artificiale, sono basate sull'osservazione da due punti di vista differenti della stessa zona della superficie terrestre (cioè fanno uso della stereometria). Sfruttando metodi telemetrici incoerenti, la quota di un punto sulla superficie terrestre è calcolata in base alla diversa posizione del punto considerato nelle due immagini prodotte, oppure in base al diverso angolo, secondo cui i due sistemi orbitali vedono il punto sulla superficie terrestre.

Invece con metodi interferometrici (coerenti), l'elevazione del terreno è calcolata confrontando la fase delle immagini acquisite dai due diversi punti di osservazione. Allo scopo di poter sfruttare l'informazione

contenuta nella fase dell'immagine SAR, si considera un sistema composto da due radar ad apertura sintetica montati su piattaforme che si muovono lungo orbite parallele. In realtà, i dati sperimentali finora ottenuti sono stati rilevati dallo stesso satellite durante passaggi successivi su orbite differenti, perché una missione tandem, seppure prevista, non è ancora stata effettuata.

2. Procedure

Il radar ad apertura sintetica (SAR) presenta alcuni vantaggi rispetto ai sistemi tradizionali. Normalmente in questo tipo di sistemi è considerato solo il modulo delle immagini prodotte; tuttavia è possibile ricavare informazioni circa l'elevazione del terreno, confrontando la fase di due immagini, relative alla stessa zona della superficie terrestre, acquisite da punti di osservazione diversi. Infatti utilizzando l'informazione contenuta nella fase dei ritorni radar, si riescono ad ottenere risoluzioni altimetriche molto più elevate dei sistemi di telemetria classici.

Questo lavoro si occupa di osservazioni da satellite. Queste sono più adatte alla generazione dei dati interferometrici per altimetria, in quanto i satelliti non sono soggetti alle rapide variazioni di assetto, tipiche di altri sistemi SAR (ad esempio, quelli montati su aereo). In particolare, se le orbite dei due satelliti su cui sono montate le antenne radar sono parallele, si dimostra che le variazioni di quota del terreno dipendono linearmente dalle variazioni di fase dell'immagine interferometrica.

Un problema di notevole difficoltà nasce dal fatto che i valori di fase sono necessariamente limitati nell'intervallo $-\pi$, $+\pi$ (chiamati valori principali della fase) e, di conseguenza, la differenza tra i valori assunti in due punti adiacenti è nota con un'indeterminazione di 2π radianti. Questa indeterminazione può essere risolta, in modo semplice, solo se le variazioni di fase sono più piccole di π; in caso contrario si verificherà equivocazione (aliasing) della fase.

Un algoritmo per determinare i valori di fase effettivi, a partire dai valori principali (phase unwrapping o svolgimento della fase), anche in presenza di equivocazione, è stato l'oggetto essenziale di questo lavoro. A questo punto, il lavoro si è indirizzato in due direzioni distinte:

- dapprima si è cercato di sfruttare sia l'ampiezza sia la fase delle immagini SAR, al fine di identificare con precisione i punti in cui si presenta equivocazione della fase;
- successivamente si è cercato di ottenere soluzioni di qualità migliore, combinando più immagini SAR rilevate da punti di osservazione molto differenti.

Il rapporto segnale/rumore migliora notevolmente, se la distanza tra i due sistemi SAR aumenta. Tuttavia a questo vantaggio, si contrappone l'inconveniente considerevole che il procedimento di svolgimento della fase è molto più difficoltoso (o addirittura impossibile), a causa del maggior numero di zone in cui si ha equivocazione. Pertanto si è cercato di combinare le soluzioni ottenute da coppie di immagini riprese da SAR vicini (100 *m* circa) e lontani (1 *km* al massimo), per sfruttare le potenzialità offerte da questa tecnica. Tutte le tecniche mostrate in via teorica sono già state applicate sperimentalmente ad immagini rilevate durante missioni spaziali effettive.

Resta da osservare come l'uso di più frequenze ovvierebbe a questo inconveniente, ma gli attuali satelliti in orbita operano con una sola frequenza e, di conseguenza, il problema dello svolgimento della fase rimane.

3. Algoritmi

I prodotti, conseguenti agli obiettivi ed ai risultati attesi, sono costituiti da un sistema di programmi capaci di operare anche con dati SAR interferometrico. In quest'ambito, alcuni algoritmi di particolare interesse e rilevanza sono, di seguito, descritti.

Algoritmo di matching

1) Traslazione lungo il range e l'azimut di un'immagine, in quanto:

- le zone osservate dai due radar non sono esattamente coincidenti;
- la distanza fra due punti successivi, lungo la direzione del range, non è la stessa per entrambe le immagini, in ragione del fatto che i punti di osservazione non coincidono;
- le orbite del satellite non sono esattamente parallele.

2) Interpolazione dei dati, dato che gli spostamenti dei punti non hanno, in generale, un valore intero.

Algoritmo di unwrapping

1) Suddivisione dei punti singolari in clusters: tutti i punti appartenenti allo stesso cluster distano tra loro meno di un certo valore (in pratica, i punti di ogni cluster sono talmente ravvicinati che è impossibile stabilire quali siano gli estremi delle ghost lines reali).
2) Unione di tutti i punti appartenenti allo stesso cluster: dato che non è possibile stabilire quali punti singolari connettere, tutti i punti singolari dello stesso cluster sono uniti (pertanto l'unione tra due punti è condizionata al fatto che in essi la circuitazione abbia segno opposto e sia effettuata con segmenti di linea retta).
3) Unione tra punti appartenenti a cluster diversi: si cerca l'insieme di tagli con distanza minima in modo da collegare tutti i cluster (tuttavia i percorsi formati da questi tagli non devono generare cicli; in caso contrario, si creerebbero zone che non possono essere raggiunte dal percorso di integrazione della fase).

Algoritmo di georeferenziazione

1) Base di dati da compensare a minimi quadrati o con procedure robuste:

- osservazioni da immagini digitali dallo spazio;
- dati telerilevati da sensori attivi;
- coordinate di punti in sistemi di riferimento diversi;
- eventuali parametri ausiliari disponibili.

2) Sistema di riferimento (cioè datum geodetico) adottato per la compensazione delle equazioni di osservazione, pseudo-osservazione e vincolo:

- sistema di riferimento geodetico, adottando l'ellissoide come superficie di riferimento;
- correzione delle quote, per tener conto dell'ondulazione del geoide, passando così dalle quote ellissoidiche alle quote ortometriche.

Detto sistema di programmi risponde positivamente agli scopi richiesti.

4. Applicazioni

I risultati sperimentali, ottenuti elaborando due immagini ERS-1 su un seminato di circa un milione e mezzo di punti, relative ad una zona nel sud della Catalonia (Spagna) di oltre 100 km^2 e con un dislivello massimo di circa 230 *m*, sono piuttosto incoraggianti. Infatti nonostante si disponga di software di tipo sperimentale (non ingegnerizzato e non ottimizzato per una effettiva produzione operativa di modelli digitali) i tempi di elaborazione manuale sono piuttosto contenuti e si limitano essenzialmente alla preparazione dei dati (cambiamenti dei formati delle immagini, lettura dei dati ausiliari, trasformazioni dei sistemi di riferimento, ecc.) ed alla fase interattiva di controllo ed edizione manuale delle fasi srotolate, mentre tutta l'elaborazione è svolta in modo completamente automatico.

Per quanto concerne la precisione, confrontando le quote del DEM ottenuto da interferometria, dopo la proiezione cartografica, il calcolo delle quote ortometriche ed il ricampionamento al passo della griglia di riferimento (fig. 4.9), con quelle di un DTM proveniente da fotogrammetria convenzionale, per cartografia alla scala 1:5000, si è ottenuto un errore quadratico medio (RMS) di 9.2 m.

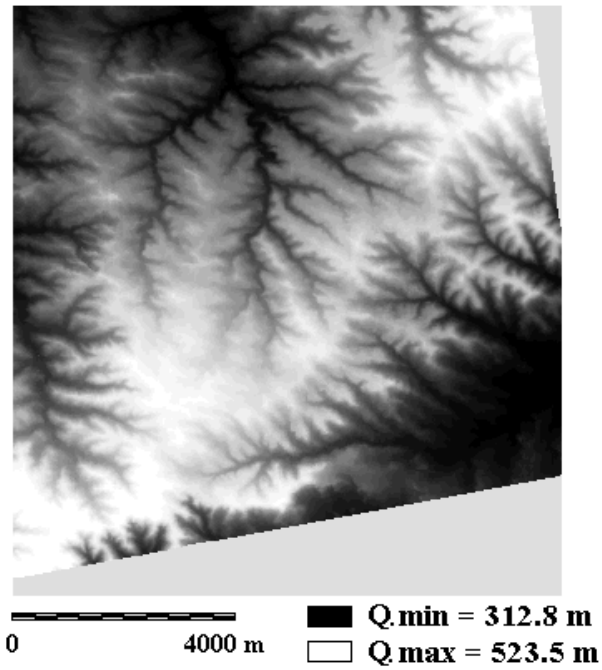


Fig. 4.9 – DEM generato mediante InSAR su zona nel sud della Catalonia (circa 100 km^2)

Le differenze, calcolate su tutta l'area rilevata, includono alcuni picchi (di circa 70 – 90 *m*, come mostrato in figura 4.10) molto localizzati e dovuti ad errori nello svolgimento delle fasi. La precisione è soddisfacente e comparabile con quella ottenuta con un DEM generato per la stessa area a partire da una coppia di immagini SPOT.

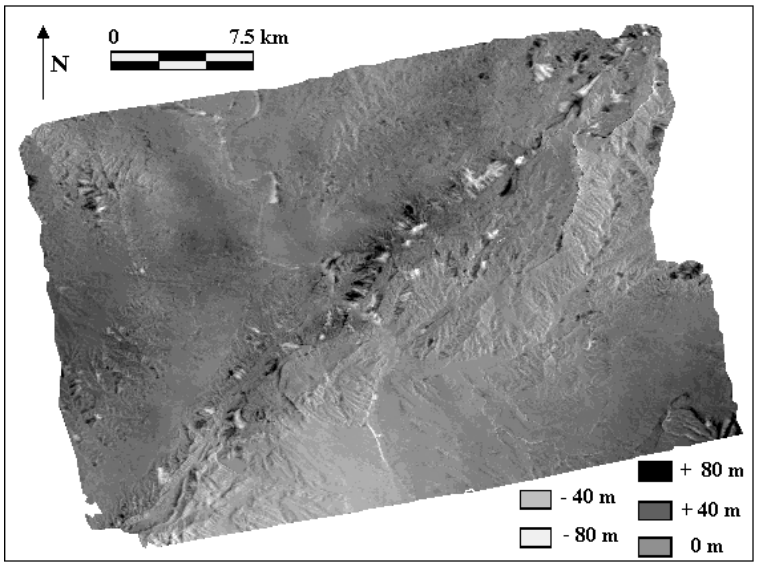


Fig. 4.10 – Differenze tra DEM di riferimento e DEM generato

Sebbene come già menzionato, i risultati ottenuti possano ritenersi incoraggianti, diversi aspetti della procedura deevono essere approfonditi e migliorati. In particolare, la procedura per lo srotolamento della fase interferometrica (fig. 4.11) può essere migliorata tenendo conto dell'informazione apportata dall'immagine di coerenza (fig. 4.12).

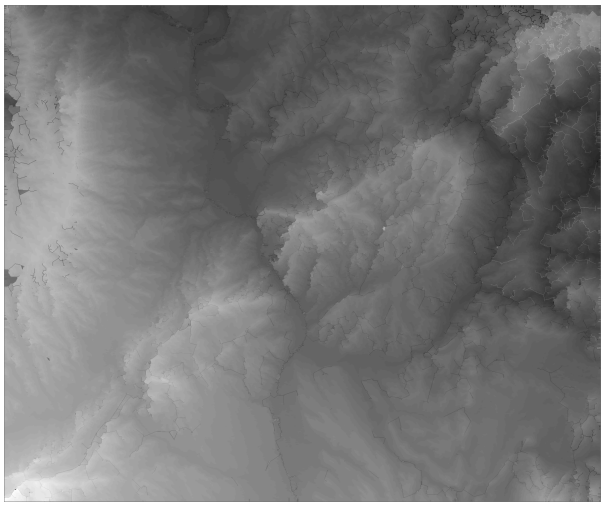


Fig. 4.11 – Fase interferometrica srotolata

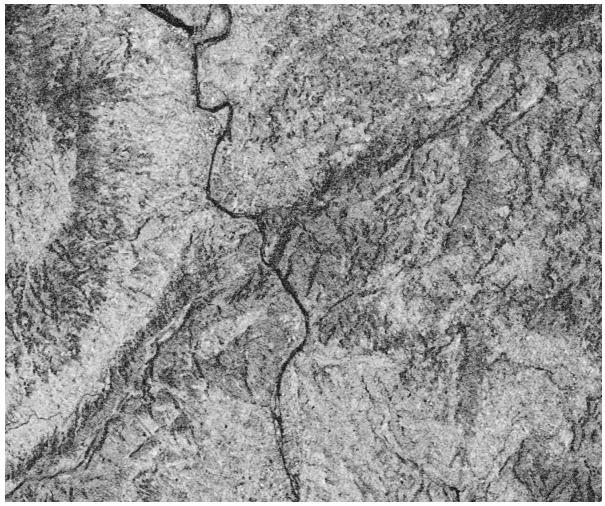


Fig. .4.12 – Immagine della coerenza

Inoltre sono ancora da indagare tutti i problemi concernenti il passaggio dalle fasi al seminato di punti, con particolare attenzione alla precisione richiesta nei dati ausiliari (ad es., le orbite, i punti di appoggio, ecc.) ed alla precisione del DEM che costituisce un parametro di fondamentale importanza, per valutare le potenzialità di impiego del prodotto finale per le diverse applicazioni.

In quest'ambito, ortoimmagini dell'ampiezza e della coerenza del segnale possono essere superimposte al DEM, producendo mappe o modelli 3D, a loro volta, completabili con la superimposizione di elementi vettoriali di interesse (parametrature, toponomastica, cartografia tecnica e/o tematica, ecc.), come insegnano le moderne tecniche della cartografia automatica, comunemente in uso.

D. RICOSTRUZIONE DI SUPERFICI CON DATI DA DIGITALIZZAZIONE CARTOGRAFICA

1. Modelli ottimali e tecniche avanzate per generare un DEM

Il principale obiettivo è stato di studiare la ricostruzione di superfici (ad es., generazione di DEM) e la validazione dei dati, perché le più importanti tecniche statistiche e numeriche sono comuni ad entrambi gli argomenti. L'idea base sottostante all'approccio usato è combinare, a livelli diversi le potenzialità di tecniche differenti. Inizialmente un'interpolazione polinomiale di basso grado serve a rimuovere una tendenza generale; successivamente il metodo degli elementi finiti è utilizzato per armonizzare le variazioni locali. Infine presupponendo che la topografia sia una realizzazione di un processo stocastico, si applica una tecnica sofisticata e raffinata di filtraggio basata su stime di covarianza e collocazione.

La generazione di DEM può essere svolta sia mediante analisi di immagini, sia mediante scansione automatica di immagini (o mappe) analogiche. Le carte forniscono informazioni utili per molti progetti (ad es., per la Valutazione dell'Impatto Ambientale). In questo esempio, la generazione di un DEM, come supporto di un GIS per la Valutazione dell'Impatto Ambientale, ha permesso la georeferenziazione di immagini rilevanti nello studio della distribuzione degli inquinanti. Si noti che il DEM non è facoltativo, ma necessario, perché le immagini sono 2D e gli oggetti (in questo esempio, il terreno) generalmente 3D.

2 Descrittori di forma

La cluster analysis, il riconoscimento di forme (mediante vettorizzazione, segmentazione e ricostruzione grafica) ed i descrittori di forma sono differenti modi per fare la compressione dei dati con una strategia ottimale ed ottenere, allo stesso tempo, un miglioramento sensibile nella qualità dei dati stessi. I descrittori di forma, in particolare, forniscono una rappresentazione analitica dei dati. Infatti differenti casi di strutture geometriche possono essere studiati mediante differenti descrittori di forma:

- una linea riferita ad un dominio monodimensionale, mediante interpolazione di linee;
- una superficie riferita ad un piano, mediante ricostruzione di superfici;
- il contorno di una figura riferita ad un'idonea approssimazione del contorno stesso (ad es., una poligonale fra i punti osservati costituenti la linea), usando due (o tre, se la curva è gobba) descrittori di forma ad una dimensione per le due (o tre) componenti del contorno;
- la superficie di un oggetto riferita a un'idonea approssimazione della superficie stessa (ad es., una triangolazione fra i punti osservati costituenti la superficie), usando tre descrittori di forma a due dimensioni per le tre componenti della superficie.

Molti approcci matematici sono utili per risolvere il problema posto; le splines polinomiali (costruite a partire dalle splines su base normalizzata) costituiscono uno dei metodi più interessanti. Il supporto locale delle B-splines è la loro maggiore proprietà. Inoltre le B-splines di un dato ordine sono continue ed anche un certo numero di loro derivate è continuo. Allora le B-splines di terzo ordine sono molto importanti, in quanto il loro supporto locale è abbastanza compatto ed esse sono continue con pendenza e con curvatura continue.

Infine si noti che le figure o gli oggetti delimitati, rispettivamente, da un contorno od una superficie sono in questo contesto semplicemente entità geometriche, indipendenti dalla loro natura fisica. Pertanto i loro contorni e superfici hanno solamente un significato geometrico e non differiscono dai due casi precedenti (ovvero una linea riferita ad un dominio monodimensionale ed una superficie riferita ad un piano).

3. Generazione del DEM

La zona, presa in considerazione per la generazione di un DEM finalizzato alla Valutazione dell'Impatto Ambientale, è stata l'area circostante la centrale Termoelettrica di Torre del Sale (Piombino). L' acquisizione dei dati in ingresso, circa 30.000 quote, è stata effettuata con la digitalizzazione di:

- una carta a scala 1:25.000 in un'area relativamente vasta, da curve di livello e punti isolati;
- una carta a scala 1:5.000 delle città di Follonica e Piombino, dove i simboli usati nella carta a scala più piccola annullano le informazioni altimetriche contenute, solamente da punti isolati.

La precisione dei dati in ingresso è di circa $5\,m$ in altezza. L'area, sita sulla costa del Mar Tirreno, è circa $30\times 20\,Km^2$ in ampiezza ed include zone piane ed accidentate, comprese tra 0 a $450\,m$ slm. La carta delle curve di livello dei dati originali è mostrata nella figura 4.13.

Data la topografia relativamente irregolare della zona, le curve di livello dei residui, dopo l'interpolazione polinomiale di basso ordine, hanno approssimativamente lo stesso andamento di quelle originarie (questo significa che la rimozione della tendenza non è particolarmente importante e può essere tralasciata. Al contrario, la lunghezza di correlazione è abbastanza grande e, per questa ragione, deve essere effettuato un prefiltraggio con il metodo degli elementi finiti. Di conseguenza, i dati sono stati interpolati mediante splines bicubiche (con nodi intervallati di $50\,m$), ottenendo residui intermedi.

Dopodiché ottenuta una ragionevole lunghezza di correlazione, si è proceduto al filtraggio mediante stima di covarianza e collocazione. La figura 4.14 mostra la rappresentazione 3D del terreno. Il rumore residuo è molto basso, con una dispersione pari a $2.8\,m$ (come può essere mostrato da una carta a curve di livello, comparata con i dati originali: fig. 4.15), e le funzioni di covarianza del rumore residuo assomigliano a quelle di un rumore bianco.

4. Validazione dei dati

Le basi di dati per la generazione di un DEM sono fornite da tecniche fotogrammetriche, di scansione delle immagini, oppure da digitalizzazione di carte. Nella scansioni fotogrammetriche, le altezze sono misurate lungo profili, dove l'intervallo spaziale e la densità delle misure possono essere adattate alla scabrosità del terreno, per ottenere (approssimativamente) una qualità uniforme. Invece la digitalizzazione di carte è solitamente eseguita mediante un campionamento delle curve di livello e, soprattutto nelle zone piane, usando punti isolati. Di conseguenza, i dati in ingresso per il processo di interpolazione non sono, in generale, regolarmente distribuiti (ma possono presentare classi vuote o sottocampionate: fig. 4.16) ed anche la forma planimetrica dell'area può essere qualsiasi. Inoltre se l'area non è semplicemente connessa (ad es., per l'inclusione di un lago), si trovano altre classi critiche. Ovviamente il numero di classi critiche dipende dalla spaziatura dei nodi: minore è l'ampiezza della spaziatura ($250\,m$ nell'esempio proposto), maggiore è il numero delle classi critiche, secondo una legge quadratica. Una strada possibile suggerisce:

- di usare inizialmente una spaziatura relativamente grande, per ottenere un'interpolazione approssimata e predire le altezze pseudo-osservate nelle classi critiche;
- di procedere successivamente con una spaziatura relativamente piccola, per avere un'interpolazione più accurata, tenendo conto delle pseudo-osservazioni precedentemente ottenute.

Purtroppo tuttavia questa procedura fallisce, quando si incontrano grandi linee di rottura. Infatti un grande gradiente in altezza, in un'area ristretta, non permette di seguire la forma della superficie interpolante. Allora l'intervallo di spaziatura dei nodi è stato ridotto a $50\,m$, cosicché la superficie interpolante possa seguire la topografia il più vicino possibile, e vincoli o pseudo-osservazioni di peso elevato, quali:

- l'altezza del punto più vicino;
- la media pesata delle altezze dei due punti più vicini;
- la media pesata delle altezze di alcuni punti circostanti;

dove i pesi sono proporzionali all'inverso delle distanze fra i nodi e ciascun punto, sono stati introdotti.
I primi due casi sono da preferirsi quando i nodi mal posti sono, rispettivamente:

- lungo linee di contorno di regioni vuote (costituite, ad es., dal mare o da laghi);
- in zone prossime alle linee di rottura,

mentre il terzo caso potrebbe essere applicato, quando le classi critiche si trovano casualmente distribuite. Si noti che la scelta delle prime due soluzioni abbisogna di un campionamento molto denso sul contorno o sulle linee di rottura, mentre condizioni errate imposte produrranno una ricostruzione non realistica della superficie. La figura 4.17 mostra un esempio dovuto alla mancanza di punti nel campionamento della linea di costa. Infine la valutazione del DEM, già fornita dall'entità del rumore, è stata più approfonditamente studiata, mediante una cross-validazine del modello, usando le altezze dei punti di controllo geodetici per effettuare una cross-validazione del modello. Di conseguenza i nuovi dati, costituiti dai punti di controllo geodetici, sono stati processati e le altezze predette sono state comparate con quelle osservate, ottenendo una serie di discrepanze con una dispersione nuovamente pari a $2.8\,m$.

5. Compressione dei dati mediante descrittori di forma

I descrittori di forma, applicati all'analisi della linea di costa e di alcune curve di livello, hanno avuto lo scopo di fornire una rappresentazione analitica di linee caratteristiche e provare la fattibilità della compressione dei dati (pari ad un terzo nell'esempio proposto: fig. 4.18), usando procedure proprie. Descrittori di forma, applicati all'intera superficie, avrebbero forse potuto mostrare livelli di compressione più elevati, ma corre il dovere di segnalare la ben maggiore complessità di un tale modo di procedere.
I risultati ottenuti sono incoraggianti dal punto di vista sia della compressione dei dati che della validità dell'interpretazione degli stessi. Le stesse procedure potrebbero essere raffinate in futuro, usando strategie adattative per la ricerca automatica di nodi in conformità alla densità dei dati ed al comportamento del fenomeno. A tale scopo, metodologie di campionamento progressivo e selettivo, tecniche multilivello, algoritmi sequenziali ed analisi della varianza offrono utili contributi.
Per quanto riguarda le tecniche applicate, approcci stocastici e/o misti (ad es., stime di covarianza e collocazione e/o combinazioni con altri metodi deterministici) potrebbero essere applicati in futuro. Tuttavia mentre questi metodi presentano dei vantaggi dal punto di vista della validità dell'interpretazione dei dati, essi risentono di alcuni svantaggi dal punto di vista della compressione degli stessi. Infatti gli approcci stocastici e misti offrono strategie più sofisticate e raffinate, per modellare il comportamento del fenomeno, ma la dimensione del segnale è uguale alla dimensione dei dati in ingresso.

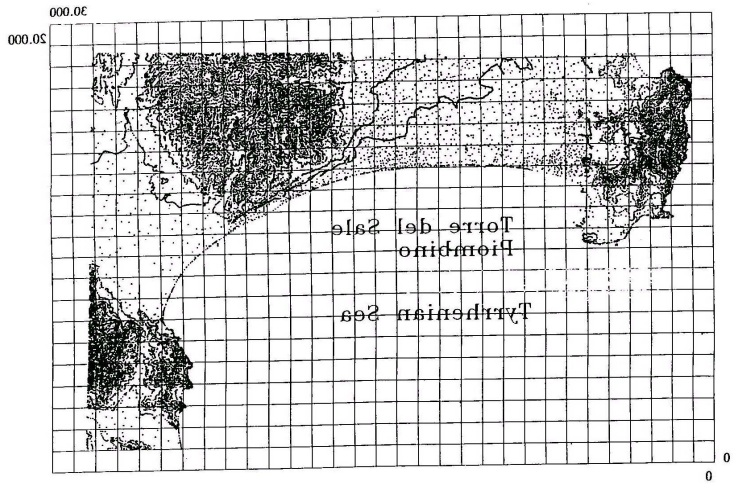


Fig. 4.13 – Area test

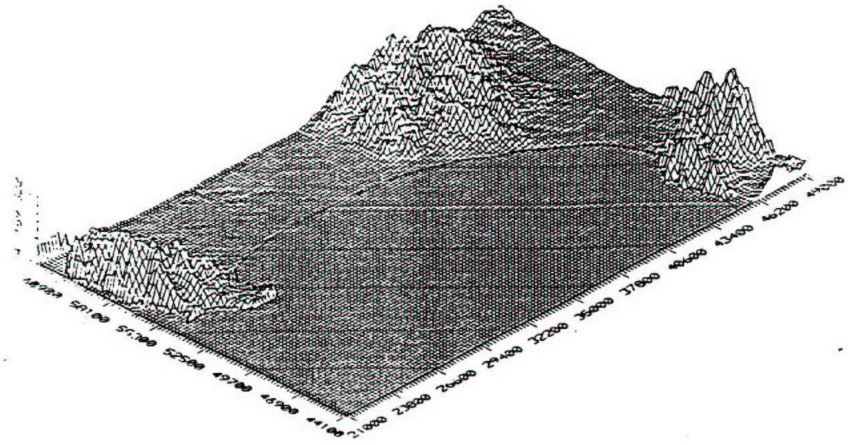


Fig. 4.14 – Rappresentazione 3D

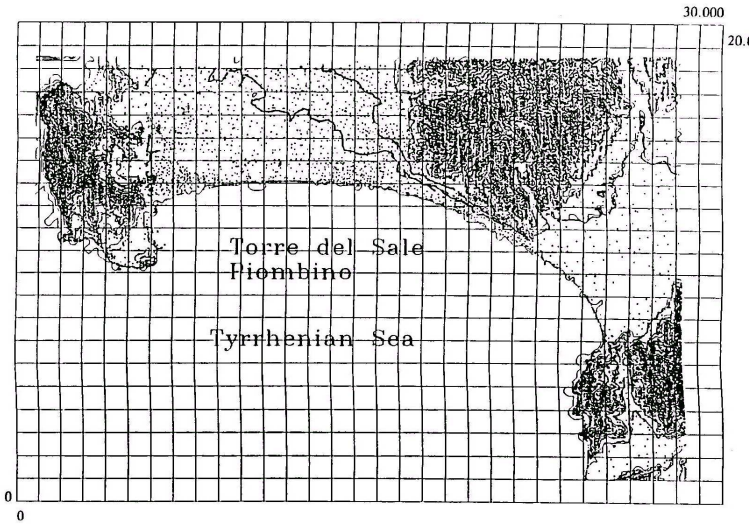


Fig. 4.15 – Rappresentazione a curve di livello confrontata con i dati in ingresso

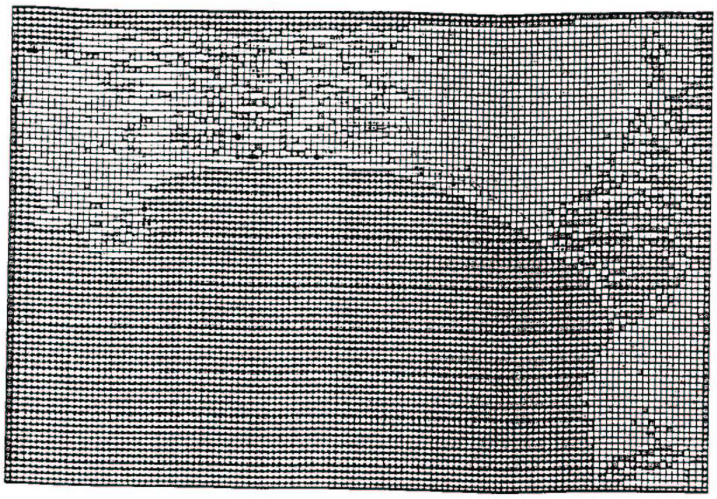


Fig. 4.16 – Classi vuote ed altezze predette errate (negative)

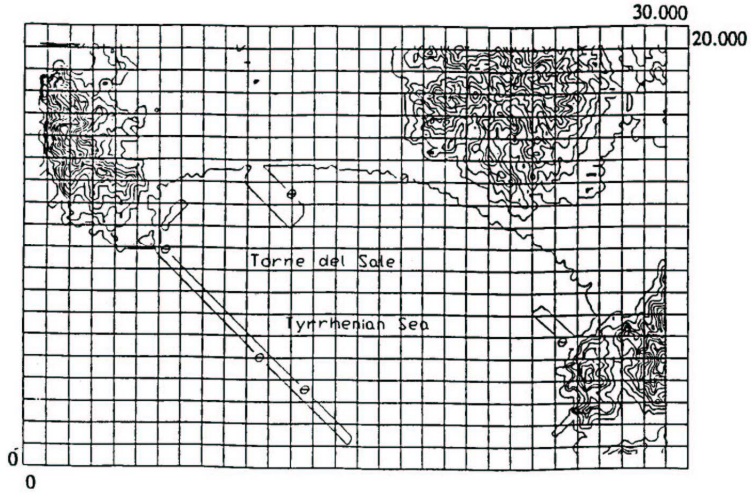


Fig. 4.17 – Rappresentazione a curve di livello con errori grossolani dovuti alla mancanza di punti nel campionamento della linea di costa

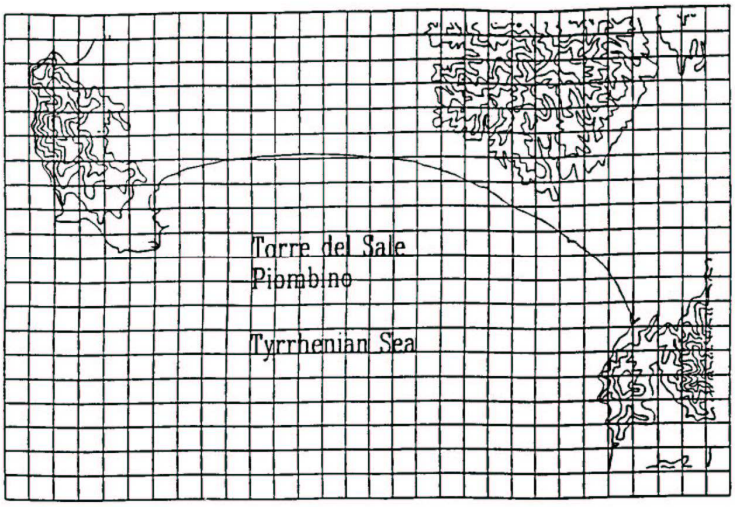


Fig. 4.18 – Descrittori di forma della linea di costa e delle curve di livello

PARTE V – TRASFORMAZIONI ELEMENTARI

Le trasformazioni elementari costituiscono quella parte, non particolarmente complessa, della modellazione di oggetti, capaci di far passare una figura od un oggetto da un sistema di coordinate ad un altro, con o senza deformazioni, comunque note a priori nei loro effetti. Fra queste, sono di particolare interesse le trasformazioni elementari piane e spaziali. Infatti ad una dimensione, le trasformazioni elementari consistono:

- in una traslazione (cioè nel calcolo di una media);
- con un eventuale cambio di scala (ovvero nel calcolo dei due coefficienti di una retta di regressione).

essendo solo tutto questo quanto rimane rispettivamente della congruenza e della trasformazione conforme (ovvero della trasformazione affine o dell'omografia, essendo così non distinguibili dalla precedente). A due dimensioni, è necessario scrivere i sistemi di equazioni di osservazione e risolverli con semplici passaggi, con l'intento di ricondurli a forma lineare, oppure quantomeno iterativamente lineare.

Data una trasformazione conforme (o una trasformazione S, ovvero di similitudine) nel piano:

$$X = \Delta X + x\lambda \cos\vartheta + y\lambda \sin\vartheta$$
$$Y = \Delta Y - x\lambda \sin\vartheta + y\lambda \cos\vartheta$$

il suo calcolo è immediato, avendo posto: $a = \lambda \cos\vartheta$, $b = \lambda \sin\vartheta$. Riscrivendo la trasformazione data:

$$X = \Delta X + xa + yb$$
$$Y = \Delta Y - xb + ya$$

si ottiene, per semplicità, avendo indicato ancora con ($x\ X\ y\ Y$) le coordinate ridotte ai baricentri:

$$a = \frac{E(xX) + E(yY)}{E(x^2) + E(y^2)} \qquad b = \frac{E(yX) - E(xY)}{E(x^2) + E(y^2)}$$

Di conseguenza, la rotazione e la variazione di scala sono:

$$\vartheta = arc\tan\frac{b}{a} \qquad \lambda = \sqrt{a^2 + b^2}$$

e le traslazioni:

$$\Delta X = \overline{X} - \overline{x}a - \overline{y}b \qquad \Delta Y = \overline{Y} + \overline{x}b - \overline{y}a$$

Anche il calcolo di una congruenza:

$$X = \Delta X + x\cos\vartheta + y\sin\vartheta$$
$$Y = \Delta Y - x\sin\vartheta + y\cos\vartheta$$

è immediato e coincide con quello di una trasformazione conforme, imponendo una scala unitaria (ovvero semplicemente trascurando il valore della variazione di scala). Questo procedimento è del tutto rigoroso, in quanto ogni trasformazione conforme è isogonica, in termini geometrici. Inoltre in termini statistici, la rotazione e la variazione di scala sono fra loro indipendenti, perché calcolate con una trasformazione ortogonale di varietà lineari indipendenti. Infatti essendo:

$$\sigma_a^2 = \sigma_b^2 = \frac{\sigma_0^2}{E(x^2) + E(y^2)}$$

ed essendo inoltre: $\sigma_{ab} = 0$, applicando la legge di propagazione della covarianza, si ha: $\sigma_{\lambda\vartheta} = 0$.

Data una <u>trasformazione affine</u>:

$$X = \Delta X + xa + yb$$
$$Y = \Delta Y + xc + yd$$

avendo indicato, per semplicità, ancora con ($x\ X\ y\ Y$) le coordinate ridotte ai baricentri, si ottiene:

$$a = \frac{E(y^2)E(xX) - E(xy)E(yX)}{E(x^2)E(y^2) - (E(xy))^2}$$
$$b = \frac{-E(xy)E(xX) + E(x^2)E(yX)}{E(x^2)E(y^2) - (E(xy))^2}$$
$$c = \frac{E(y^2)E(xY) - E(xy)E(yY)}{E(x^2)E(y^2) - (E(xy))^2}$$
$$d = \frac{-E(xy)E(xY) + E(x^2)E(yY)}{E(x^2)E(y^2) - (E(xy))^2}$$

Di conseguenza, avendo posto: $a = \lambda \cos\vartheta$, $b = \lambda \sin\vartheta$, $c = -\mu \sin(\vartheta + \gamma)$, $d = \mu \cos(\vartheta + \gamma)$, ovvero:

$$X = \Delta X + x\lambda \cos\vartheta + y\lambda \sin\vartheta$$
$$Y = \Delta Y - x\mu \sin(\vartheta + \gamma) + y\mu \cos(\vartheta + \gamma)$$

la rotazione, la torsione e le variazioni di scala sono:

$$\vartheta = arc\tan\frac{b}{a} \qquad \gamma = arc\tan\frac{-c}{d} - \vartheta \qquad \lambda = \sqrt{a^2 + b^2} \qquad \mu = \sqrt{c^2 + d^2}$$

e le traslazioni:

$$\Delta X = \overline{X} - \overline{x}a - \overline{y}b \qquad \Delta Y = \overline{Y} - \overline{x}c - \overline{y}d$$

Il calcolo di una <u>trasformazione omografica (od omografia)</u>:

$$X = \frac{ax + by + c}{gx + hy + 1} \qquad\qquad Y = \frac{dx + ey + f}{gx + hy + 1}$$

si ottiene, dapprima riducendo la trasformazione allo stesso denominatore, ovvero:

$$X = ax + by + c - gxX - hyX \qquad\qquad Y = ax + by + c - gxY - hyY$$

e successivamente costruendo la matrice normale ed il vettore termine noto normale[10].

$$C = \begin{bmatrix} E(x^2) & E(xy) & E(x) & 0 & 0 & 0 & -E(x^2X) & -E(xyX) \\ & E(y^2) & E(y) & 0 & 0 & 0 & -E(xyX) & -E(y^2X) \\ & & 1 & 0 & 0 & 0 & -E(xX) & -E(yX) \\ & & & E(x^2) & E(xy) & E(x) & -E(x^2Y) & -E(xyY) \\ & & & & E(y^2) & E(y) & -E(xyY) & -E(y^2Y) \\ & & & & & 1 & -E(xY) & -E(yY) \\ & sim. & & & & & E(x^2X^2)+E(y^2X^2) & E(xyX^2)+E(xyY^2) \\ & & & & & & & E(y^2X^2)+E(y^2Y^2) \end{bmatrix}$$

$$d^T = \begin{bmatrix} E(xX) & E(yX) & E(X) & E(xY) & E(yY) & E(Y) & -E(xX^2)-E(xY^2) & -E(yX^2)-E(yY^2) \end{bmatrix}$$

Ottenuti i parametri della trasformazione, la rotazione, la torsione e le variazioni di scala si ricavano come per la trasformazione affine, mentre le traslazioni sono rispettivamente: $\Delta X = c$, $\Delta Y = f$. Per quanto riguarda invece gli angoli di convergenza, se questi sono relativamente piccoli, valgono le seguenti approssimazioni: $\alpha = g$, $\beta = -h$; altrimenti gli stessi angoli hanno valore: $\beta = arc\,tan(-h)$, $\alpha = arc\,tan(g\,cos\,\beta)$.

Al termine di questa presentazione, corre il dovere di rilevare come, differenziando rispetto a tutti i parametri incogniti e procedendo iterativamente con i minimi quadrati non – lineari, si possa ottenere una soluzione differente, in quanto i pesi delle osservazioni, se unitari, sono diversamente assegnati. D'altra parte, anche per la semplice retta di regressione si è soliti assegnare pesi alle ordinate, considerando le ascisse parametri e, in moltissimi altri casi, il modello stocastico è comunque gestito in modo alquanto semplificato; pertanto anche accettare, come soluzione dell'omografia, quella ricondotta a forma lineare, non dovrebbe costituire una violazione troppo grave cui si può sempre rimediare come sopra specificato.

Un caso speciale, più complesso, è dato dalla <u>trasformazione affine particolare</u>:

$$X = \Delta X + x\lambda\,cos\,\vartheta + y\lambda\,sin\,\vartheta$$
$$Y = \Delta Y - x\mu\,sin\,\vartheta + y\mu\,cos\,\vartheta$$

dove saranno indicate, per semplicità, sempre con ($x\ X\ y\ Y$) anche le coordinate ridotte ai baricentri. Di conseguenza, avendo posto: $a = \lambda\,cos\,\vartheta$, $b = \lambda\,sin\,\vartheta$, $c = \mu\,cos\,\vartheta$, $d = \mu\,sin\,\vartheta$, ovvero:

[10] Infine la matrice normale, invertita e moltiplicata per il vettore termine noto normale, fornisce la soluzione cercata. Ovvi motivi di brevità impediscono di riportare formule chiuse dei suddetti passaggi di algebra matriciale.

$$X = xa + yb$$
$$Y = xc + yd$$

occorre aggiungere il vincolo: $ac + bd = 0$. Indicati con z il vettore dei parametri, R la seguente matrice:

$$R = \begin{bmatrix} x & y & 0 & 0 \\ 0 & 0 & x & y \end{bmatrix}$$

e s il vettore delle coordinate nell'altro sistema di riferimento, il sistema di equazioni di osservazione ed il vincolo possono essere così riscritti:

$$s - Rz = 0$$
$$z^T \Gamma z = 0$$

avendo introdotto, allo scopo, una matrice di doppio scambio, ovvero:

$$\Gamma = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Applicando il criterio dei minimi quadrati e differenziando rispetto al vettore dei parametri ed al moltiplicatore di Lagrange k (scritto per tenere debitamente conto del vincolo suddetto), si ottengono le equazioni normali:

$$g - (C + k\Gamma)z = 0$$
$$z^T \Gamma z = 0$$

avendo posto: $g = \sum_i R_i^T s_i$, $C = \sum_i R_i^T R_i$.

Queste equazioni sono poi risolte iterativamente; dalla prima equazione si ottiene:

$$z = C^{-1}(g - k\Gamma z)$$

che, sostituito nella seconda equazione normale, dà la seguente equazione di secondo grado:

$$(z^T \Gamma C^{-1} \Gamma C^{-1} \Gamma z)k^2 + 2(gC^{-1} \Gamma C^{-1} \Gamma z)k + gC^{-1} \Gamma C^{-1} g = 0$$

A sua volta, una delle radici di questa equazione serve per calcolare nuovamente z e così via, fino a convergere alla soluzione cercata. Esperienze numeriche hanno indicato come valore iniziale del ciclo iterativo: $k = 0$, quale valore prescelto per il proseguimento del ciclo iterativo: la più piccola delle radici dell'equazione di secondo grado (ovvero quella ottenuta sottraendo la radice quadrata del discriminante della suddetta equazione) e come numero di iterazioni necessarie: appena un ricalcolo dei parametri della trasformazione.

Al termine del ciclo iterativo, sostituite le espressioni iniziali nell'espressione della soluzione trovata ed eseguiti alcuni passaggi, si ottengono i parametri della trasformazione:

$$a = \frac{E(y^2)(E(xX) - kc) - E(xy)(E(yX) - kd)}{E(x^2)E(y^2) - (E(xy))^2} \qquad b = \frac{-E(xy)(E(xX) - kc) + E(x^2)(E(yX) - kd)}{E(x^2)E(y^2) - (E(xy))^2}$$

$$c = \frac{E(y^2)(E(xY) - ka) - E(xy)(E(yY) - kb)}{E(x^2)E(y^2) - (E(xy))^2} \qquad d = \frac{-E(xy)(E(xY) - ka) + E(x^2)(E(yY) - kb)}{E(x^2)E(y^2) - (E(xy))^2}$$

e inoltre: $k = \frac{V - \sqrt{V^2 - UW}}{U}$, avendo posto:

$$U = 2\left(E(xy)^2 + E(y^2)^2\right)ac - 2\left(E(x^2) + E(y^2)\right)E(xy)(ad + bc) + 2\left(E(x^2)^2 + E(xy)^2\right)bd$$

$$V = \left(E(xy)^2 + E(y^2)^2\right)(aE(xX) + cE(xY)) - \left(E(x^2) + E(y^2)\right)E(xy) \cdot$$
$$\cdot (bE(xX) + dE(xY) + aE(yX) + cE(yY)) + \left(E(x^2)^2 + E(xy)^2\right)(bE(yX) + dE(yY))$$

$$W = 2\left(E(xy)^2 + E(y^2)^2\right)E(xX)E(xY) - 2\left(E(x^2) + E(y^2)\right)E(xy)(E(xX)E(yY) + E(xY)E(yX)) +$$
$$+ 2\left(E(x^2)^2 + E(xy)^2\right)E(yX)E(yY)$$

Le espressioni della rotazione, delle variazioni di scala e delle traslazioni sono identiche a quelle della trasformazione affine. Inoltre nel caso, non infrequente, di una disposizione perfettamente simmetrica dei dati, è possibile calcolare una trasformazione affine a quattro parametri (oltre alle traslazioni) e ricavare da essi la rotazione e le variazioni di scala, imponendo una torsione nulla, perché in termini statistici, la rotazione, la torsione e le variazioni di scala risultano fra loro indipendenti, grazie alla simmetria dei dati.

Anche a tre dimensioni, è necessario scrivere i sistemi di equazioni di osservazione e risolverli con alcuni artifici, con l'intento di ricondurli a forma lineare, oppure quantomeno iterativamente lineare, anche se in questo caso i suddetti artifici sono talvolta più complessi e, in particolare, l'eventuale superamento del quarto grado nell'equazione risolvente dei suddetti sistemi mette in luce i limiti imposti dai teoremi di Abel e di Galois alla risoluzione delle equazioni algebriche.
Data l'equazione di una trasformazione conforme (o una trasformazione S, ovvero di similitudine) nello spazio 3D, dove R è una matrice di rotazione nello spazio, λ la variazione di scala e t il vettore delle traslazioni, sottraendo ad essa l'equazione media:

$$y = \lambda Rx + t$$
$$\underline{\bar{y} = \lambda R\bar{x} + t}$$
$$y - \bar{y} = \lambda R(x - \bar{x})$$

dalla quale, a posteriori, ricavare le traslazioni: $t = \bar{y} - \lambda R\bar{x}$, è possibile calcolare inizialmente la variazione di scala, grazie al prodotto scalare dell'equazione risultante per se stessa, sotto riportato:

$$y^T y = \lambda^2 x^T R^T R x = \lambda^2 x^T x \Rightarrow \lambda = \sqrt{\frac{y^T y}{x^T x}}$$

In questa espressione, per comodità di scrittura, si sono indicate ancora con ($x\ X\ y\ Y$) le coordinate ridotte ai baricentri e nelle seguenti saranno indicate sempre con ($x\ X\ y\ Y$) le stesse moltiplicate per un fattore di scala: $y = \lambda R x$.

Dopodiché per poter procedere al calcolo della rotazione nello spazio, occorre scegliere un sistema comodo di rappresentazione della stessa, come quella offerta dalla matrice di rotazione razionale di Rodriguez:

$$R = (I - S)^{-1}(I + S)$$

dove: I è una matrice identità e S una matrice emisimmetrica, così definita

$$S = \begin{vmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{vmatrix}$$

Allora il sistema di equazioni di osservazione diventa:

$$y = Rx = (I - S)^{-1}(I + S)x \ \Rightarrow \ (I - S)y = (I + S)x$$

ovvero in termini scalari:

$$\begin{bmatrix} 1 & -c & b \\ c & 1 & -a \\ -b & a & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & c & -b \\ -c & 1 & a \\ b & -a & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

e riorganizzando opportunamente matrici e vettori, così da raggruppare i parametri della trasformazione cercata:

$$\begin{bmatrix} 0 & Z-z & -(Y-y) \\ -(Z-z) & 0 & X-x \\ Y-y & -(X-x) & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} X-x \\ Y-y \\ Z-z \end{bmatrix} = 0$$

la cui soluzione è ovviamente immediata. Infine dati questi parametri e calcolata la matrice di rotazione razionale di Rodriguez:

$$R = \frac{1}{1 + a^2 + b^2 + c^2}\begin{bmatrix} 1 + a^2 - b^2 - c^2 & 2(ab - c) & 2(ac + b) \\ 2(ab + c) & 1 - a^2 + b^2 - c^2 & 2(bc - a) \\ 2(ac - b) & 2(bc + a) & 1 - a^2 - b^2 + c^2 \end{bmatrix}$$

è possibile procedere al calcolo tanto degli angoli di Cardano (ω rollio, φ beccheggio, k imbardata o rotazione propria), quanto di quelli di Eulero ($\vartheta = \lambda$ precessione o longitudine, $\zeta = \varphi$ nutazione o latitudine, α anomalia o azimut), esplicitando opportunamente le matrici di rotazione espresse in funzione dei suddetti angoli:

$$R = \begin{vmatrix} \cos k & \sin k & 0 \\ -\sin k & \cos k & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{vmatrix} \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{vmatrix} =$$

$$= \begin{vmatrix} \cos\varphi\cos k & \cos\omega\sin k + \sin\omega\sin\varphi\cos k & \sin\omega\sin k - \cos\omega\sin\varphi\cos k \\ -\cos\varphi\sin k & \cos\omega\cos k - \sin\omega\sin\varphi\sin k & \sin\omega\cos k + \cos\omega\sin\varphi\sin k \\ \sin\varphi & -\sin\omega\cos\varphi & \cos\omega\cos\varphi \end{vmatrix}$$

$$R = \begin{vmatrix} \cos k & \sin k & 0 \\ -\sin k & \cos k & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos\zeta & \sin\zeta \\ 0 & -\sin\zeta & \cos\zeta \end{vmatrix} \begin{vmatrix} \cos\vartheta & \sin\vartheta & 0 \\ -\sin\vartheta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} =$$

$$= \begin{vmatrix} \cos\vartheta\cos\alpha - \sin\vartheta\cos\zeta\sin\alpha & \sin\vartheta\cos\alpha + \cos\vartheta\cos\zeta\sin\alpha & \sin\alpha\sin\zeta \\ -\cos\vartheta\sin\alpha - \sin\vartheta\cos\zeta\cos\alpha & -\sin\vartheta\sin\alpha + \cos\vartheta\cos\zeta\cos\alpha & \cos\alpha\sin\zeta \\ \sin\vartheta\sin\zeta & -\cos\vartheta\sin\zeta & \cos\zeta \end{vmatrix}$$

Con specifico riferimento agli angoli di Cardano, l'esplicitazione dei loro valori si attua nel modo seguente:

- φ da r_{31} facendo: $arc\sin r_{31}$
- k da r_{11} e r_{21} facendo: $arc\tan(-r_{21}/r_{11})$
- ω da r_{32} e r_{33} facendo: $arc\tan(-r_{32}/r_{33})$

Tuttavia se φ è $\pm\pi/2$ ed il suo coseno nullo, si procede annullando κ, per definizione (data la coincidenza fra gli assi x e z ed il sommarsi delle due rotazioni), e ricavando ω da r_{22} e r_{23} , come $arc\tan(r_{23}/r_{22})$.

Analogamente per gli angoli di Eulero, l'esplicazione dei loro valori si attua nel modo seguente:

- ζ da r_{33} facendo: $arc\cos r_{33}$
- ϑ da r_{31} e $-r_{32}$ facendo: $arc\tan(r_{31}/-r_{32})$
- α da r_{13} e r_{23} facendo: $arc\tan(r_{13}/r_{23})$

Tuttavia se ζ è 0 oppure π ed il suo seno nullo, si procede annullando α, per definizione (come nel caso degli angoli di Cardano), e ricavando ϑ da r_{12} e r_{11}, come $arc\tan(r_{12}/r_{11})$.

Come a due dimensioni, anche il calcolo di una congruenza è immediato e coincide con quello di una trasformazione conforme, imponendo una scala unitaria (ovvero semplicemente trascurando il valore della variazione di scala). Questo procedimento è del tutto rigoroso, anche a tre dimensioni, in quanto ogni trasformazione conforme è isogonica, in termini geometrici.
Inoltre in termini statistici, le rotazioni e la variazione di scala sono fra loro indipendenti, perché calcolate con una trasformazione ortogonale di varietà lineari indipendenti. In entrambi i casi, potrebbe poi essere opportuno procedere iterativamente ad un affinamento della soluzione, sviluppando in serie di Taylor arrestata al primo ordine il sistema di equazioni di osservazione, utilizzando come valori approssimati dei parametri risultati ottenuti con i procedimenti esposti.

Analogamente, data una trasformazione affine:

$$X = \Delta X + xa + yb + zc$$
$$Y = \Delta Y + xd + ye + zf$$
$$Z = \Delta Z + xg + yh + zl$$

avendo indicato, per semplicità, ancora con ($x\ X\ y\ Y$) le coordinate ridotte ai baricentri, si ottiene:

$$a = \frac{AE(xX) - BE(yX) + CE(zX)}{H} \qquad b = \frac{-BE(xX) + DE(yX) - FE(zX)}{H}$$

$$c = \frac{CE(xX) - FE(yX) + GE(zX)}{H}$$

$$d = \frac{AE(xY) - BE(yY) + CE(zY)}{H} \qquad e = \frac{-BE(xY) + DE(yY) - FE(zY)}{H}$$

$$f = \frac{CE(xY) - FE(yY) + GE(zY)}{H}$$

$$g = \frac{AE(xZ) - BE(yZ) + CE(zZ)}{H} \qquad h = \frac{-BE(xZ) + DE(yZ) - FE(zZ)}{H}$$

$$k = \frac{CE(xZ) - FE(yZ) + GE(zZ)}{H}$$

dove i fattori ($A\,B\,C\,D\,F\,G$) sono dati dai minori complementari:

$$A = E(y^2)E(z^2) - (E(yz))^2 \qquad B = E(xy)E(z^2) - E(zx)E(yz)$$
$$D = E(x^2)E(z^2) - (E(zx))^2 \qquad C = E(xy)E(yz) - E(zx)E(y^2)$$
$$G = E(x^2)E(y^2) - (E(xy))^2 \qquad F = E(x^2)E(yz) - E(zx)E(xy)$$

ed il denominatore H è dato dal determinante:

$$H = E(x^2)E(y^2)E(z^2) + 2E(xy)E(yz)E(zx) - E(x^2)(E(yz))^2 - E(y^2)(E(zx))^2 - E(z^2)(E(xy))^2$$

Ottenuti i nove parametri della trasformazione, oltre alle traslazioni si ottengono dalle equazioni medie, a posteriori, le tre variazioni di scala sono ricavate dalla radice quadrata della somma quadratica di tre parametri alla volta (una tripletta per ogni equazione):

$$\Delta X = \overline{X} - \overline{x}a - \overline{y}b - \overline{z}c \qquad \lambda = \sqrt{a^2 + b^2 + c^2}$$
$$\Delta Y = \overline{Y} - \overline{x}d - \overline{y}e - \overline{z}f \qquad \mu = \sqrt{d^2 + e^2 + f^2}$$
$$\Delta Z = \overline{Z} - \overline{x}g - \overline{y}h - \overline{z}l \qquad \nu = \sqrt{g^2 + h^2 + l^2}$$

Per quanto riguarda invece, le tre rotazioni e le tre torsioni, riscritta la matrice di rotazione / torsione R per gli angoli di Cardano ($\omega\ \varphi\ k$), evidenziando in essa le torsioni ($\gamma\ \delta\ \varepsilon$), e dopo aver diviso tutti i suoi parametri per le scale corrispondenti, si ottiene:

$$\begin{vmatrix} \cos(\varphi+\delta)\cos k & \cos\omega\ \sin k + \sin\omega\ \sin(\varphi+\delta)\cos k & \sin\omega\ \sin k - \cos\omega\ \sin(\varphi+\delta)\cos k \\ -\cos\varphi\ \sin(k+\gamma) & \cos\omega\cos(k+\gamma) - \sin\omega\ \sin\varphi\ \sin(k+\gamma) & \sin\omega\cos(k+\gamma) + \cos\omega\ \sin\varphi\ \sin(k+\gamma) \\ \sin\varphi & -\sin(\omega+\varepsilon)\cos\varphi & \cos(\omega+\varepsilon)\cos\varphi \end{vmatrix}$$

Dopodiché l'esplicitazione delle tre rotazioni e delle tre torsioni procede nel seguente modo, per brevità, avendo chiamato ($A\ B\ C\ D\ E\ F\ G\ H\ L$) i nove coefficienti della suddetta matrice:

$$\varphi = arc\ sin\ G \qquad \omega + \varepsilon = arc\ tan\frac{-H}{L} \qquad k + \gamma = arcsin\frac{-D}{\cos\varphi}$$

Inoltre dapprima da ($E\ F$) si ricava ω e successivamente da ($B\ C$) si ricava k ; infatti sottraendo membro a membro, e risolvendo le espressioni ottenute, si ha:

$$\omega = arc\ sin\frac{F\cos(k+\gamma) - E\sin\varphi\ \sin(k+\gamma)}{\cos^2(k+\gamma) + \sin^2\varphi\ \sin^2(k+\gamma)} = arc\ cos\frac{E\cos(k+\gamma) + F\sin\varphi\ \sin(k+\gamma)}{\cos^2(k+\gamma) + \sin^2\varphi\ \sin^2(k+\gamma)}$$

$$k = arc\ sin(B\cos\omega + C\sin\omega) \qquad \text{infine:} \qquad \varphi + \delta = arccos\frac{A}{\cos k}$$

Si noti come sia necessario avere cura di verificare sempre i quadranti di appartenenza degli archi calcolati, anche al fine di ricavare correttamente da queste espressioni angolari, per differenza, le torsioni ($\gamma\ \delta\ \varepsilon$).

Il calcolo di una <u>trasformazione omografica (od omografia)</u>:

$$X = \frac{ax + by + cz + d}{px + qy + rz + 1} \qquad Y = \frac{ex + fy + gz + h}{px + qy + rz + 1} \qquad Z = \frac{kx + ly + mz + n}{px + qy + rz + 1}$$

si ottiene, dapprima riducendo la trasformazione allo stesso denominatore, ovvero:

$$X = ax + by + cz + d - pxX - qyX - rzX \qquad Y = ex + fy + gz + h - pxY - qyY - rzY$$

$$Z = kx + ly + mz + n - pxZ - qyZ - rzZ$$

e successivamente costruendo la matrice normale ed il vettore termine noto normale[11].

$$C = \begin{bmatrix} T & 0 & 0 & U \\ & T & 0 & V \\ & & T & W \\ sim. & & & S \end{bmatrix}$$ composta come di seguito definita (essendo l'ultima matrice la matrice S):

$$T = \begin{bmatrix} E(x^2) & E(xy) & E(xz) & E(x) \\ & E(y^2) & E(yz) & E(y) \\ & & E(z^2) & E(z) \\ & & & 1 \end{bmatrix} \qquad U = \begin{bmatrix} -E(x^2X) & -E(xyX) & -E(xzX) \\ -E(xyX) & -E(y^2X) & -E(yzX) \\ -E(xzX) & -E(yzX) & -E(z^2X) \\ -E(xX) & E(yX) & E(zX) \end{bmatrix}$$

$$V = \begin{bmatrix} -E(x^2Y) & -E(xyY) & -E(xzY) \\ -E(xyY) & -E(y^2Y) & -E(yzY) \\ -E(xzY) & -E(z^2Y) & -E(z^2Y) \\ -E(xY) & E(yY) & E(zY) \end{bmatrix} \qquad W = \begin{bmatrix} -E(x^2Z) & -E(xyZ) & -E(xzZ) \\ -E(xyZ) & -E(y^2Z) & -E(yzZ) \\ -E(xzZ) & -E(yzZ) & -E(z^2Z) \\ -E(xZ) & E(yZ) & E(zZ) \end{bmatrix}$$

$$\begin{bmatrix} E(x^2X^2)+E(y^2X^2)+E(z^2X^2) & E(xyX^2)+E(xyY^2)+E(xyZ^2) & E(xzX^2)+E(xzY^2)+E(xzZ^2) \\ & E(x^2Y^2)+E(y^2Y^2)+E(z^2Y^2) & E(yzX^2)+E(yzY^2)+E(yzZ^2) \\ & & E(x^2Z^2)+E(y^2Z^2)+E(z^2Z^2) \end{bmatrix}$$

$d^T = \begin{bmatrix} u^T & v^T & w^T & s^T \end{bmatrix}$ composto come di seguito definito (essendo l'ultimo vettore il vettore s^T):

$$u^T = [E(xX) \quad E(yX) \quad E(zX) \quad E(X)] \qquad v^T = [E(xY) \quad E(yY) \quad E(zY) \quad E(Y)]$$

$$w^T = [E(xZ) \quad E(yZ) \quad E(zZ) \quad E(Z)]$$

$$\begin{bmatrix} -E(xX^2)-E(yX^2)-E(zX^2) & -E(xY^2)-E(yY^2)-E(zY^2) & -E(xZ^2)-E(yZ^2)-E(zZ^2) \end{bmatrix}$$

Ottenuti i parametri della trasformazione, la rotazione, la torsione e le variazioni di scala si ricavano come per la trasformazione affine, mentre le traslazioni sono rispettivamente: $\Delta X = d$, $\Delta Y = h$, $\Delta Z = n$.

Per quanto riguarda invece i tre angoloidi (o angoli solidi) di convergenza, gli stessi sono funzioni abbastanza complesse di tutti i parametri dell'omografia. L'algoritmo di Oosterom e Strackee calcola l'angolo solido, sotteso ad un triangolo di vertici $R_1\ R_2\ R_3$ visto dall'origine (che forma una piramide con la stessa origine):

11 Infine la matrice normale, invertita e moltiplicata per il vettore termine noto normale, fornisce la soluzione cercata. Ovvi motivi di brevità impediscono di riportare formule chiuse dei suddetti passaggi di algebra matriciale. D'altra parte, i suddetti calcoli di algebra matriciale sono proprio standard e non richiedono ulteriori competenze specifiche. A riguardo, si veda anche: G.B., Mussio L. (1984): Algoritmi esatti per la risoluzione di grandi sistemi. In: M. Cunietti (Ed) *Ricerche di Geodesia, Topografia e Fotogrammetria*, n. 4. CLUP, Milano, p. 5–46, a sua volta, ripreso nell'articolo: Crespi M., Mussio L. (1989): Problemi di calcolo. In: F. Crosilla, L. Mussio (Ed's) *Progettazione ed ottimizzazione del rilievo topografico e fotogrammetrico di controllo*. CISM, Udine, p. 255–282.

$$\Omega = 2\,arc\tan\frac{det(R_1R_2R_3)}{R_1R_2R_3 + (R_1 \cdot R_2)R_3 + (R_2 \cdot R_3)R_1 + (R_3 \cdot R_1)R_2}$$

Allora calcolate tre rette convergenti in un vertice (origine) per tre coppie di punti, i tre vettori, i loro moduli e le loro componenti, è possibile procedere all'applicazione di questa espressione al caso specifico [12].
Un caso speciale, più complesso, è dato dalla trasformazione affine particolare:

$$\begin{aligned} X &= \Delta X + x\lambda\cos\varphi\cos k + y\lambda(\cos\omega\sin k + \sin\omega\sin\varphi\cos k) + z\lambda(\sin\omega\sin k - \cos\omega\sin\varphi\cos k) \\ Y &= \Delta Y - x\mu\cos\varphi\sin k + y\mu(\cos\omega\cos k - \sin\omega\sin\varphi\sin k) + z\mu(\sin\omega\cos k + \cos\omega\sin\varphi\sin k) \\ Z &= \Delta Z + x\nu\sin\varphi - y\nu\sin\omega\cos\varphi + z\nu\cos\omega\cos\varphi \end{aligned}$$

dove saranno indicate, per semplicità, sempre con ($x\ X\ y\ Y$) anche le coordinate ridotte ai baricentri. Di conseguenza, avendo posto:

$$\begin{aligned} X &= xa + yb + zc \\ Y &= xd + ye + zf \\ Z &= xg + yh + zl \end{aligned}$$

occorre aggiungere i tre vincoli: $ad + be + cf = 0$, $ag + bh + cl = 0$, $dg + eh + fl = 0$. Indicati con z il vettore dei parametri, R la seguente matrice:

$$R = \begin{bmatrix} x & y & z & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x & y & z \end{bmatrix}$$

e s il vettore delle coordinate nell'altro sistema di riferimento, il sistema di equazioni di osservazione ed i tre vincoli possono essere così riscritti:

$$s - Rz = 0 \qquad z^T\Gamma_1\, z = 0 \qquad z^T\Gamma_2\, z = 0 \qquad z^T\Gamma_3\, z = 0$$

avendo introdotto, allo scopo, tre matrici di scambio (fra tre elementi per volta):

$$\Gamma_1 = \begin{bmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Gamma_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{bmatrix} \qquad \Gamma_3 = \begin{bmatrix} 0 & 0 & I \\ 0 & 0 & 0 \\ I & 0 & 0 \end{bmatrix}$$

Applicando il criterio dei minimi quadrati, differenziando rispetto al vettore dei parametri ed ai moltiplicatori di Lagrange ($k_1\ k_2\ k_3$), scritti per tenere debitamente conto dei vincoli suddetti, ed avendo posto: $g = \sum_i R_i^T s_i$, $C = \sum_i R_i^T R_i$. si ottengono le equazioni normali:

[12] Anche al termine di questa presentazione, corre ancora il dovere di rilevare come, differenziando rispetto a tutti i parametri incogniti e procedendo iterativamente con i minimi quadrati non – lineari, si ottenga sempre una soluzione differente, in quanto i pesi delle osservazioni, se unitari come usuale, sono diversamente assegnati.

$$g - (C + k_1\Gamma_1 + k_2\Gamma_2 + k_3\Gamma_3)z = 0$$

$$z^T\Gamma_1\, z = 0 \qquad z^T\Gamma_2\, z = 0 \qquad z^T\Gamma_3\, z = 0$$

Queste equazioni sono poi risolte iterativamente; dalla prima equazione si ottiene:

$$z = C^{-1}\left(g - (k_1\Gamma_1 + k_2\Gamma_2 + k_3\Gamma_3)z\right)$$

che, sostituito nelle altre equazioni normali, dà un sistema di equazioni di secondo grado in ($k_1\ k_2\ k_3$). A sua volta, una delle radici di questo sistema serve per calcolare nuovamente z e così via, fino a convergere alla soluzione cercata. Purtroppo un sistema di tre equazioni complete di secondo grado dà un'equazione risolvente di sesto grado, cosa che impedisce, in generale, il calcolo di una soluzione esatta. Pertanto la soluzione del suddetto sistema, ove non – razionale, dovrà essere cercata solo mediante tecniche di soluzione approssimata. A tale proposito, si può procedere:

- alla ricerca di approssimazioni razionali delle radici con la formula di Ruffini, minimizzando l'errore residuo;
- ad operazioni con le basi di Groebner, un metodo più moderno, capace di trovare anche le radici non – razionali, a partire da equazioni a coefficienti razionali (cosa in generale vera, quando si opera in ambito numerico; altrimenti occorre approssimare opportunamente i coefficienti non – razionali delle equazioni).

Tuttavia essendo tutti questi i problemi argomenti precipui di calcolo numerico, tutto quanto riguarda la soluzione di sistemi di equazioni non – lineari è considerato estraneo agli scopi del presente lavoro.
Dopodichè le espressioni della rotazione, delle variazioni di scala e delle traslazioni sono identiche a quelle della trasformazione affine. Inoltre nel caso, non infrequente, di una disposizione perfettamente simmetrica dei dati, è possibile calcolare una trasformazione affine a nove parametri (oltre alle traslazioni) e ricavare da essi la rotazione e le variazioni di scala, imponendo una torsione nulla, perché in termini statistici, la rotazione, la torsione e le variazioni di scala risultano fra loro indipendenti, grazie alla simmetria dei dati.

APPENDICE B – OSSERVAZIONI CON DIFETTO DI RANGO

Un aspetto originale ed innovativo riguarda un problema, solo apparentemente, collaterale: l'acquisizione d'informazioni da dati con difetto di rango. Infatti se i dati rilevati sono distanze o punti immagine, mentre ovviamente le informazioni richieste sono le coordinate 3D dei punti oggetto, allora l'informazione acquisita presenta un difetto di rango, rispettivamente pari a due od uno che deve essere sanata attraverso opportune condizioni geometriche di regolarizzazione.
Questo fatto è davvero concreto, in quanto le scansioni (laser, radar e sonar), misurano distanze, in forma di nuvole di punti, senza una precisa indicazione sulla direzione (spaziale) d'osservazione, mentre i punti immagine sono proiettati in uno spazio bidimensionale.
In entrambi i casi, la continuità locale della superficie dell'oggetto e, nel primo caso, anche la continuità locale della linea di scansione (delle distanze) permettono di superare l'impasse nel quale si cadrebbe altrimenti.

Le condizioni geometriche di regolarizzazione:

Una prima condizione geometrica di regolarizzazione, abbastanza semplice, impone medie mobili locali, per garantire la ragionevole continuità della superficie. Infatti l'usuale espressione della media aritmetica è calcolata, per ogni punto dell'insieme dato, su tutti i punti appartenenti ad un opportuno intono del punto, di volta in volta, in esame.

A riguardo, occorre notare come un intorno troppo piccolo (a rischio di essere quasi vuoto) sia troppo sensibile alla casuale variabilità dei singoli dati, mentre un intorno troppo grande (pressoché coincidente con la totalità dei dati stessi) finisca con ridurre le medie mobili quasi alla sola media generale.

Una seconda condizione geometrica di regolarizzazione, ben più complessa, serve a garantire la continuità locale della linea di scansione (delle distanze). La ragione di questa seconda continuità è insita nelle modalità stesse della scansione effettuata dal sensore nel suo movimento (rotatorio) di rilevamento dei punti oggetto.

- Allora se la superficie può essere considerata localmente piana, allora il cono retto delle distanze la interseca con qualche retta e molte coniche diverse tra loro, al variare dell'inclinazione del cono rispetto alla superficie piana.
- Se invece la superficie ha una forma localmente più complessa, pur fermandosi alle quadriche, le intersezioni sono curve, piane o gobbe, di terzo o quarto grado di cui non è ancora nota una classificazione completa.

Come evidente, questa ultima questione, riguardante essenzialmente problemi di geometria analitica, esula dagli scopi del presente lavoro e merita uno studio apposito.

La fig. B.1 mostra le coniche, già studiate nell'antichità da Apollonio di Perga, ottenute come sezioni di un cono retto, essendo le stesse: una ellisse (od una circonferenza, come caso particolare), una parabola ed una iperbole (od una iperbole equilatera, come caso particolare, se il cono ha un angoloide retto). E' noto che due rette coincidenti, parallele od intersecantesi (come gli asintoti di una iperbole) sono coniche degeneri, comunque estranee ad i casi in esame.

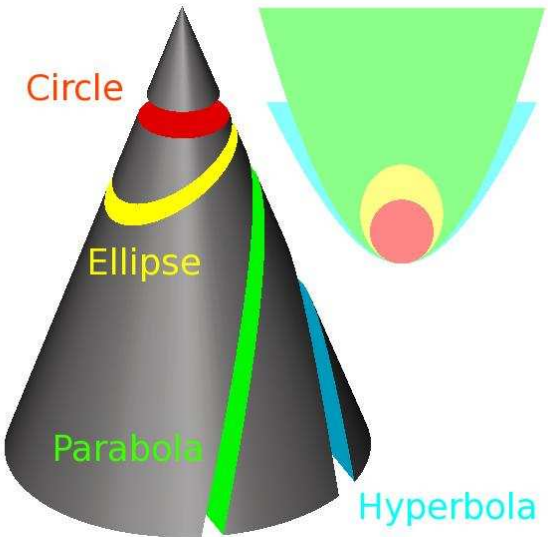


Fig. B.1 – L e coniche come intersezione di un piano con un cono retto

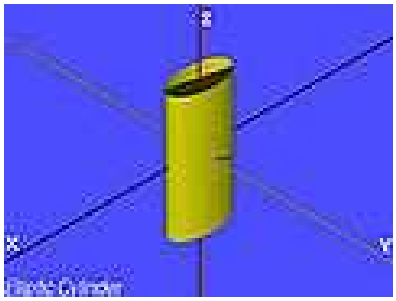
Per ragioni di semplicità, la condizione d'appartenenza dei punti immagine alla linea di scansione è ottenuta imponendo che la differenza seconda finita sia conforme a quella di una conica prescelta (nella maggior parte dei casi: una iperbole [13]):

$$\Delta^2 x = (x_1 - 2x_2 + x_3) / \Delta^2$$

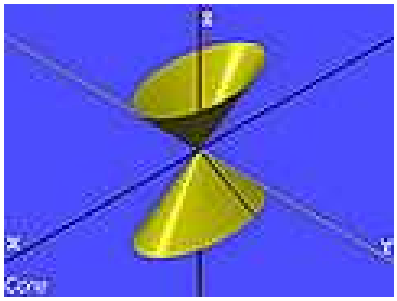
Come già detto in precedenza, prendere in considerazione una superficie più complessa esula dagli scopi del presente lavoro; pertanto quanto segue è solo una digressione sulle quadriche ed alcune loro combinazioni, volta a dare qualche indicazione a riguardo, gettando uno sguardo anche sul mondo affascinante delle simmetrie.
Le quadriche sono, come noto, superfici di secondo grado [14], con l'eccezione di quelle degeneri (date da piani coincidenti, piani paralleli e piani intersecantisi, comunque estranei ad i casi in esame), in numero di sette [15]:

1. cilindri;
2. coni;
3. ellissoidi;
4. paraboloidi ellittici;
5. paraboloidi iperbolici;
6. iperboloidi ad una falda;
7. iperboloidi a due falde.

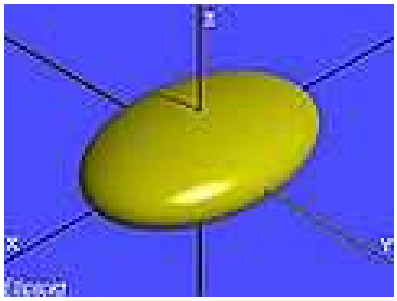
La fig. B.2 illustra le sette quadriche non-degeneri.



Cilindro



Cono

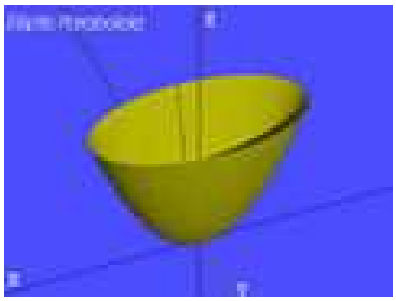


Ellissoide

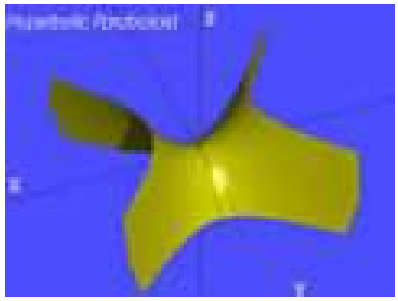
[13] La ragione di questa affermazione sta nel parallelismo tra l'asse verticale del sensore e la verticalità del piano contenente i punti oggetto. Una ortogonalità genera circonferenze, ma richiede di ruotare il sensore (nel caso di un piano verticale) o di avere un sensore capace di esplorare anche viste molto alte o molto basse (nel caso di un piano verticale). Ovviamente non parallelismi e non ortogonalità determinano ellissi e parabole, ma sono casi più rari, in questo contesto specifico.

[14] Alcune di queste superfici sono euclidee (cilindri e coni), cioè a semplice curvatura, mentre tutte le altre sono rimaniane, perché a doppia curvatura. A loro volta, le superficie rimaniane si suddividono in ellittiche ed iperboliche, rispettivamente secondo la non-esistenza di una parallela, oppure l'esistenza di infinite parallele. Entrambi i casi costituiscono infatti una violazione del quinto postulato di Euclide che impone l'esistenza di una sola parallela, ad una retta, per un punto ad essa esterno.

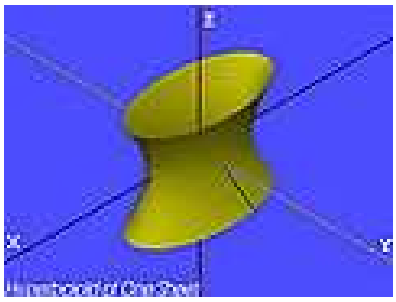
[15] Una precisazione è doverosa riguardo le classificazioni ed i loro numeri. Ogni classificazione è solo un tentativo d'interpretazione e, di conseguenza, abbastanza arbitraria (ad esempio, non sono presi in considerazione casi inclinati: rari, ma non impossibili). E' comunque significativo rilevare, come qualche classificazione porti a conteggi già noti in altri contesti (come ad esempio, nelle dimensioni dei gruppi di simmetrie lineari, planari e spaziali, caratterizzati rispettivamente da numeri sette, diciassette e trentadue). Escludendo tentazioni cabalistiche, occorre fare riferimento a forme biologiche (ad esempio, destra e sinistra o le dita della mano) e/o naturali che influenzano il modo di pensare.



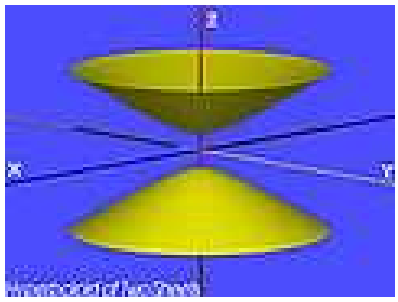
Paraboloide ellittico



Paraboloide iperbolico



Iperboloide ad una falda



Iperboloide a due falde

Fig. B.2 – Le quadriche non-degeneri

L'intersezione di un cono retto con le sette quadriche, aggiunte al piano, già preso in considerazione (che notoriamente individua, quali intersezioni, rette, ellissi o circonferenze, parabole ed iperboli o iperboli equilatere), determina i seguenti diciassette casi (relativamente ai quali devono essere studiate le varie linee d'intersezione):

1. piano parallelo [16];
2. piano perpendicolare;
3. cilindro ellittico parallelo;
4. cilindro ellittico perpendicolare;
5. cilindro parabolico parallelo;
6. cilindro parabolico perpendicolare;
7. cilindro iperbolico parallelo;
8. cilindro iperbolico perpendicolare;
9. cono parallelo;
10. cono perpendicolare;
11. ellissoide;
12. paraboloide ellittico;
13. parabolide iperbolico parallelo;
14. parabolide iperbolico perpendicolare;
15. iperboloide ad una falda parallelo;
16. iperboloide ad una falda perpendicolare;
17. iperboloide a due falde.

[16] Gli aggettivi: parallelo, inclinato e perpendicolare, qui come nel seguito, sono riferiti all'asse di rotazione del cono retto.

Facendo riferimento a particolari intersezioni tra quadriche, come quelle tra due cilindri, tra un cilindro ed una cupola, e tra due cilindri ed una cupola, prendendo in considerazione le tre forme ellittico, parabolico ed iperbolico (senza mescolarle tra loro), si ottengono altri quindici casi, per un totale di trentadue casi distinti (relativamente ai quali devono ancora essere studiate le varie linee d'intersezione):

18. due cilindri ellittici per unione;
19. due cilindri ellittici per intersezione;
20. due cilindri parabolici per unione;
21. due cilindri parabolici per intersezione;
22. due cilindri iperbolici per unione;
23. due cilindro iperbolici per intersezione;
24. un cilindro ellittico ed una cupola ellittica;
25. un cilindro parabolico ed una cupola parabolica;
26. un cilindro iperbolico ed una cupola iperbolica;
27. due cilindri ellittici ed una cupola ellittica per unione;
28. due cilindri ellittici ed una cupola ellittica per intersezione;
29. due cilindri parabolici ed una cupola parabolica per unione;
30. due cilindri parabolici ed una cupola parabolica per intersezione;
31. due cilindri iperbolici ed una cupola iperbolica per unione;
32. due cilindro iperbolici ed una cupola iperbolica per intersezione.

La fig. B.3 mostra le volte a crociera ed a padiglione, ottenute rispettivamnte per unione ed intersezione di due cilindri ortogonali tra loro [17].

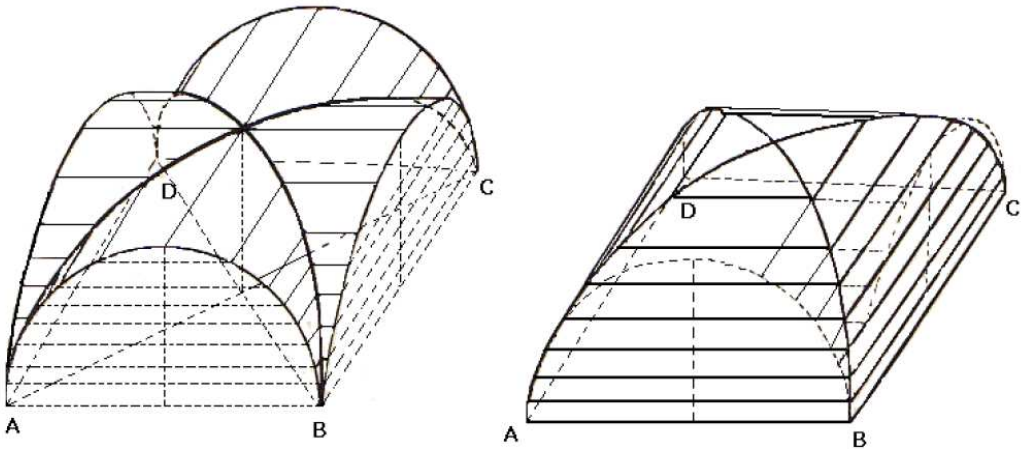
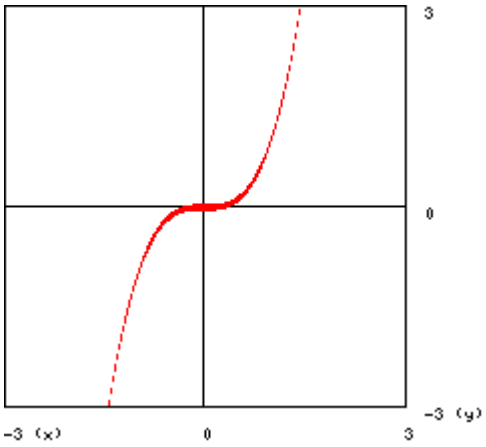


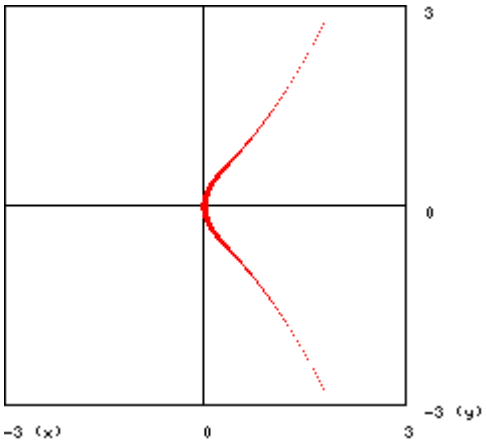
Fig. B.3 – Volte a crociera ed a padiglione

Quali esempi di linee d'intersezione (con l'evidente limite di essere solo curve piane e non anche gobbe), i grafici a seguire, riportati nella fig. B.4, mostrano alcune delle 78 cubiche, studiate già da Newton e successivamente da Stirling: in particolare, oltre alla parabola cubica semplice, le parabole cubiche pura, con ovale, puntuta, annodata e cuspidata.

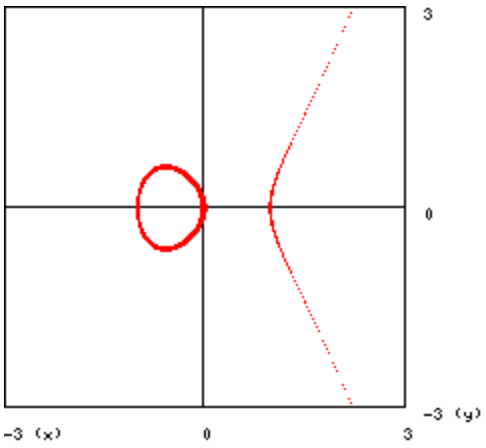
[17] L'unione di due cilindri ortogonali tra loro considera di ciascuno solo la regione esterna all'altro, mentre l'intersezione di due cilindri considera di ciascuno solo la regione interna all'altro.



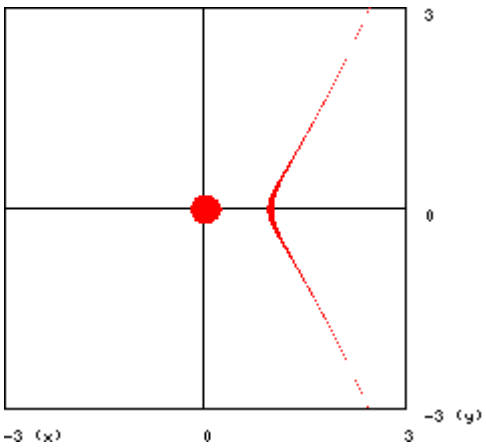
Parabola cubica semplice



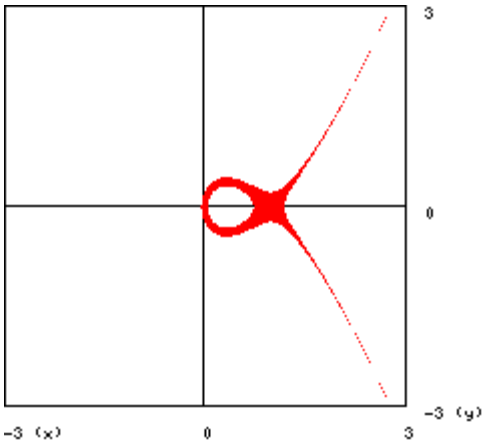
Parabola cubica pura



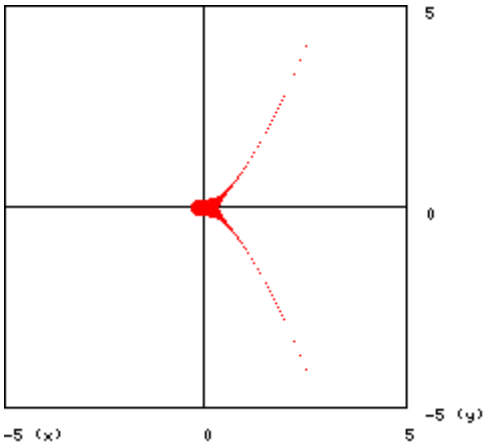
Parabola cubica con ovale



Parabola cubica puntuta



Parabola cubica annodata



Parabola cubica cuspidata

Fig. B.4 – Esempi di parabole cubiche

BIBLIOGRAFIA DI RIFERIMENTO [18]

[18] Il presente lavoro è debitore nei confronti di una lunga serie di articoli, dell'autore e di alcuni suoi amici e colleghi, di cui si dà elenco e ragione, nel prosieguo:

- per la prima parte, Crippa B., Forlani G. (1990): Il calcolo con il metodo degli elementi finiti. Ricerche di Geodesia Topografia e Fotogrammetria, vol. 6, p. 92–137, CLUP, Milano;

1. Bajaj L. C., IHM I: (1992): Smoothing Polyhedra using Implicit Algebraic Splines. Computer Graphics 26, p. 79–86.
2. Bartelme N. (1996): Data structures and data models in GIS. In: Data Acquisition and analysis for multimedia GIS, Springer, NewYork.
3. Bartles R., J. Beatty and B.Barsky (1987): An introduction to the Serpent user Interface Management System. Software Engineering Institute. Carnegie–Mellon University, Pittsburgh (Pa).
4. Bézier P. (1974): Mathematical and Practical Possibilities of UNISURF. Computer Aided Geometric Design, Academic Press, New York.
5. Boissonant J.D. (1984): Geometric structures for Three–dimensional shape representation. ACM Trans. Graphics 3, p. 266–286.
6. Catmull E., Rom R., (1974): A class of Local Interpolating Splines. In: Barnhill R., Riesenfeld R. (Ed's) Computer Aided Geometric Design, p. 317–326, Academic Press, San Francisco.
7. Egenhofer M. J., Mark D. M. (1995): Naive Geography, COSIT, XXX.
8. Fang T.P., Piegl L.A. (1993): Delaunay Triangulation Using a Uniform Grid. IEEE Computer Graphics Applications 13(3), p. 36–47.
9. Fang T.P., Piegl L.A. (1994): Algorithm for Constrained Delaunay Triangulation. The Visual Computer 10, p. 255–265.
10. Floriani D. L. (1987): Surface representation based on triangular grids. The Visual Computer 3, p. 27–50.
11. Floriani D. L., Bruzzone E. (1990): Two data structures for building tetrahedralizations. The Visual Computer 3, p. 27–50.
12. Foley Th. A., Nielson G. M. (1989): Knot Selection for Parametric Spline Interpolation. In: Lyche T., Shumaker L. (Ed's) Mathemathical Methods in Computer Aided Geometric Design, p. 261–272, Academic Press, San Francisco.
13. Giubas L., Stolfi J. (1985): Primitive for the manipulation of general subdivision and the computation of Voronoi diagrams. ACM Trans. Graphics 4, p. 74–123.
14. Green P.J., Sibson R. (1978): Computing Dirichlet tesselations in the plane. Computer Journal 21, p.168–173.
15. Halstead M., Kass M., De Rose T. (1993): Efficient Fair Interpolation using Catmull–Clark surface. Computer Graphics, Annual Conference Series, p. 35–43.
16. Hannan B. (1994): A data reduction scheme for triangulated surface. Computer Aided Geometric Design 11, p. 197–214.
17. Hoppe H., De Rose T., Duchamp (1992): Surface Reconstruction from Unorganized Points. Computer Graphics 26, p. 71–78.
18. Hoppe H., De Rose T., Duchamp (1993): Mesh Optimization. Proc. Computer Graphics, Annual Conference Series, p. 19–25.
19. Joe B. (1989): Three dimensional triangulation from local transformations. SIAM J. Sci. Stat. Comput. 10, p. 718–741.
20. Joe B. (1991): Construction of three dimensional Delaunay triangulation using local transformations. Computer Aided Geometric Design 8, p. 123–142.
21. Kunii T.L., Satoh T., Yamaguchi K. (1995): Generation of topological representations for octree encoding. IEEE CG&A.
22. Laurini R., Thompson D. (1996): Fundamentals of Spatial Information Systems, Academic Press, London.
23. Lawson C.L. (1977): Software for C^1 surface interpolation. In: Rice J.R. (Ed) Mathematical software III, p. 161–194, Academic Press, San Francisco.
24. Lee D.T.,Sachter B.J. (1980): Two algorithmis for costructing Delaunay triangulations. Int. Journal of Computer and Information Sciences 9, p. 219–242.

- per la seconda e terza parte, Bellone T., Mussio L. (1997): Modellizzazione 3D. Atti della Conferenza *Le Immagini e le Informazioni Territoriali* – 1° Conferenza Nazionale dell'ASITA, p. 90–97, Pa rma;
- per la quarta parte, Mussio L., Tucci G. (2000): Modellazione solida e GIS tridimensionali. Atti della Conferenza *Informazione Geografica: innovazione e formazione* – 4° Conferenza Nazionale dell'ASITA, vol. 2, p.113 7–1142, Genova;
- Crippa B., De Haan A., Mussio L.(1992): The Creation of a DEM from SPOT Data. In: T. D. Guyenne, J.J. Hunt (Ed's) *Environment Observation and Climate Modeling through International Space Projects.* ESA ISY-2, ESTEC, Noordwijk, p. 433-438;
- Mussio L. (2002): Georeferencing and integration of digital images from space with remote sensored data from active sensors and geopotential observations. In: B. Betti, F. Sansò, V. Tornatore (Ed's) *Ricerche di Geodesia, Topografia e Fotogrammetria*, n. 13. CLUP, Milano, p. 129-136;
- Mussio L., Bellone T., Malinverni E.S. (1995): Validazione di dati di rilevamento degli inquinanti raccolti in differenti istanti e/o diverse località nel corso di periodi ed in aree geografiche significativi. Boll. della SIFET, n. 4, 1995;
- per la quinta parte, Mussio L. (1984): Il calcolo delle trasformazioni piane elementari. Ricerche di Geodesia Topografia e Fotogrammetria, vol. 4, p. 253–273, CLUP, Milano.

25. Letcher J.S. Jr, Shook D.M., Shepherd S.G.(1995): Relational Geometric Synthesis. Computer Aided Design, 27.
26. Li D. (1996): Three Dimensional GIS and Geoinformatics. IntArchPRS, Wuhan Technical University of Survay and Mapping (China).
27. Loop C. (1994): A G^1 triangular spline surface of arbitrary topological type. Computer Graphics 11, p. 303–330.
28. Mark D. M. (1993): Human spatial cognition. In: Medyckyj–Scott D. and Hearnshaw H. M. (Ed's) Human Factors in Geographical Information Systems, Belhaven Press, p. 51–60.
29. Meagher D.J.(1982): Geometric modeling using octree encoding. Computer graphics and Image Processing, 19, 129-147.
30. Meyers D., Skinner Sh. (1992): Surface from Contours, ACM Trans. Graphics 11/3.
31. Mirante A., Weingarten N. (1982): The radial sweep algorithm for constructing triangulated irregular networks. IEEE Computer Graphics & Applications 2, p. 11–21.
32. Moolenar M., De Hoop S. (1994): Advanced geographic moddelling – Spatial data modelling and query languages for 2D and 3D application, Delft.
33. Mortenson M. E. (1997): Geometric Modeling. John Wiley and Sons Inc., New York.
34. Preparata F. P., Shamos M. I. (1985): Computational Geometry, an Introduction, Springer, New York.
35. Requicha A.A.G. (1977): Mathematical models of rigid solid objects. Technical Memo, No. 28, Production Automation Project, University of Rochester.
36. Rikkers R., Moolenar M., Stuiver J. (1994): A query oriented implementation of a topologic data structure for 3-dimensional vector maps. Int. J. Geographical Information System, vol. 8, n. 3, 243 -260.
37. Schumaker L. (1987): Triangulation methods. In: Chui C.K., Schumaker L., Utreas F.I. (Ed's) Topics in Multivariate Approximation, p. 219–232, Academic Press, San Francisco.
38. Thurston W. (1997): Three-dimensional Geometry and Topology. Princeton University Press, Princeton
39. Weeks J. (1985): The Shape of Space, Dekker, New York.

40. Interopérabilité pour l'utilisation généralisée de la Géoinformation (2005), Carosio A. (Ed), ETHZ.

Fig. N.1 - China Central TeleVision (CCTV) di Beijing
(concavo, non stellato e multiconnesso)