

IST. EL. INF.
BIBLIOT. A
Posiz. ARCOFLO

UN CONFRONTO OPERATIVO FRA VARI METODI
DI INTERPOLAZIONE GRAFICA

F.Barsi, L.Milanesi

Rapporto B4-47
Contratto Selenia-CNR
Settembre 1986

convenzione

SELENIA - CNR

ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

PISA

Introduzione

Il problema di interpolare "in maniera morbida" insiemi ordinati di punti e' stato affrontato da tempo sia con lo scopo di estrarre la maggior quantita' di informazione possibile da dati sperimentali sia, in applicazioni piu' propriamente grafiche, con lo scopo di riprodurre e memorizzare forme date con un numero limitato di punti. In questa sede, l'attenzione sara' rivolta al secondo degli aspetti citati.

I metodi di interpolazione finora proposti forniscono tutti valide soluzioni ma presentano comportamenti assai diversi fra loro. Tali comportamenti sono agevolmente quantificabili dall'analisi delle espressioni matematiche che definiscono i singoli metodi : da tali espressioni si possono cioe' ricavare le caratteristiche di continuita' delle curve di interpolazione e delle loro derivate, gli andamenti della curvatura, la presenza o meno di proprieta' locale.

Tutto cio' e' di estrema importanza, ma, da un punto di vista d'utente, non esaurisce l'insieme delle conoscenze necessarie a chi, dovendo riprodurre o memorizzare una forma, deve scegliere il metodo di interpolazione migliore e, in conseguenza, determinare un insieme adeguato di punti che gli consenta di ricostruire la forma data con un errore limitato.

E' bene precisare che tale "errore" non e' riconducibile, quando si considerano applicazioni grafiche, al solo scarto fra la forma riprodotta e l'originale ma richiede la considerazione di molteplici altri elementi quali l'assenza di "rumore grafico" (che si traduce in incertezza del tratto nel disegno) e di variazioni brusche e sgradevoli della curvatura .

Queste osservazioni portano a concludere che un "giudizio visivo" sul risultato ottenuto applicando un metodo di interpolazione e' in definitiva l'unico valido metro della bonta' di una riproduzione grafica . Ed e' importante osservare che e' a tale criterio che si riferiscono i principali studiosi.

Acquista quindi notevole interesse un confronto pratico fra i risultati forniti dai vari metodi nel riprodurre curve fondamentali quali circonferenze e

coniche in genere, poligoni e sinusoidi . La riproduzione di tali curve , gia' di per se' significativa per molte applicazioni , assume ancor piu' importanza ove si osservi che figure complesse possono pensarsi "scomponibili" in tratti di esse.

L'interesse per un tale confronto e' ulteriormente rafforzato dalla mancanza di qualcosa di analogo nella letteratura corrente. Beninteso, cio' non e' senza motivo. Le conclusioni che da tali confronti possono essere tratte devono essere molto caute per le ragioni che seguono:

- metro di "giudizio visivo" : infatti, anche se tutti concordano nel ritenerlo come l'unico valido in applicazioni grafiche esso ha la caratteristica di non essere che scarsamente quantificabile e quindi conserva un'impronta soggettiva che non consente mai conclusioni drastiche ;
- la scelta dei punti su cui deve operare il metodo e' ancora affidata all'esperienza dell'utente. Da questo punto di vista particolare interesse rivestono i metodi di interpolazione che godono di proprieta' locale, ovvero quei metodi in cui lo spostamento di un punto influisce non su tutta la curva interpolata, ma solo su una ben definita porzione di essa .

Nonostante gli argomenti precedenti consiglino quindi notevole cautela nel valutare i risultati di un confronto pratico sul comportamento dei vari metodi di interpolazione, e' parso che, non spingendosi oltre certi dettagli nella valutazione , potessero essere messi in evidenza (e quindi confrontati) almeno i seguenti aspetti :

- numero dei punti necessari per riprodurre una data forma
- capacita' dei singoli metodi di riprodurre forme anche molto diverse tra loro (es. circonferenze e poligoni)
- facilita' di impiego da parte di un utente non particolarmente esperto .

Con le premesse elencate, i risultati di questo lavoro possono essere di interesse non solo accademico, ma anche

pratico e possono stimolare ad ulteriore attivita' tesa sia al miglioramento dei metodi in genere sia alla selezione di insiemi ottimali o pseudo-ottimali di punti da interpolare .

RASSEGNA DEI PRINCIPALI METODI DI INTERPOLAZIONE

Negli ultimi venti anni i problemi concernenti la generazione e la riproduzione di curve e superfici hanno visto radicalmente mutata la loro impostazione attraverso l'introduzione e l'uso massiccio degli elaboratori elettronici.

Diversi rami dell'informatica attuale si occupano esplicitamente di studiare e progettare architetture Hardware e di sviluppare pacchetti Software per una gestione ottimale di tali problemi e di altre problematiche connesse: CAD (Computer Aided Design), CAM (Computer Aided Manufacturing), CAGD (Computer Aided Geometric Design) ecc. ecc.

La generazione di curve e superfici e' di fondamentale importanza nelle attivita' piu' svariate come la progettazione di carrozzerie di auto, scafi di navi, fusoliere ed ali di aereo, pale di eliche, soles di scarpe, bottiglie ecc. ma anche per la descrizione di fenomeni geologici, fisici e medici.

Prima dell'avvento dei computers si utilizzavano strumenti di "geometria descrittiva. Una superficie era definita da un insieme di curve: questa informazione era sufficiente a determinare una sagoma dalla quale si otteneva poi il prodotto finale.

Oggi invece e' possibile "pilotare" direttamente la generazione delle forme attraverso il controllo di un elaboratore: per far cio' e' pero' necessario "convertire" le definizioni delle curve e delle superfici in una rappresentazione computer-compatibile.

Come passo iniziale le forme sono "digitalizzate" cioe' sostituite con delle punteggiate con procedimenti che possono essere manuali od automatici.

I metodi che consentono di ripercorrere il cammino a ritroso, cioe' di ricostituire la forma originaria a partire da un insieme discreto di punti saranno riconsiderati limitandosi pero' al loro studio nel piano: infatti, nonostante gli oggetti del mondo reale siano sempre tridimensionali, la maggioranza delle tecniche di disegno con cui si e' soliti reappresentare questa realta' e' bidimensionale.

Nel prendere in esame i metodi che a partire da un insieme discreto di punti, definiscono un modello matematico per la generazione di curve, e' importante chiarire come tali curve debbano essere legate a questi

punti: ci sono metodi che generano curve che passano "per" tutti i punti dati (e allora si parla di INTERPOLAZIONE) e metodi nei quali le curve passano "vicino" ai punti dati (APPROSSIMAZIONE).

In cio' che segue ci occuperemo esclusivamente di metodi di interpolazione con la sola eccezione di un metodo molto usato che non rientra esattamente in nessuna delle due categorie citate. Diciamo subito che tali metodi saranno da considerarsi tanto migliori quanto piu' presenteranno alcune delle caratteristiche sotto elencate:

- LOCALITA': la modifica di una porzione di curva non si ripercuote sull'intera figura. Per ottenere cio', uno dei metodi piu' semplici e' realizzare l'interpolazione giustapponendo archi di curva ciascuno dei quali e' interamente determinato da un sottoinsieme di punti dati: in generale tali archi congiungono coppie di punti adiacenti e prendono il nome di SPLINES.
- FLESSIBILITA': capacita' di riprodurre le forme piu' varie.
- CONTINUITA': e' richiesta la continuita' almeno della derivata prima. La continuita' delle derivate di ordine superiore e' desiderabile ma non essenziale in molte applicazioni.
- INVARIANZA della curva di interpolazione rispetto a rotazioni, traslazioni, cambiamenti di fattore di scala.
- la scelta dei PARAMETRI DI CONTROLLO deve essere agevole anche per utenti non particolarmente esperti.

1 INTERPOLAZIONE LINEARE

Sia

$$P_i = (x_i, y_i)$$

il generico punto del piano e sia dato l'insieme ordinato di punti

$$\{P_1, P_2, \dots, P_n\}$$

L'interpolazione lineare connette una qualsiasi coppia

$$\{P_i, P_{i+1}\}$$

con un segmento

$$P(t) = t P_i + (1 - t) P_{i+1}$$

cioè:

$$x(t) = t x_i + (1 - t) x_{i+1}$$

$$y(t) = t y_i + (1 - t) y_{i+1}$$

con $t \in [0, 1]$ e $i = 1, \dots, n-1$.

Notiamo che la curva di interpolazione che così si ottiene ha le seguenti caratteristiche:

- passa per i punti dati
- è continua
- è invariante per trasformazioni affini quali rotazioni, traslazioni, cambiamenti di fattore di scala.

Questo metodo trova la sua applicazione ove sia necessario riprodurre con una procedura semplice una forma piana; scegliendo opportunamente la successione dei punti, è possibile ottenere un errore assoluto di riproduzione arbitrariamente piccolo.

2 INTERPOLAZIONE POLINOMIALE

L'aggettivo "polinomiale" deriva dal fatto che in questo approccio i punti di un insieme ordinato sono raccordati mediante una curva le cui componenti $x(t)$ e $y(t)$ sono funzioni polinomiali di un parametro reale t .

Dato un insieme ordinato di punti

$$\{P_1, P_2, \dots, P_n\}$$

esiste uno ed un solo polinomio p di grado minore od uguale ad $n-1$ tale che

$$p(x_i) = y_i$$

per ognuno dei punti $P(x_i, y_i)$ dell'insieme.

Se interpolare n punti con un unico polinomio e' concettualmente facile, tuttavia richiede calcoli complessi per determinare i coefficienti; inoltre la curva risultante puo' presentare oscillazioni indesiderate. Le difficoltà citate crescono con il valore di n .

Conviene percio', in generale, usare un polinomio di grado k ($k < n$) opportuno che interpoli al piu' $(k+1)$ punti con tratti polinomiali adiacenti: se tale curva polinomiale a tratti di grado k ha derivate continue fino all'ordine $(k-1)$ nei punti di raccordo di due tratti, essa prende il nome di SPLINE MATEMATICA.

Le diverse scelte del grado k hanno portato alla definizione di vari tipi di spline.

CURVE DI BEZIER

Storia

le curve oggi conosciute come curve di Bezier, furono separatamente sviluppate da P. de Casteljaou nel 1959 e da P. Bezier nel 1962.

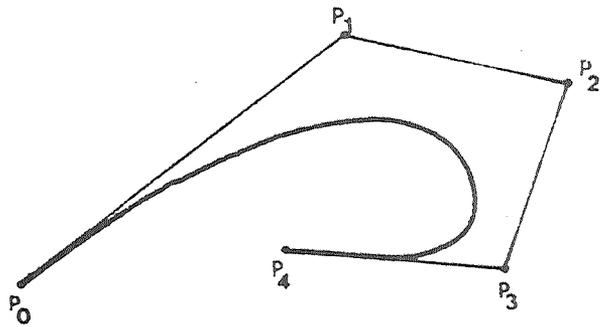
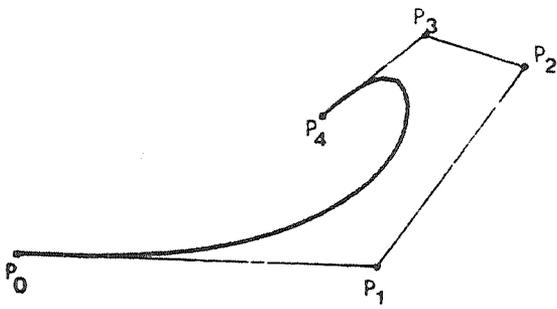
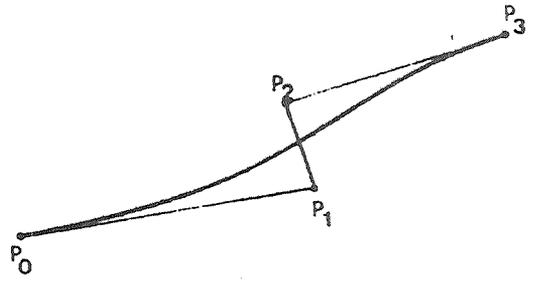
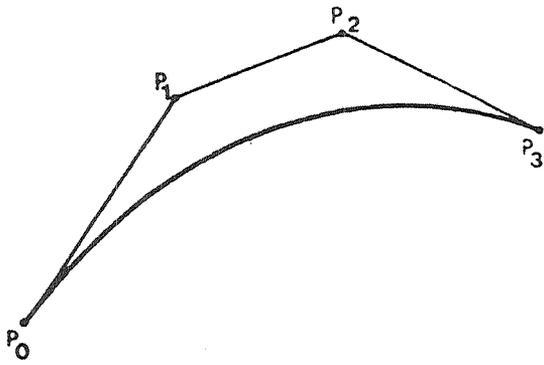


FIG 1

La teoria matematica che sta a monte del metodo e' basata sulla definizione dei polinomi di Bernstein anche se fu solo nel 1970 che R. Forrest scopri' le relazioni fra il lavoro svolto da Bezier e tali polinomi.

Bezier e de Casteljaou svilupparono le loro teorie per un sistema CAD assemblato in Francia dalle fabbriche di auto Renault e Citroen.

Le curve di Bezier costituiscono tuttora la base di molti sistemi CAD nonche' di nuovi metodi per la descrizione di forme e superfici.

Il metodo

Una curva di Bezier viene definita mediante una successione di punti che in generale non appartengono ad essa (tranne il primo e l'ultimo) ma che costituiscono i vertici di un poligono che, in qualche modo, decide l'andamento della curva (figura 1).

Sia

$$\{P_0, P_1, P_2, \dots, P_n\}$$

la successione dei punti. Allora il generico punto della curva $P(t)=(x(t),y(t))$ sara' espresso come:

$$P(t) = \sum_{i=0}^n P_i J_{n,i}(t) \quad 0 \leq t \leq 1$$

dove:

$$J_{n,i}(t) = \binom{n}{i} t^i (1-t)^{n-i}$$

sono i polinomi di Bernstein.

Il grado dei polinomi di Bernstein e' quindi uguale al numero di vertici del poligono di controllo meno uno.

Inoltre, poiche':

$$J_{n,0}(0) = 1 \quad \text{e} \quad J_{n,n}(1) = 1$$

si ha che

$$P(0) = P_0 \quad \text{e} \quad P(1) = P_n$$

cioè i punti iniziale e finale della sequenza da interpolare appartengono alla curva di Bezier.

Si può dimostrare che i polinomi di Bernstein godono delle seguenti proprietà:

$$1) \quad \sum_{k=0}^n J_{n,k}(t) = 1 \quad t \in [0,1]$$

che assicura che la relazione esistente tra la curva e i suoi vertici di definizione è invariante per trasformazioni affini.

$$2) \quad J_{n,i}(t) \geq 0 \quad \text{per ogni} \quad t \in [0,1]$$

$$J_{n,i}(t) = 0 \quad \text{per ogni} \quad i < 0, i > n$$

questa proprietà, insieme alla (1), garantisce che il segmento di curva $P(t)$ con $t \in [0, 1]$ è interamente contenuto nell'involucro convesso dei punti P_i .

$$3) \quad J_{n,i}(t) = (1-t) J_{n,i-1}(t) + t J_{n-1,i-1}(t)$$

questa proprietà è utile poiché fornisce un pratico metodo per il calcolo dei polinomi di Bernstein.

Inoltre si osserva che le tangenti agli estremi sono parallele ai lati iniziale e finale del poligono di definizione.

B-SPLINE, CUBIC SPLINE, CARDINAL SPLINE

Il metodo B-spline

Il metodo B-spline, che non e' altro che una generalizzazione del metodo di Bezier, ne migliora alcuni aspetti e cioe':

- il grado del polinomio che definisce la curva non e' piu' rigidamente legato al numero dei vertici del poligono di controllo
- lo spostamento di un vertice non influisce sull'intero tratto di curva definita dal poligono a cui il vertice appartiene.

I polinomi caratteristici della B-spline differiscono da quelli di Bernstein per il loro comportamento "non-globale", cioe' sono funzioni $N_{i,k}(t)$ diverse da 0 solo su un sottointervallo $[x_i, x_{i+k+1}]$ dell'intervallo di definizione del parametro.

In questo modo ogni vertice del poligono di definizione P_i e' associato ad un'unica funzione: da qui discende la non globalita' del metodo aumentando di fatto la capacita' di localita' del metodo stesso.

Inoltre le B-splines permettono di cambiare l'ordine della curva senza dover cambiare il numero dei vertici del poligono di definizione.

L'equazione parametrica della B-spline e':

$$P(t) = \sum_{i=0}^n P_i N_{i,k}(t)$$

dove i P_i sono gli $n+1$ punti di interpolazione che formano i vertici del poligono di definizione.

Le quantita' $N_{i,k}(t)$ sono ricorsivamente definite a partire da una serie ordinata di numeri reali detti "nodi" (knots):

$$a_1 \leq a_2 \leq \dots \leq a_n$$

che a loro volta formano il cosiddetto "vettore dei nodi"

$$[a_1, a_2, \dots, a_n]$$

Oltre al vettore dei nodi, un altro parametro fondamentale del metodo e' rappresentato dalla scelta dell'ordine della curva: si puo' dimostrare che fissando l'ordine della curva k uguale al numero dei vertici del poligono di controllo, si ottiene una curva di Bezier.

Al decrescere dell'ordine la curva ottenuta tende ad "adagiarsi" sul poligono di definizione fino a coincidere con esso per $k=2$.

L'ordine della curva e' importante anche dal punto di vista della "morbidezza" in quanto influisce sulla continuita' delle derivate dei vari ordini.

In generale la B-spline e' detta funzione polinomiale di ordine k (e grado $k-1$) in quanto verifica le seguenti condizioni:

- la funzione di interpolazione $P(t)$ e' polinomiale di grado $k-1$ su ciascun intervallo $(a_i \leq t \leq a_{i+1})$
- $P(t)$ e le sue derivate di ordine $1, 2, \dots, k-2$ sono tutte continue sull'intera curva.

In conclusione le B-spline sono nettamente piu' flessibili delle curve di Bezier poiche' l'utente, per modificare la forma della curva puo' apportare ai parametri di controllo modifiche di tipo differente: cambiando il grado k della curva, con $2 \leq k \leq n+1$, oppure cambiando il numero o la posizione dei vertici del poligono di definizione o la molteplicita' dei nodi.

D'altra parte il loro uso e' piu' complesso e

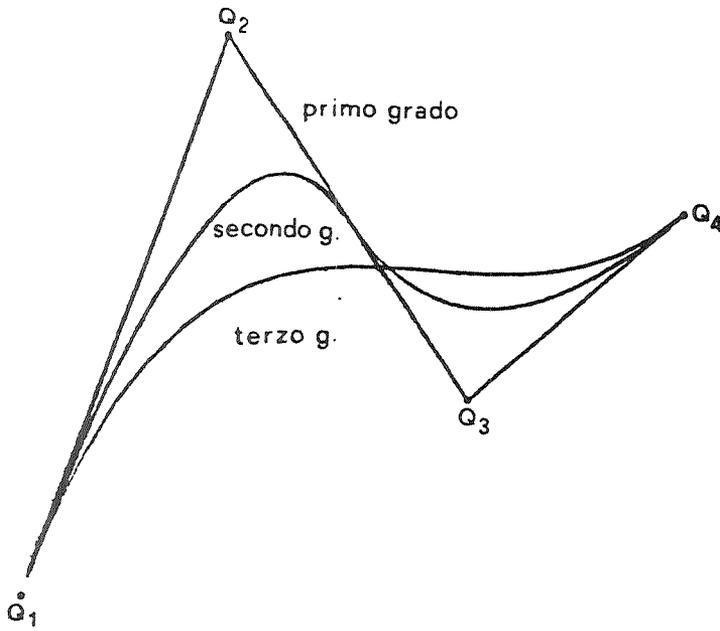


figura 2

richiede una maggiore esperienza da parte dell'utente.

A titolo di esempio, la figura 2 mostra come varia una curva al variare dell'ordine.

Il metodo Spline Cubica (Cubic Spline)

In generale una spline matematica e' formata, come detto, da un insieme di tratti polinomiali di grado k con derivate continue fino all'ordine k-1 nei punti di giunzione dei vari tratti.

Come si puo' intuire dal nome, la spline cubica e' definita da un polinomio di grado tre, per cui possiede continuita' della derivata prima e seconda nei punti di raccordo dei vari tratti.

E' comune usare splines definite da un polinomio di grado basso (come nel caso della cubica) poiche' questo riduce sia i tempi di calcolo che le instabilita' proprie di ordine elevato.

D'altra parte, poiche' i metodi con curve di piccolo ordine non possono interpolare un numero arbitrario di punti, e' necessario interpolare dei tratti adiacenti di curva.

Proprio sulla base di queste considerazioni, le spline cubiche sono usate per interpolare una serie di punti dati considerandoli due alla volta: cioe', presi due punti, si trova il tratto di spline che li interpola e si passa alla coppia successiva che ha come primo punto il secondo punto del tratto precedente.

Un ulteriore vantaggio delle cubiche consiste nel fatto che esse sono le curve di ordine piu' piccolo con capacita' di possedere flessi e di produrre serpentine.

L'equazione di un singolo tratto di spline (che unisce due punti) e':

$$P(t) = \sum_{i=1}^4 B_i t^{i-1} \quad t_1 \leq t \leq t_2 \quad (1)$$

dove i coefficienti B_i sono determinati dalle condizioni agli estremi del tratto .

In forma estesa la funzione $P(t)$ e' scritta come

$$P(t) = B_1 + B_2 t + B_3 t^2 + B_4 t^3 \quad (2)$$

Vediamo ora come si ottengono i coefficienti B_i .

Siano i punti P_1 e P_2 gli estremi di un tratto e P_1' e P_2' le derivate in questi punti rispetto al parametro t e supponiamo infine che il parametro t vari fra t_1 e t_2 nel tratto tra P_1 e P_2 . Per semplicita' si puo' porre $t_1 = 0$: le condizioni da imporre agli estremi del tratto di interpolazione riguardano i punti stessi e le tangenti in tali punti

$$\begin{array}{ll} P(t_1) = P_1 & P(t_2) = P_2 \\ dP/dt (t=t_1) = P_1' & dP/dt (t=t_2) = P_2' \end{array}$$

Queste quattro relazioni, applicate alla definizione (1) della spline portano alla determinazione dei coefficienti:

$$\begin{array}{ll} B_1 = P_1 & B_2 = P_1' \\ B_3 = [3(P_2 - P_1)/t_{12}] - [2 P_1'/t_{12}] - [P_2'/t_{12}] & \\ B_4 = [2(P_2 - P_1)/t_{12}] + [P_1'/t_{12}] + [P_2'/t_{12}] & \end{array}$$

da cui si ricava che i coefficienti dipendono dai punti estremi del tratto e dalle derivate prime in tali punti.

Per maggior chiarezza notiamo che nel piano i coefficienti B_i hanno due componenti, una per l'ascissa, che indicheremo con B_{ix} , e una per l'ordinata indicata con B_{iy} per cui le componenti di $P(t)$ sono

$$x(t) = B_{1x} + B_{2x} t + B_{3x} t^2 + B_{4x} t^3$$

$$y(t) = B_{1y} + B_{2y} t + B_{3y} t^2 + B_{4y} t^3$$

Supponendo ora di avere n punti dati e di conoscere le derivate nel punto iniziale e finale, imponiamo la continuita' della derivata seconda nei punti di giunzione dei vari tratti di spline: per far questo e' necessario che la curvatura rimanga costante in tali punti. Si ottiene in tal modo un sistema di n-2 equazioni in n incognite, cioe' le tangenti nei punti di interpolazione: e' quindi necessario fissare due ulteriori valori delle tangenti (e precisamente nel punto iniziale e finale dell'intera curva) affinche' il sistema risulti risolubile. Tale sistema puo' essere espresso in forma matriciale:

$$A p = b$$

dove A e' una matrice tridiagonale contenente i valori dei parametri iniziali e finali dei vari tratti, p e' il vettore colonna delle tangenti incognite:

$$p = [P'_1, P'_2, \dots, P'_n]$$

e b rappresenta il vettore colonna dei termini noti.

Ricapitolando quanto detto finora abbiamo che :

- il metodo prevede il calcolo dei coefficienti B che sono differenti da tratto a tratto (un tratto unisce due punti consecutivi).
- per il calcolo dei coefficienti e' necessario determinare le tangenti nei punti di interpolazione intermedi, calcolo che prevede l'inversione di una matrice e il suo prodotto per un'altra (p = A b).

A questo punto non resta che esaminare i modi in cui

vengono fissate le tangenti nel primo e nell'ultimo punto di interpolazione affinche' il sistema $A p = b$ sia risolubile.

Il primo modo, che e' quello piu' diretto, e' chiamato CLAMPED END e prevede di fissare a priori il valore di P_1 e P_n affidando questa scelta all'utente : cioe', e' l'utente stesso che deve fornire i valori di queste tangenti immettendoli come dati del problema.

Questo modo di procedere presenta vantaggi e svantaggi abbastanza chiari :

- il vantaggio e' la massima liberta' che l'utente ha di "pilotare" l'andamento della curva nel punto iniziale e finale, cosa importante poiche', come vedremo, questi due punti si sono rivelati particolarmente critici negli esempi fatti.
- lo svantaggio e' l'aumento dei parametri che devono essere fissati dall'utente, cosa che, nel caso di un'utenza non particolarmente esperta, si rivela senz'altro poco desiderabile.

Il secondo metodo e' chiamato RELAXED e deriva dall'imposizione della condizione :

$$\frac{d^2 P}{dt^2} = 0$$

Il terzo e il quarto metodo usano le cosiddette CYCLIC e ANTICYCLIC CONDITIONS.

La spline ciclica si usa di solito per generare curve chiuse o porzioni di una curva che si ripete a intervalli.

Le condizioni per la ciclica (CYCLIC) sono :

$$P'_1 = P'_n$$

$$P''_1 = P''_n$$

cioe' la pendenza e la curvatura all'inizio e alla fine della curva coincidono.

Come si puo' intuire dal nome, le condizioni per la anticiclica sono asimmetriche rispetto a quelle della ciclica e cioe':

$$P'_1 = - P'_n$$

$$P''_1 = - P''_n$$

Questo tipo di spline e' usato per produrre forme particolari dove ad esempio sono presenti spigolosita' o "cuspidi" molto accentuate.

Oltre agli innegabili vantaggi gia' citati per le cubiche, fra i quali quello di essere il polinomio di grado piu' basso che possa interpolare una serie di punti, esiste anche qualche nota negativa.

Intanto la continuita' della derivata seconda nei punti intermedi si puo' imporre solo se si conosce l'intero insieme dei punti prima di generare un qualsiasi segmento di curva .

Questo e' un notevole inconveniente nel caso si debba modificare la curva localmente in modo interattivo, cioe' e' soprattutto di localita' che la cubica difetta : lo spostamento di un punto modifica la forma dell' intera curva .

Il metodo Cardinal Spline.

Questo metodo conclude la rassegna dei metodi piu' usati. Le cardinal spline sono principalmente usate in ambito industriale e presentano riferimenti bibliografici notevolmente carenti.

L'idea che sta alla base del metodo Cardinal e' la seguente : supposto di avere una serie di punti da interpolare, si desidera condizionare in qualche modo le tangenti alla spline in tali punti.

Esiste una rappresentazione (detta di Hermite) nella quale il polinomio di interpolazione di terzo grado assume la forma:

$$P(t) = [t^3 \ t^2 \ t \ 1] \begin{matrix} M & G \\ & h \quad h \end{matrix}$$

in cui M e' una matrice detta "di base" e G e' un vettore colonna contenente informazioni geometriche:

$$G = [P_i, P_{i+1}, P'_i, P'_{i+1}]$$

In questo metodo un ruolo di primaria importanza e' rivestito da un parametro "a" che compare nella matrice ed il cui valore influisce pesantemente sull'andamento della curva consentendo ampie possibilita' di scelta. Questa liberta' di scelta ha i suoi lati positivi ma e' indubbio che, soprattutto ove si vogliano riprodurre forme note, il valore di a non e' affatto di banale deduzione.

Nonostante questo il metodo si rivela valido sia per generare separatamente sia tratti morbidi che profili molto frastagliati: e' invece nella riproduzione di profili "misti" che il metodo accusa delle difficolta'.

Infatti un determinato valore del parametro si adatta bene ad un profilo ma non ad un' altro : in presenza di profili dall'andamento estremamente vario, quindi, bisognerebbe essere in grado di variare dinamicamente il valore di a in modo da adattare la curva al profilo stesso.

Anche la Cardinal difetta senz'altro di localita' : l'aggiunta, la modifica o la soppressione di un punto ha conseguenze talvolta imprevedibili sul risultato finale.

3 Il metodo Angular Spline

Il metodo Angular Spline e' stato recentemente sviluppato presso l'IEI di Pisa nell'ambito di un Contratto di ricerca con la Societa' Selenia di Roma.

Partendo da considerazioni geometriche sulla forma della curva e scegliendo un criterio per ottenere ulteriori parametri di controllo dall'insieme dei punti dati, e' stato possibile ricavare l'espressione analitica di una spline che si e' rivelata particolarmente adatta ad interpolare in modo dolce una qualsiasi successione di punti.

La spline interpola due punti adiacenti e la sua forma e' completamente determinata da un sottoinsieme di quattro punti consecutivi .

La curva che si ottiene congiungendo queste splines e' continua con la sua tangente in ogni punto e risulta invariante rispetto a tutte le trasformazioni affini del piano (rotazioni, traslazioni, ecc.).

Trattandosi di un metodo nuovo, daremo di esso un numero di dettagli maggiore di quanto non si sia fatto nella presentazione degli altri metodi e cioe', oltre la definizione, anche le sue principali proprieta' analitiche e grafiche corredate da esempi esplicativi.

Supponiamo di avere un sistema di assi cartesiani x e y e un insieme ordinato di punti del piano $P = \{P_i = (x_i, y_i), i=1, \dots, n\}$ con $P_i \neq P_{i+1}$ per ogni i .

Assumiamo che il vettore tangente alla curva in ogni punto P_i coincida con la tangente al cerchio C_i che passa per i punti P_{i-1}, P_i, P_{i+1} orientata concordemente al verso determinato sulla circonferenza dalla successione ordinata dei punti P_{i-1}, P_i, P_{i+1} .

L'espressione analitica della curva di interpolazione tra due punti consecutivi P_i, P_{i+1} si determina operando in un sistema di riferimento locale cosi' definito (vedi fig. 3)

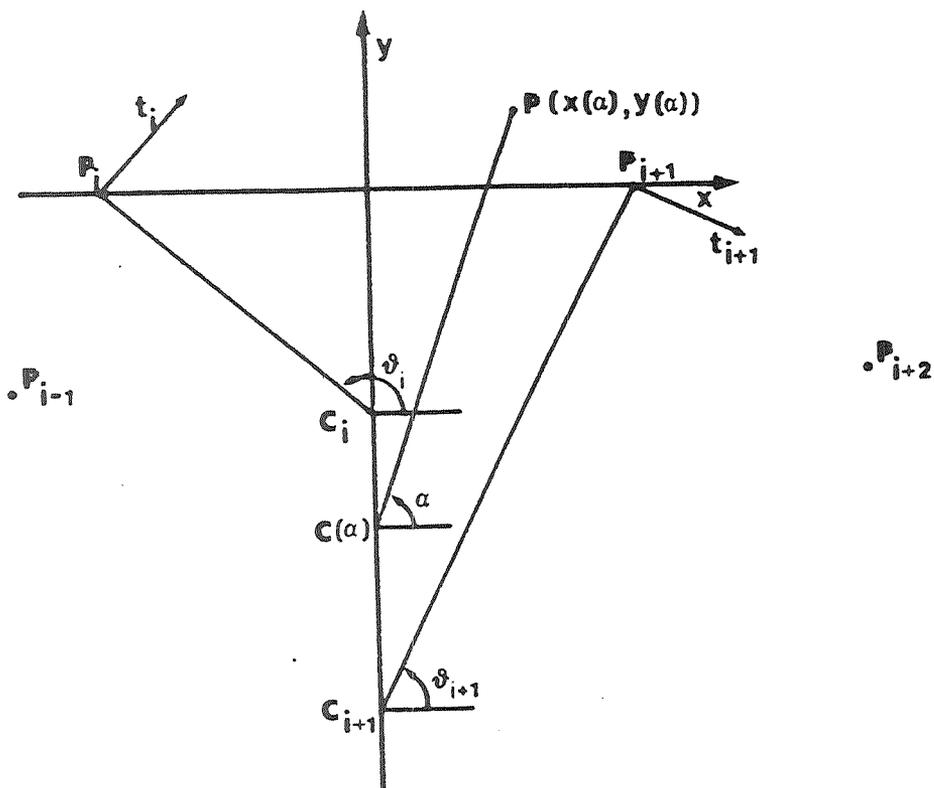


Figura 3

- origine nel punto di mezzo del segmento $P_i P_{i+1}$.
- asse x coincidente con la retta per P_i e P_{i+1} ed orientato in modo che l'ascissa di P_i risulti negativa.

Si osservi che i centri delle circonferenze C_i e C_{i+1} appartengono all'asse y: chiameremo c_i e c_{i+1} le loro ordinate.

Occorre notare che non lede la generalita' supporre che P_{i-1} e P_{i+2} abbiano entrambi ordinata negativa potendosi tutti i casi possibili ricondurre agevolmente a questo.

Sotto questa ipotesi valgono le seguenti relazioni :

- L'angolo $\phi_i = \widehat{x t_i}$ tra l'asse x e la tangente alla curva in P_i e' positivo, mentre $\phi_{i+1} = \widehat{x t_{i+1}}$ e' negativo.
- Gli angoli ϑ_i e ϑ_{i+1} tra l'asse x e i raggi delle circonferenze in P_i e P_{i+1} variano negli intervalli :

$$\vartheta_i \in (\pi/2 , 3\pi/2)$$

$$\vartheta_{i+1} \in (-\pi/2 , \pi/2)$$

La spline angolare viene definita come il luogo geometrico dei punti del piano che appartengono all'intersezione tra le circonferenze che passano per P_i con centro in $(0 , c(\alpha))$ ed il loro raggio inclinato dell'angolo α rispetto all'asse x .

Le equazioni parametriche di un generico punto P della curva sono, in funzione di α , le seguenti:

$$x(\alpha) = \rho(\alpha) \cos \alpha$$

$$y(\alpha) = \rho(\alpha) \sin \alpha + c(\alpha)$$

dove $\rho(\alpha) = \sqrt{[c(\alpha)^2 + d^2]}$, d e' la distanza di P_i dall'origine e $c(\alpha)$ e' una funzione tale che

$$c(\vartheta_i) = c_i$$

$$c(\vartheta_{i+1}) = c_{i+1}$$

e $c(\alpha)$ e' compreso tra c_i e c_{i+1} per qualunque $\alpha \in (\vartheta_i, \vartheta_{i+1})$.

La scelta della funzione $c(\alpha)$ e' di importanza fondamentale per ottenere una curva di interpolazione appropriata. Poiche' risulta

$$c_{i+1} = -d \operatorname{tg} \vartheta_{i+1} \quad c_i = d \operatorname{tg} \vartheta_i = -d \operatorname{tg}(\pi - \vartheta_i)$$

la funzione $c(\alpha)$ scelta e' stata:

$$c(\alpha) = -d \operatorname{tg}(k\alpha + \psi)$$

dove

$$k = (\vartheta_{i+1} + \vartheta_i - \pi) / (\vartheta_{i+1} - \vartheta_i)$$

$$\psi = (\vartheta_{i+1} + \vartheta_i - \pi) \vartheta_i / (-\vartheta_{i+1} + \vartheta_i) + \pi - \vartheta_i$$

Per semplicita' si puo' porre $k\alpha + \psi = \alpha'$.

Si puo' osservare che, in questo sistema di riferimento, l'angolo ϑ_i e' sempre maggiore di ϑ_{i+1} , mentre $\pi - \vartheta_i$ e' maggiore, uguale o minore di ϑ_{i+1} a seconda che c_i sia minore, uguale o maggiore di c_{i+1} rispettivamente: la funzione $\alpha'(\alpha)$ e' pertanto monotona (crescente o decrescente) per α che varia nell'intervallo $[\vartheta_{i+1}, \vartheta_i]$ ed assume valori compresi nell'intervallo $(-\pi/2, \pi/2)$.

Finora si e' supposto che le ordinate di P_{i-1} e P_{i+2} fossero entrambi negative: vediamo ora di ricondurre a

questo tutti gli altri casi.

- Se $P_{i-1} = (x_{i-1}, y_{i-1})$ e $P_{i+2} = (x_{i+2}, y_{i+2})$ hanno entrambi ordinata positiva, la curva di interpolazione viene determinata considerando i punti:

$$P'_{i-1} = (x_{i-1}, -y_{i-1})$$

$$P'_{i+2} = (x_{i+2}, -y_{i+2})$$

ed invertendo poi il segno dell'ordinata dei punti di interpolazione (fig.4).

- se le ordinate dei punti P_{i-1} e P_{i+2} hanno segno discorde, allora, secondo le ipotesi fatte sull'orientamento delle tangenti, i coefficienti angolari delle tangenti in P_i e P_{i+1} rispetto all'origine del sistema di riferimento hanno segno discorde.

In questo caso tra P_i e P_{i+1} esiste un flesso. Attraverso una procedura euristica si determinano il punto di flesso P_F e il valore della tangente in esso: l'interpolazione tra P_i e P_{i+1} viene ricondotta all'interpolazione tra P_i e P_F e tra P_F e P_{i+1} .

- Vediamo ora alcune proprietà del metodo, che dimostrano come esso possieda le caratteristiche necessarie a fornire una interpolazione "morbida": la dimostrazione di queste proprietà è omessa per non appesantire la trattazione.

Ricordiamo che le coordinate del generico punto della spline sono, sostituendo per $c(\alpha)$ nel caso standard:

$$x(\alpha) = d \cos \alpha / \cos \alpha'$$

$$y(\alpha) = d (\sin \alpha - \sin \alpha') / \cos \alpha'$$

Si verifica immediatamente che $x(\alpha)$ ed $y(\alpha)$ possiedono derivate continue di ogni ordine e la curva

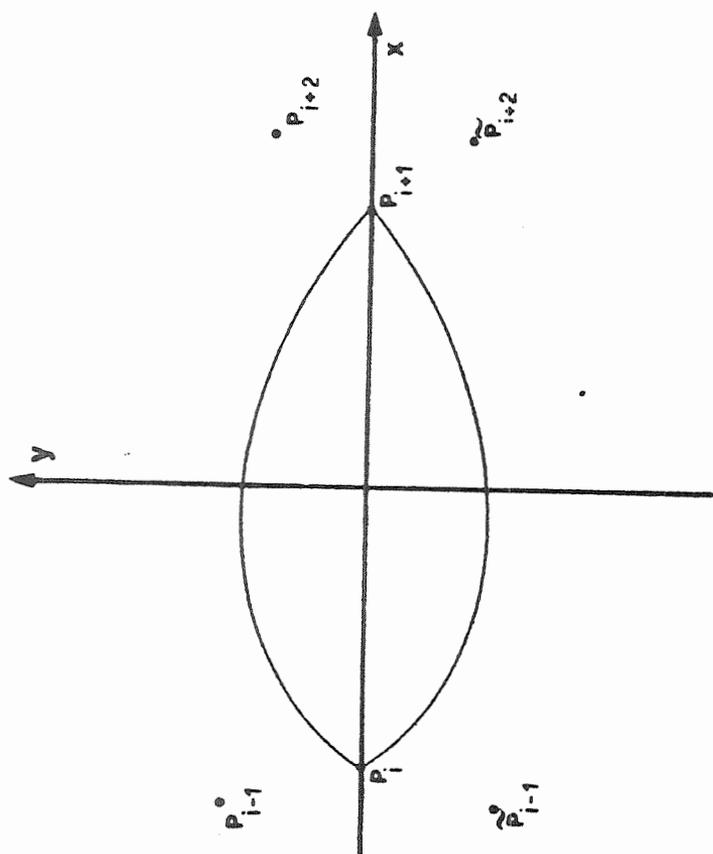


Figura 4

passa per i punti P_i e P_{i+1} .

PROPRIETA' 1 : per qualsiasi coppia di valori ϑ_i ($\pi/2 < \vartheta_i < 3\pi/2$) e ϑ_{i+1} ($-\pi/2 < \vartheta_{i+1} < \pi/2$), $c(\alpha)$ e' compreso tra c_i e c_{i+1} .

PROPRIETA' 2 : la spline angolare giace nel semipiano $y \geq 0$.

PROPRIETA' 3 : la spline angolare ha in ogni punto curvatura positiva.

PROPRIETA' 4 : la spline angolare e' una curva regolare : essa e' cioe' derivabile con continuita' e il modulo del vettore tangente e' diverso da 0.

Inoltre la spline angolare e' invariante per rotazioni e traslazioni poiche' ogni singolo arco viene determinato in un sistema di riferimento locale, dipendente solo dalla posizione dei punti interpolati.

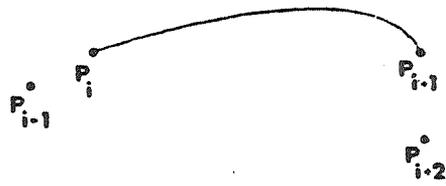
L'interpolazione di una successione di $(n + 2)$ punti e' ricondotta ad $(n - 1)$ passi elementari di interpolazione ciascuno dei quali e' completamente definito dalle posizioni relative di quattro punti della sequenza.

Il primo e l'ultimo punto della successione servono per determinare la tangente nei punti P_1 e P_n , rispettivamente, (se la sequenza va da 0 a $n + 1$): questi punti vengono chiamati "punti guida" del tratto iniziale e finale ed hanno la caratteristica di non appartenere alla curva ma di determinarne l'andamento in tali tratti.

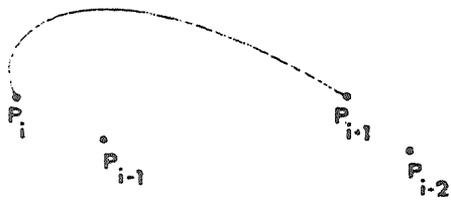
In figura 5 sono mostrati alcuni esempi di interpolazione con l'Angular spline per varie coppie di punti P_{i-1} , P_{i+2} .



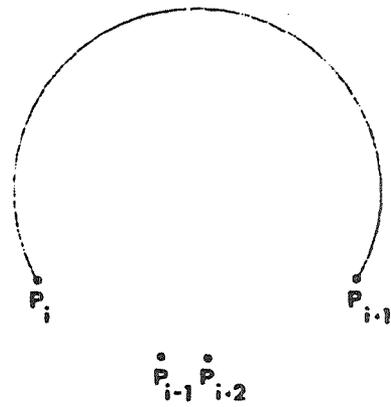
- a -



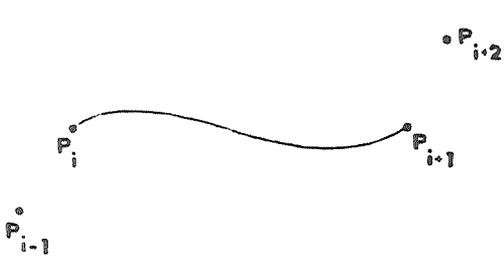
- b -



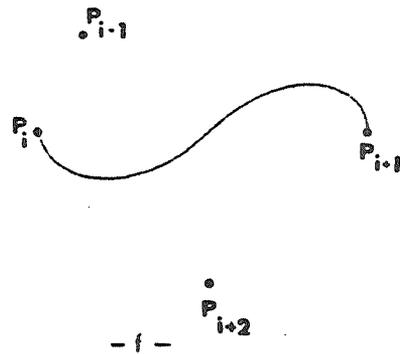
- c -



- d -



- e -



- f -

Fig. 5

ANALISI COMPARATIVA SUL COMPORTAMENTO GRAFICO DEI VARI METODI DI INTERPOLAZIONE

Come e' stato gia' osservato nella parte introduttiva, i metodi di interpolazione vengono impiegati essenzialmente per uno dei seguenti due scopi:

- evidenziare la maggior quantita' di informazione possibile a partire da insiemi di punti che rappresentano dati relativi all'andamento di un qualche fenomeno fisico. E' evidente che in tal caso, ove non esistano ulteriori informazioni relative all'andamento del fenomeno stesso, non esistono concrete possibilita' di individuare metodi ottimali di interpolazione.
- riprodurre o memorizzare forme date a partire da insiemi quanto piu' limiti possibile di punti.

Come gia' detto, cio' che interessa in questa sede e' il secondo degli aspetti citati. Ricordiamo anche che per bonta' di riproduzione si intendera' sia lo scarto limitato tra la forma riprodotta e l'originale, sia il suo gradevole andamento visivo.

Questo criterio porta a stabilire il giudizio visivo come metro unico di valutazione che tiene conto non solo della precisa riproduzione di una forma data ma anche del buon andamento della curva riprodotta .

Giova ricordare che tale giudizio visivo, pur non essendo rigorosamente quantificabile, e' ormai assunto quale metro di valutazione della bonta' di un metodo di interpolazione dalla totalita' degli studiosi .

L'interesse per un confronto fra i comportamenti dei vari metodi di interpolazione e' evidente, anche se e' necessaria una certa cautela nel giudicare i risultati.

Occorre a questo punto stabilire quali forme usare come base di confronto : nonostante il pensiero resti rivolto alla riproduzione di forme complesse, si e' dato, nel seguito, ampio risalto alla riproduzione di forme semplici, sia in quanto di per se' significative, sia in quanto queste possono essere considerate come elementi costitutivi di forme ben piu' complesse. In tal senso tale confronto equivale alla valutazione di un'ampia casistica di figure complesse.

le forme di confronto

Come base di confronto sono state prese figure particolarmente significative ed usate (coniche, poligoni, sinusoidi) e figure complesse di notevole interesse dal punto di vista della loro riproduzione come profili geografici e caratteri tipografici.

Come errore assoluto, da non superare nella riproduzione, si e' fissato uno scarto pari approssimativamente alla larghezza del tratto del disegno. Per ogni metodo usato, sono stati scelti punti appropriati e specifici del metodo stesso in numero quanto minore possibile.

Sulla base del risultato ottenuto e sul numero dei punti che sono stati necessari per conseguirlo, e' stato formulato un giudizio di massima sulla validita' del metodo. Sono state aggiunte anche osservazioni sulla facilita' di impiego dei metodi stessi da parte di utenti non particolarmente esperti.

forme semplici: la mappa dei risultati

Le numerose prove di riproduzione di forme semplici sono state organizzate secondo l'ordine riportato nella Tabella I, che potra' aiutare il lettore nella consultazione dettagliata delle prove stesse.

La tabella indica in colonna le figure riprodotte e in riga i metodi utilizzati: le caselle corrispondenti non sono altro che le figure riportate di seguito alla tabella nell'ordine di scansione previsto. In ciascuna casella sono presenti le informazioni seguenti:

- la posizione dei punti di interpolazione (indicati con crocette)
- il numero di tali punti (non sempre coincidente con il numero delle crocette - punti multipli)
- il valore degli ulteriori parametri di cui il metodo puo' necessitare.

Una precisazione necessaria riguarda la spline cubica: delle 4 versioni ne sono state utilizzate solo tre e precisamente la Relaxed, la ciclica e la aciclica in quanto la versione Clamped richiede la precisazione di

ulteriori parametri di tutt'altro che facile definizione . Inoltre, delle tre prescelte, si e' ogni volta selezionata la versione che forniva i risultati migliori per ciascuna figura. Ciascuna delle caselle relative a queste figure precisa comunque il tipo di cubica usato.

considerazioni particolareggiate sui risultati ottenuti

Le figure che risultano dall'interpolazione sono state riportate insieme alla forma originale consistente in una punteggiata (e quindi facilmente riconoscibile). Le considerazioni che seguono seguono ancora l'ordine della Tabella I: sono cioe' per figura/metodo.

cerchio

angular: nessun parametro aggiuntivo da specificare, numero minimo di punti di interpolazione, nessun vincolo sulla loro posizione, risultato perfetto (la forma analitica diviene quella di una circonferenza) Fissati i tre punti che definiscono la circonferenza, basta replicarli in modo da permettere la richiusura. Come sappiamo, infatti, la Angular necessita di quattro punti per poter disegnare un tratto: percio', fatti i conti, per riprodurre un cerchio con tre tratti occorrono sei punti (che in effetti sono tre considerati due volte) .

La replica di alcuni punti di interpolazione come punti di richiusura e' sfruttata sia per la Angular che per la Cardinal in tutte le figure chiuse.

cardinal: riproduzione piu' difficoltosa. La posizione dei punti non puo' essere qualsiasi ed il valore del parametro deve essere accuratamente scelto in funzione del raggio della circonferenza e della posizione dei punti.

cubica: necessita' di molti punti di interpolazione.

B-spline: e' necessario fissare un ulteriore parametro (ordine della curva); nelle figure chiuse, i migliori risultati si ottengono con l'ordine massimo (il che fa coincidere la B-spline con una curva di Bezier). Nella figura e' illustrato il poligono di definizione della curva (ogni vertice rappresenta un punto di interpolazione). Fra tutti i metodi esaminati, quest'ultimo e' quello che richiede il maggior sforzo da parte dell'utente. Se cio' e' tollerabile per profili

semplici come le coniche, non lo e' piu' quando si tratta di riprodurre forme complesse .

ellisse

Sono state usate due ellissi differenti per il rapporto fra gli assi (0.4 e 0.7).

angular: buon comportamento. Pochi punti necessari (soprattutto con rapporto 0.7).

cardinal: buon comportamento, identico numero di punti, difficolta' nell'individuazione del valore esatto del parametro.

cubica: analogia di comportamento con il caso della circonferenza. Le versioni RELAXED e CYCLIC sono quelle che hanno fornito i risultati migliori.

B-spline: facile la riproduzione dell'ellisse piu' schiacciata, accettabile nell'altro caso.

parabola

Anche per la parabola si sono utilizzate due figure: una piu' "aperta" ed una piu' "chiusa" per avere un giudizio piu' completo sui vari comportamenti.

angular: molto valida. Tre punti (piu' due punti guida) sono stati sufficienti per ottenere un ottimo risultato> Facile la loro scelta.

cardinal: buon risultato. Qualche difficolta' (al solito) per la scelta del parametro.

cubica: buona la versione relaxed, soprattutto con la parabola piu' aperta.

B-spline: buoni risultati, facile la riproduzione (ordine massimo e quindi curva di Bezier).

iperbole

Esattamente come per la parabola, anche l'iperbole sono state considerate due figure, una piu' aperta e l'altra piu' chiusa .

angular: valgono le considerazioni fatte per la parabola.

cardinal: come sopra.

cubica: comportamento migliore rispetto al caso della parabola. Facile l'interpolazione.

B-spline: ordine corrispondente al massimo, buoni risultati.

rettangolo

Questa e' una delle figure che piu' ha messo in evidenza le differenze di comportamento dei vari metodi: la presenza contemporanea di tratti rettilinei e spigoli ha messo in difficolta' piu' di un metodo.

angular: comportamento molto buono, numero di punti ragionevole (dodici piu' tre di richiusura) addensati agli spigoli. Qualche difficolta' nella scelta dei punti.

cardinal: comportamento eccellente, con valore del coefficiente prossimo allo zero; necessari solo quattro punti (piu' tre di richiusura).

cubica: notevoli difficolta' di riproduzione, molti punti necessari, incapacita' della curva di mantenere la forma rettilinea e riprodurre spigoli.

B-spline: comportamento eccellente; con ordine della curva uguale a due, la curva coincide con il poligono guida (e necessita di soli quattro punti).

sinusoide

buon comportamento con tutti i metodi. In dettaglio:

angular: risultato buono, ma con alcune difficolta' probabilmente dovute al metodo euristico impiegato per il trattamento del flesso.

cardinal: risultato buono con basso valore del coefficiente.

cubica: buon comportamento con la versione ACYCLIC; la riproduzione con la versione RELAXED ha portato a scarti leggermente superiori ma a soli quattro punti di

interpolazione. Alcune difficoltà nella scelta dei punti stessi.

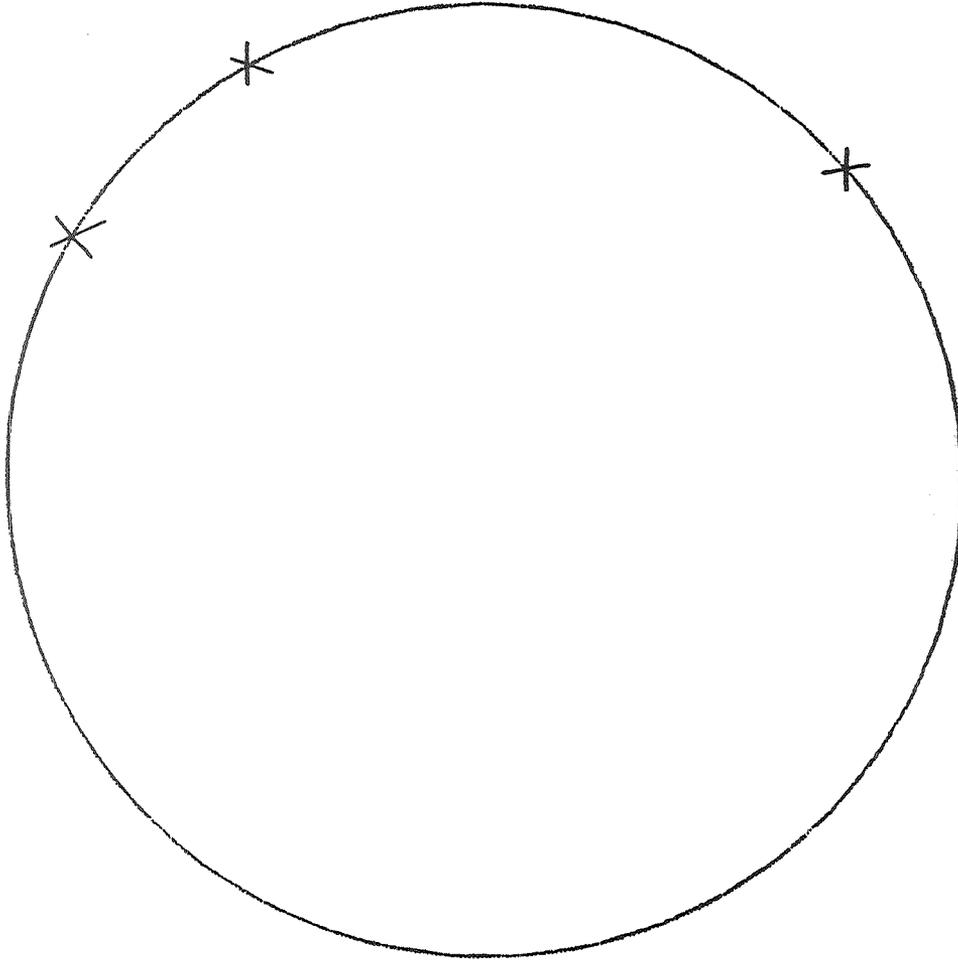
B-spline: risultati discreti, ma notevoli difficoltà nella scelta dei punti.

Metodo usato	Circonferenza	Ellisse 1	Ellisse 2	Parabola 1	Parabola 2	Iperbole 1	Iperbole 2	Sinusoida	Rettangolo
Angular									
Cardinal									
Cubica									
B-spline									

(Le figure riprodotte con i vari metodi, corrispondenti alle caselle della Tabella, sono riportate in dettaglio nelle pagine che seguono. L'ordine di presentazione seguito è quello corrispondente ad una scansione per colonne della Tabella suddetta)

TABELLA I

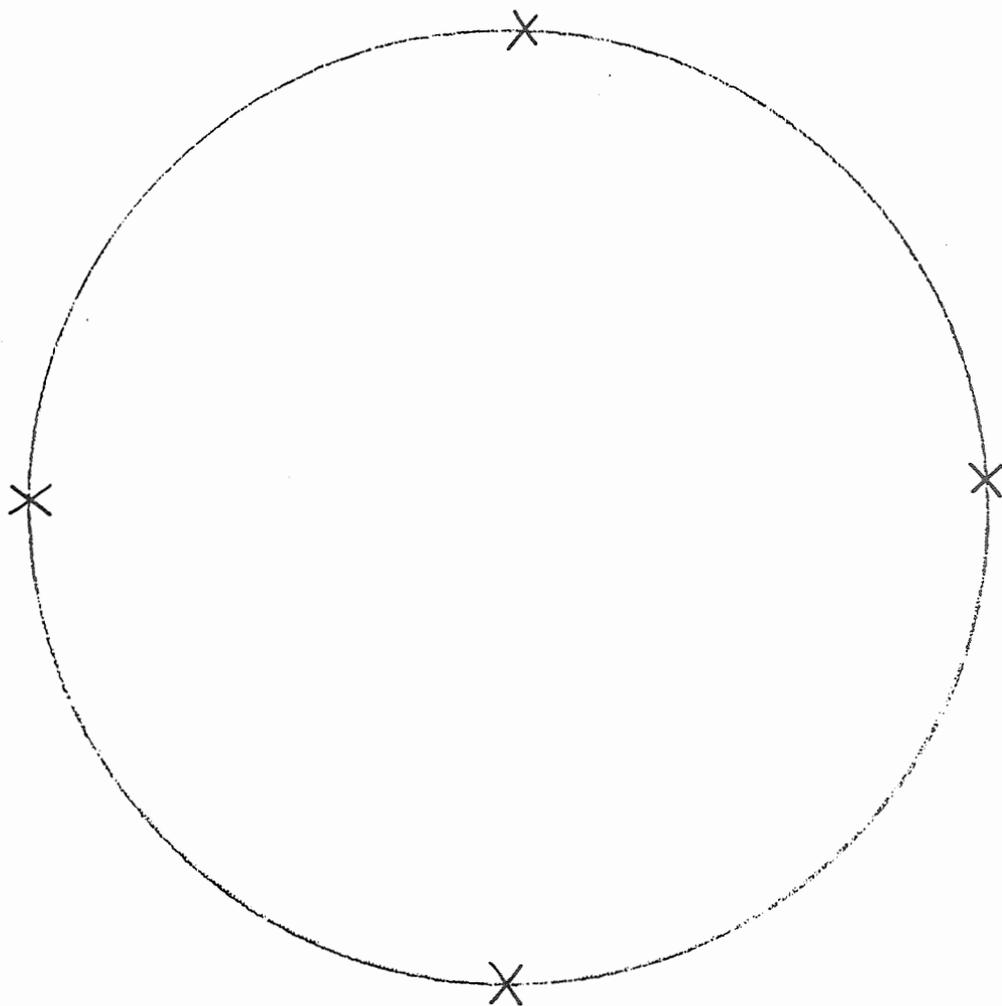
ANGULAR



6 PUNTI

cardinal

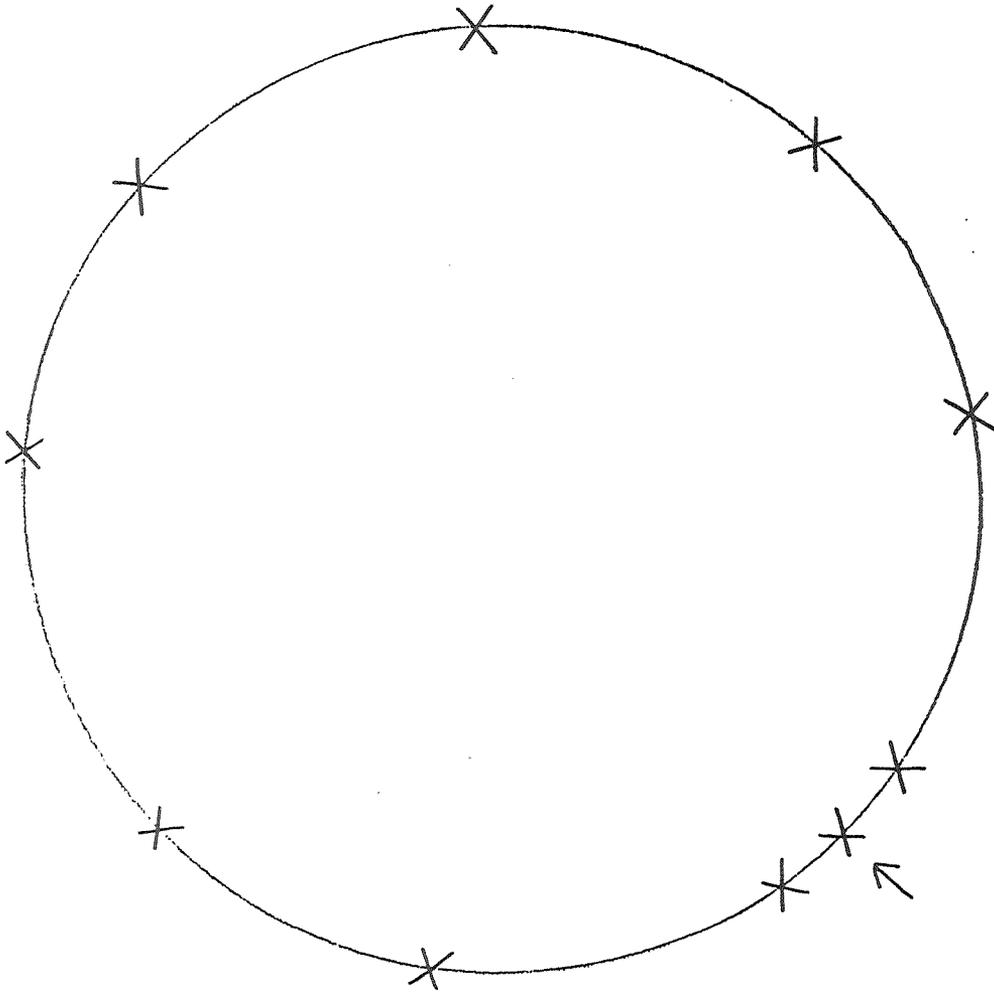
$a = 0.83$



7 punti

cubica

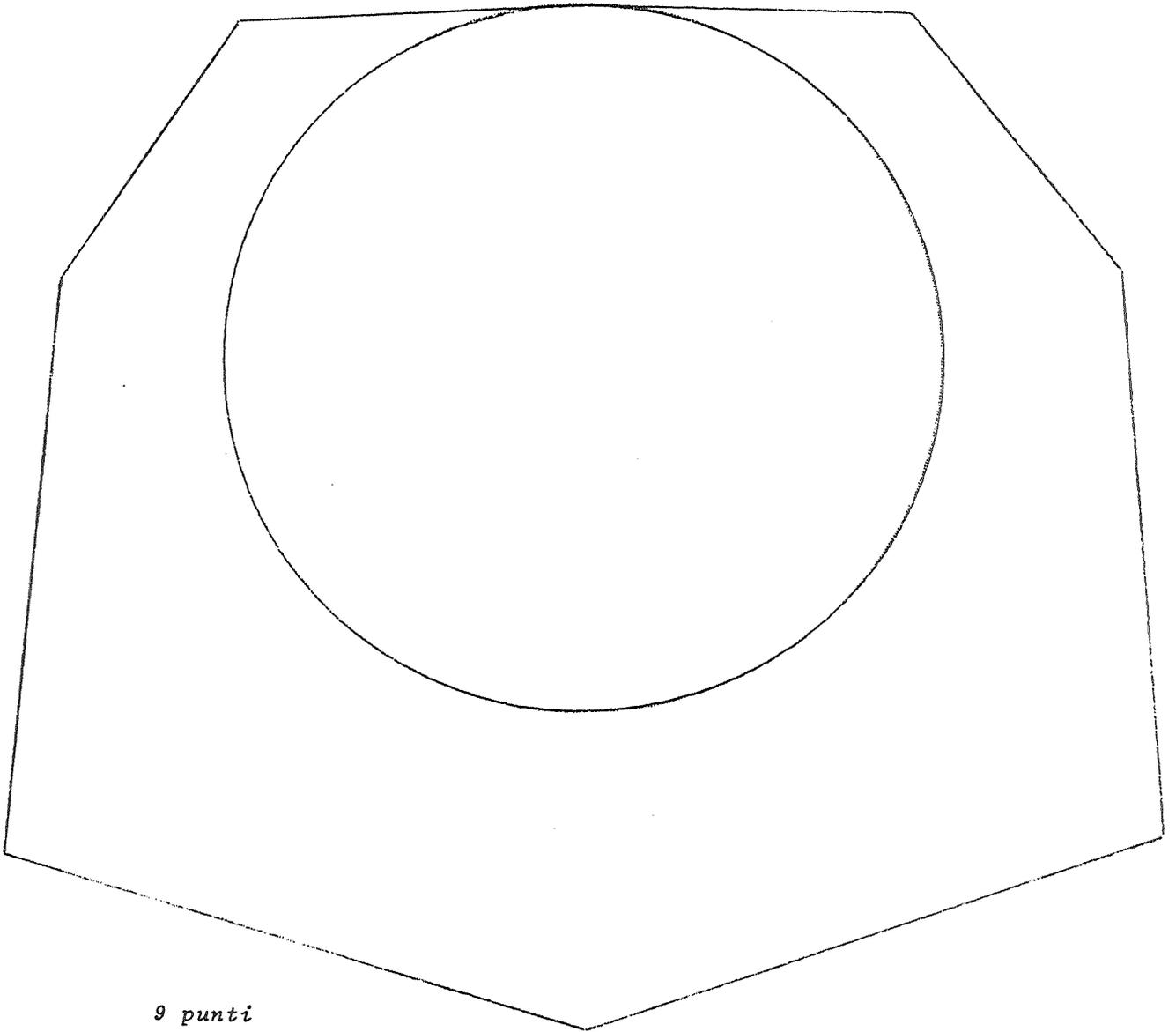
Relaxed



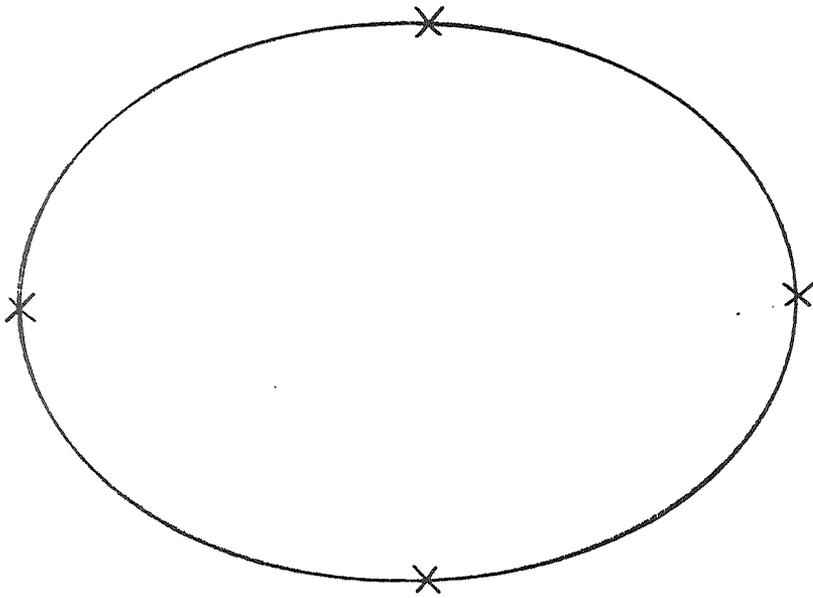
11 punti

B-spline

ord. = 9
(Bezier)



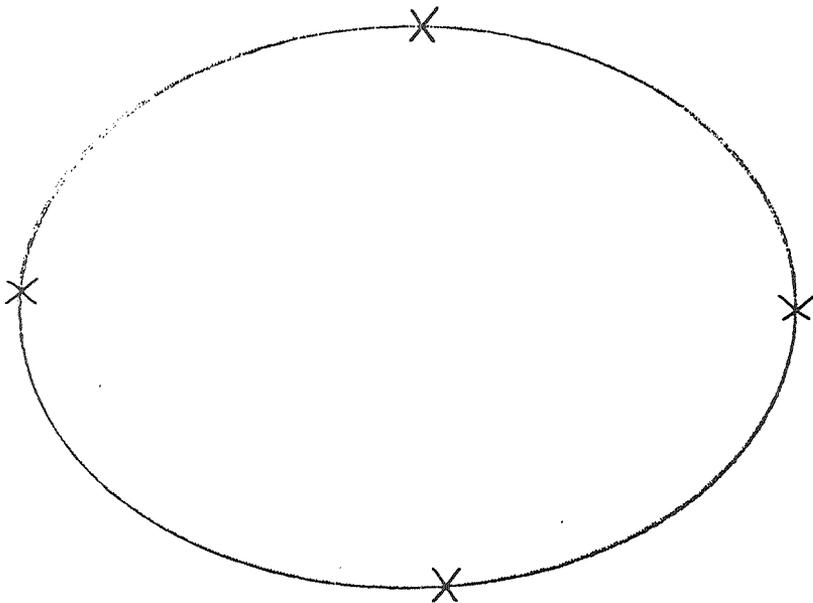
angular



5 punti

cardinal

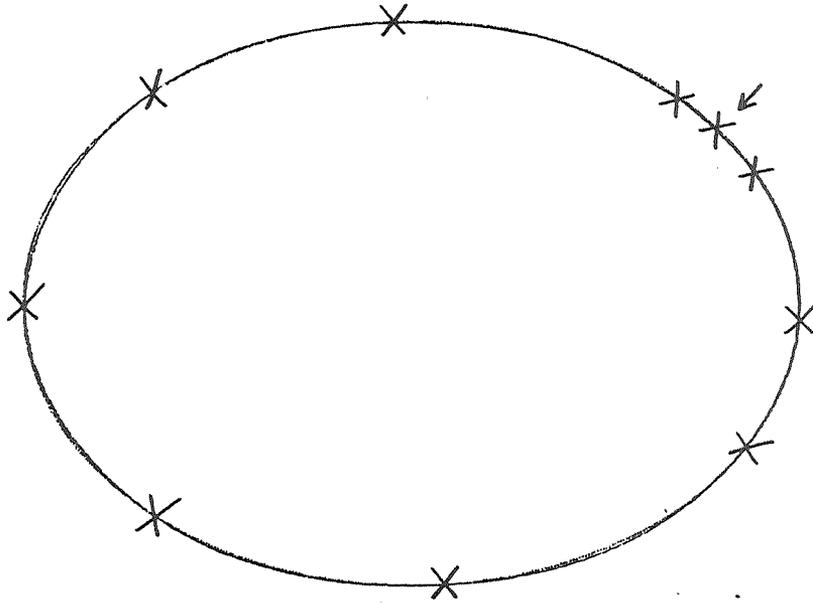
$a = 0.83$



7 punti

cubica

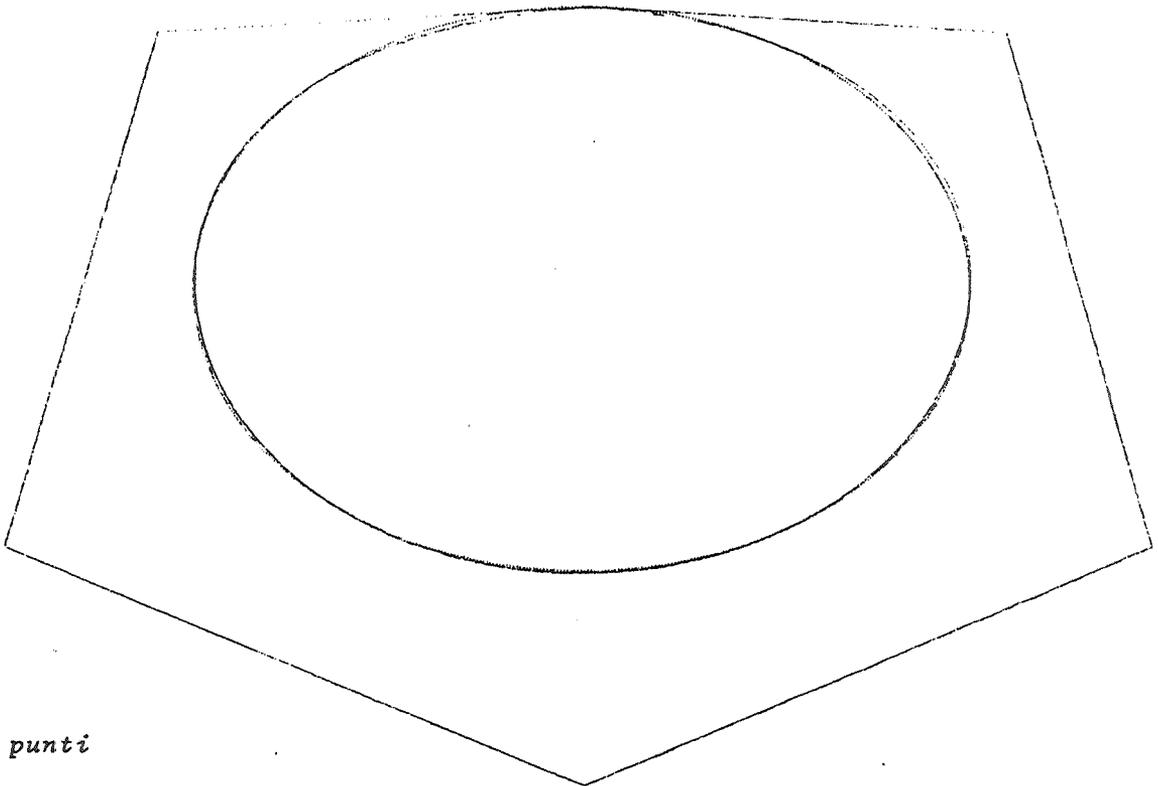
Relaxed



11 punti

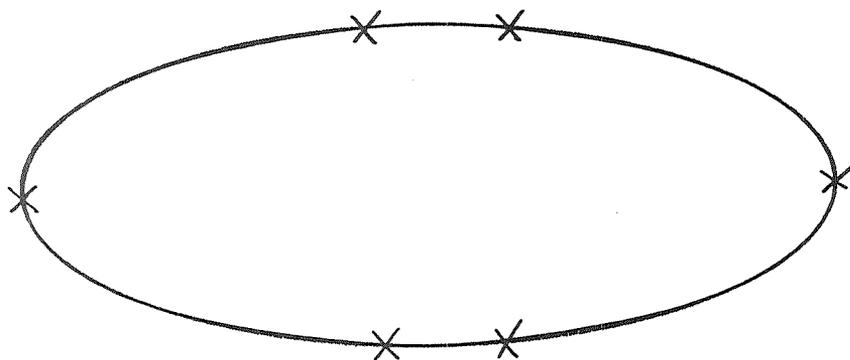
B-spline

ord. = 6



7 punti

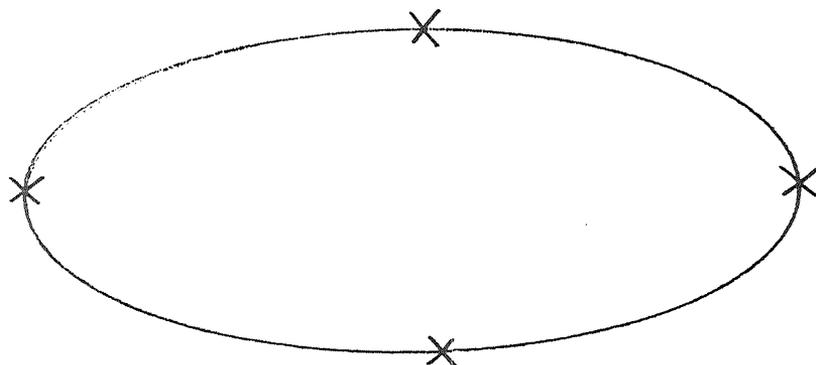
angular



9 punti

cardinal

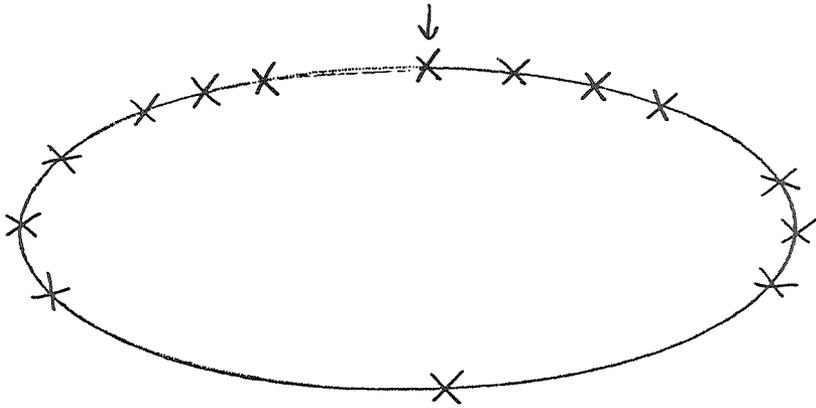
$a = 0.83$



7. punti

cubica

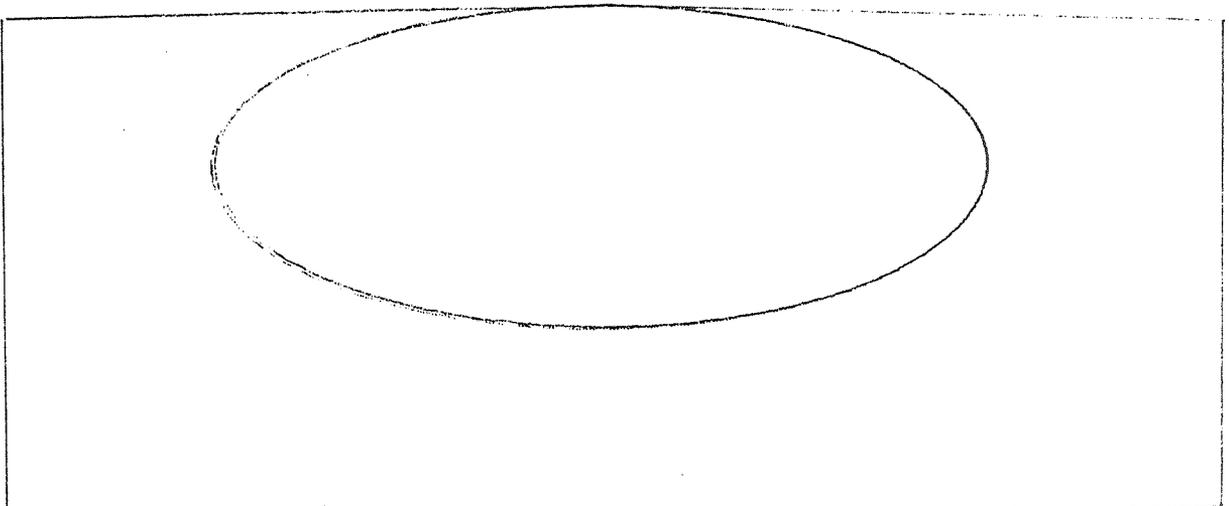
ciclica



15 punti

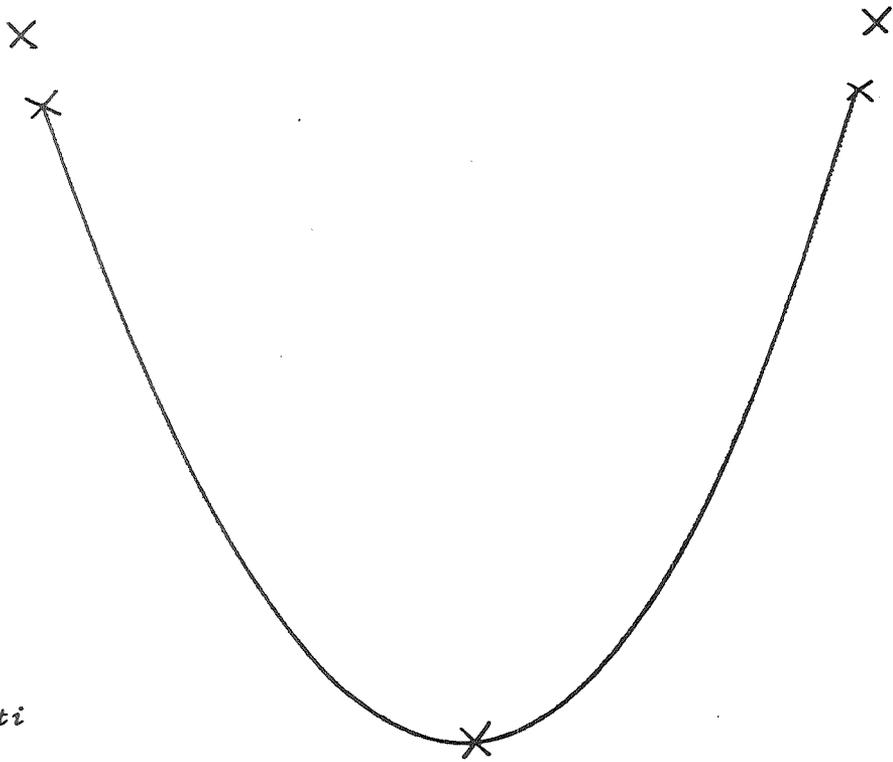
B-spline

ord. = 5



5 punti

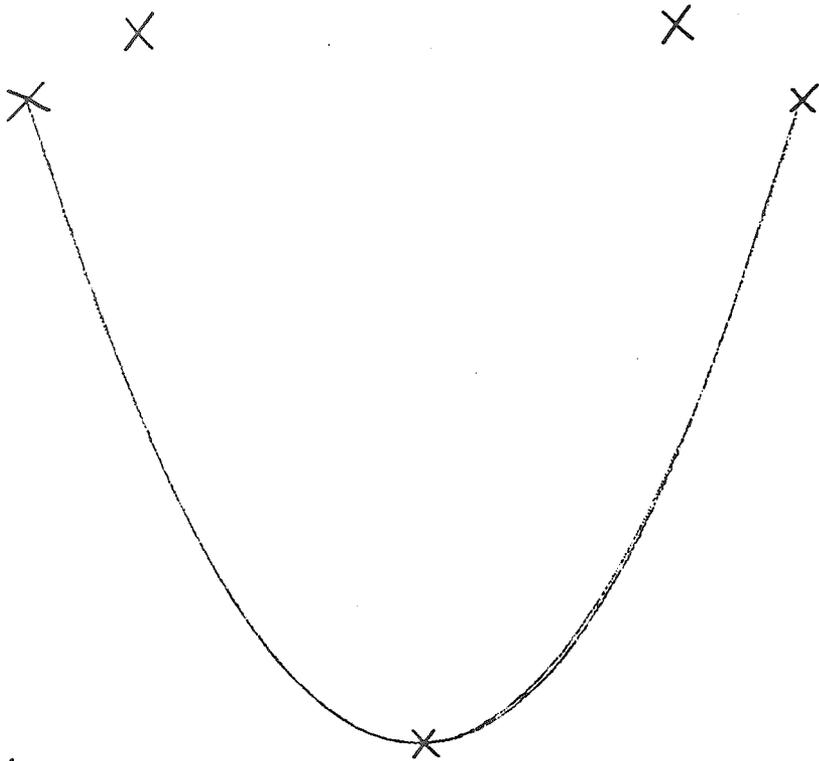
angular



5 punti

cardinal

$\alpha = 0.75$



5 punti

cubica

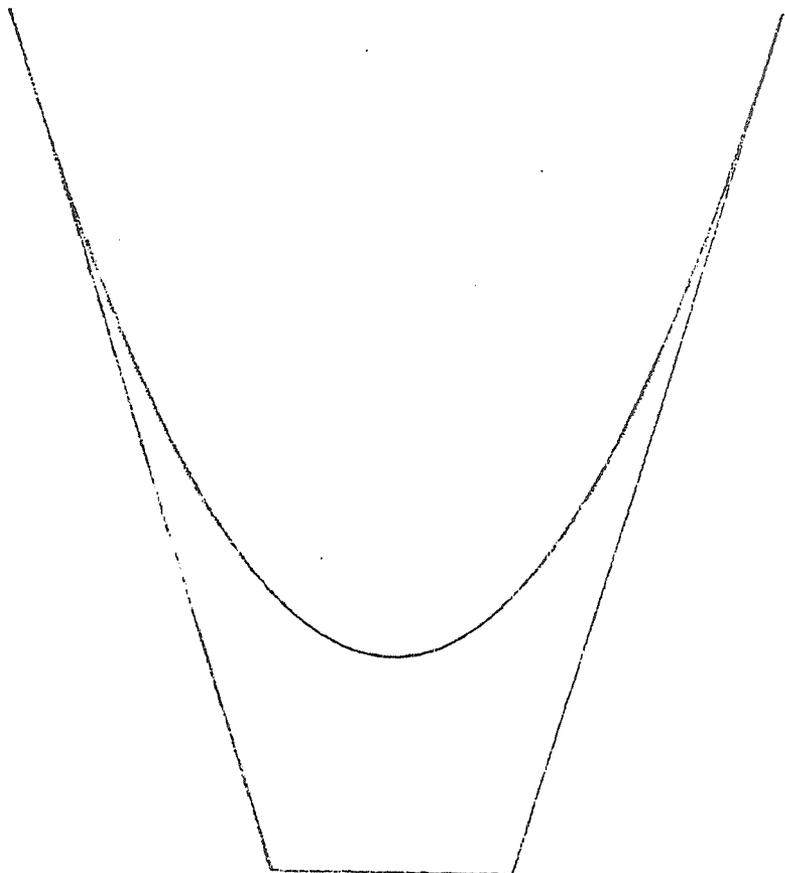
Relaxed



7 punti

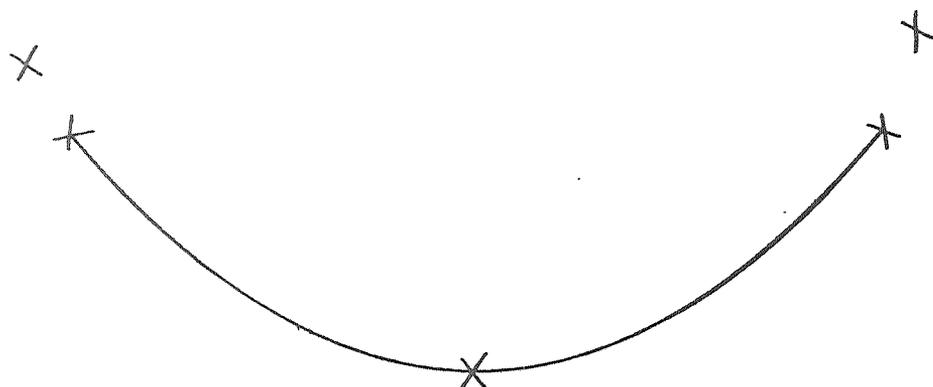
B-spline

ord. = 4



4 punti

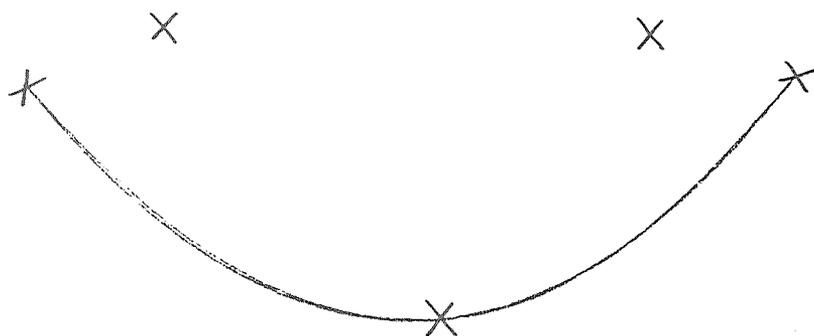
angular



5 punti

cardinal

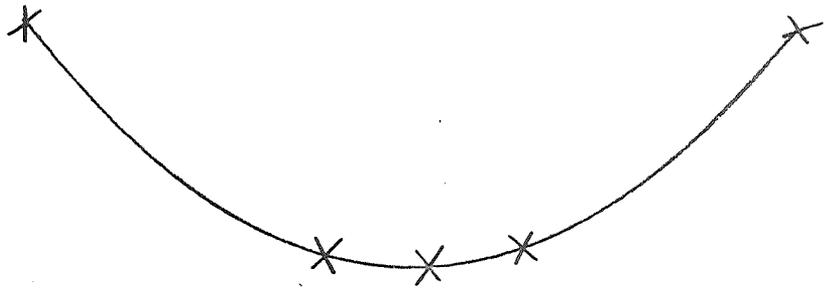
$\alpha = 0.70$



5 punti

cubica

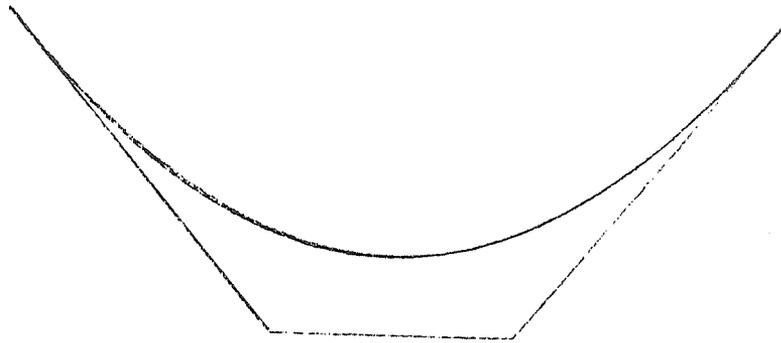
Relaxed



5 punti

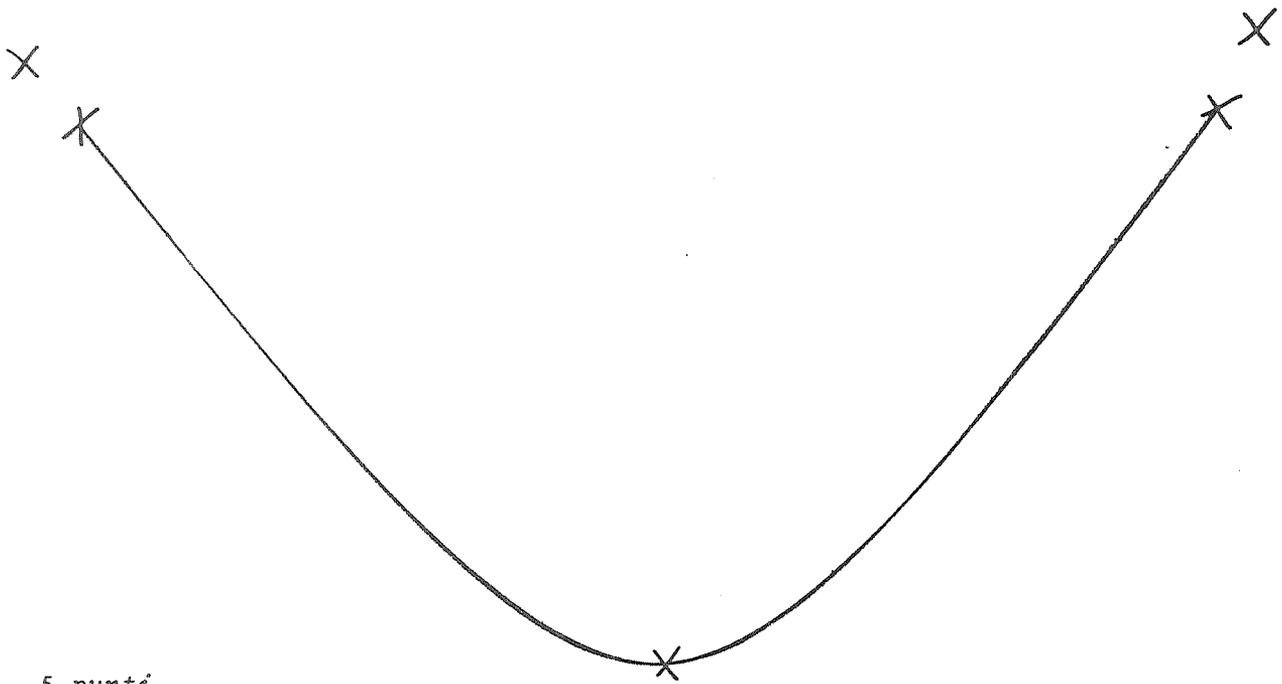
B-spline

ord. = 4



4 punti

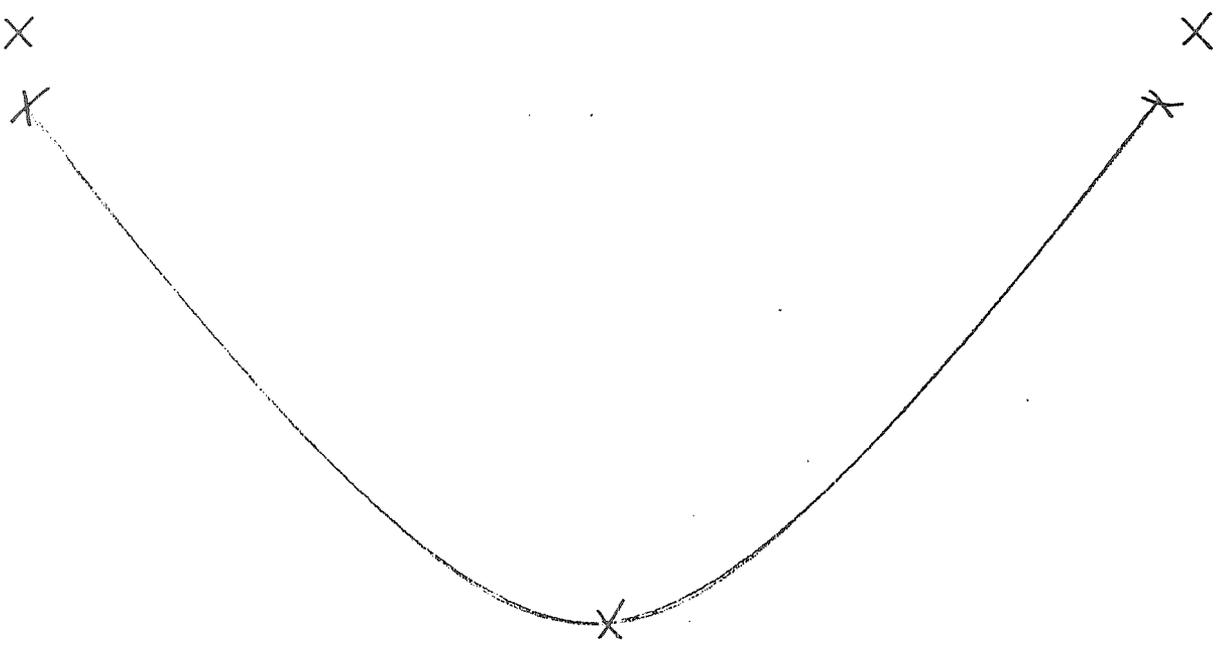
angular



5 punti

cardinal

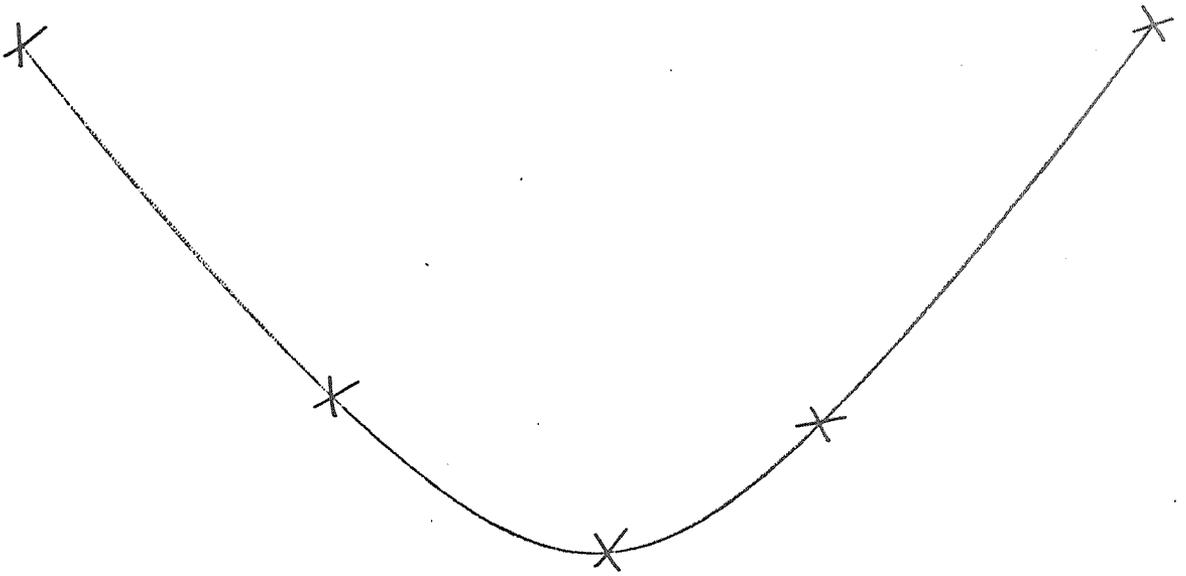
$a = 0.5$



5 punti

cubica

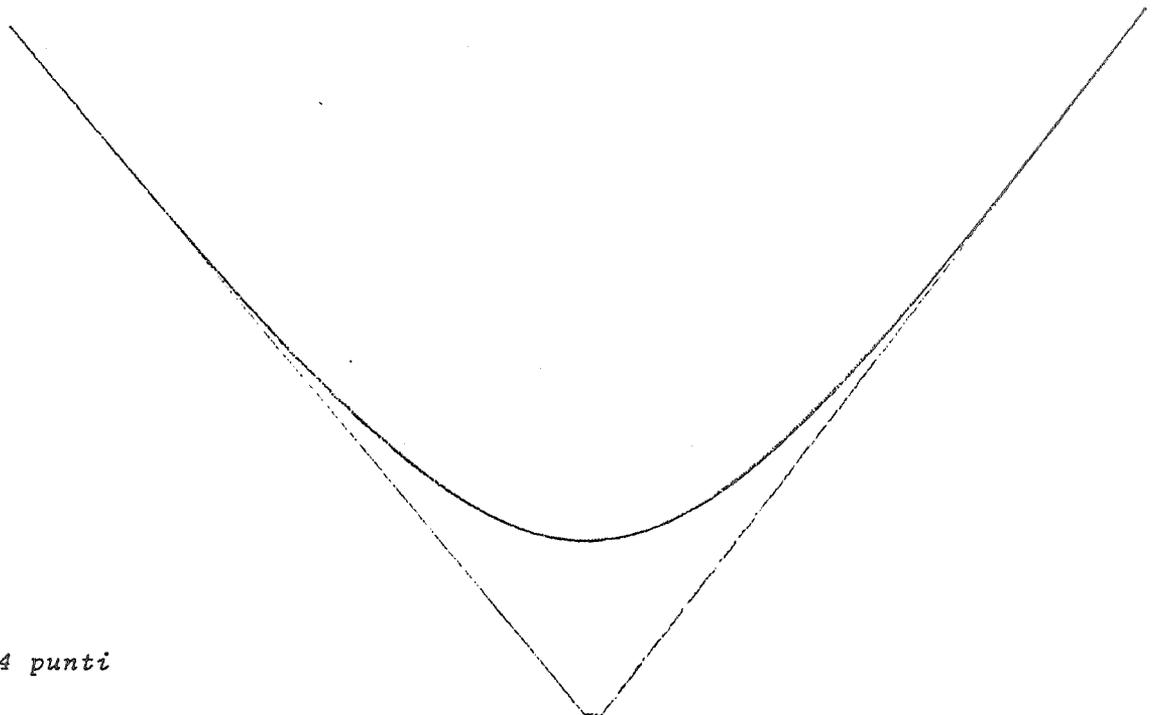
Relaxed



5 punti

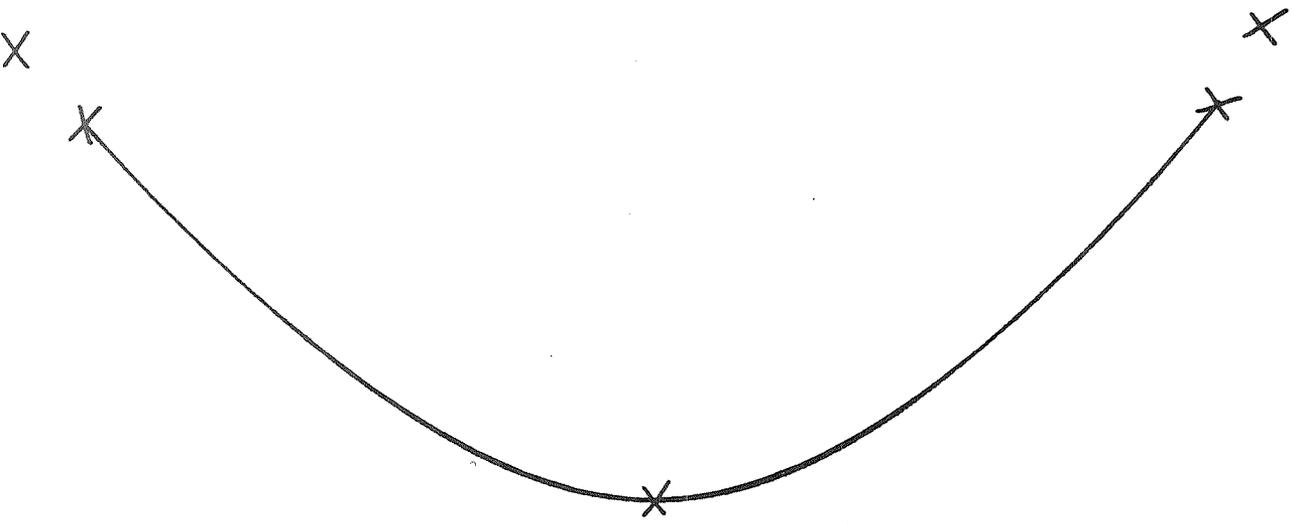
B-spline

ord. = 4



4 punti

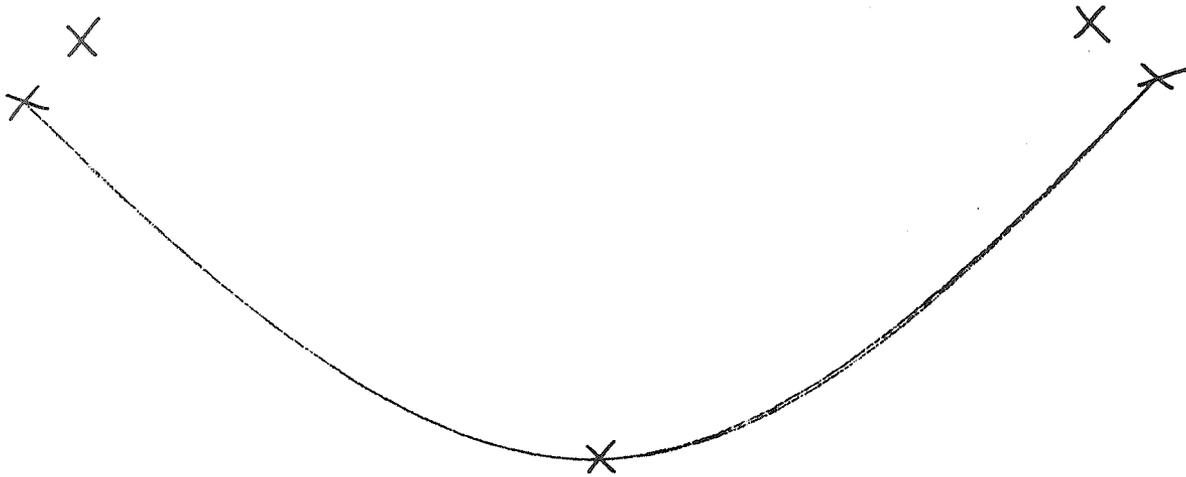
angular



5 punti

cardinal

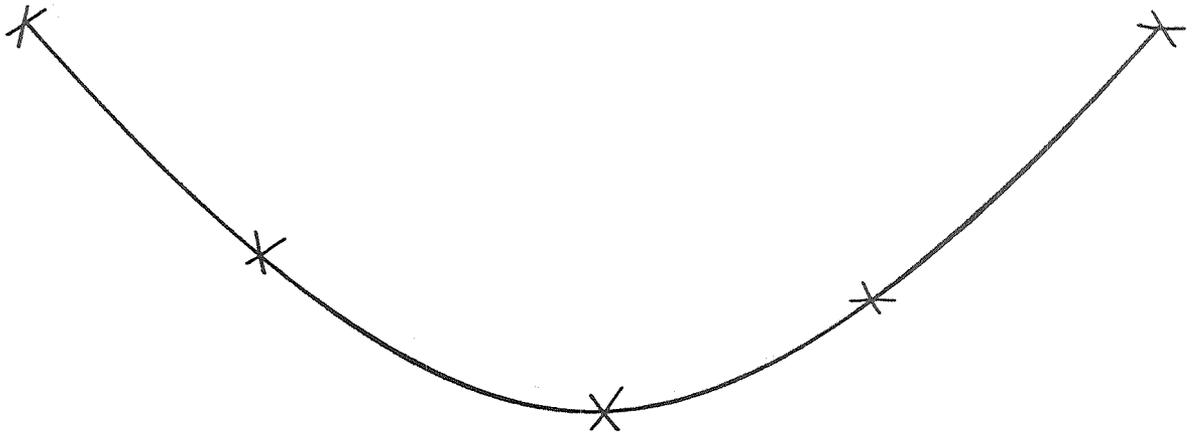
$a = 0.6$



5 punti

cubica

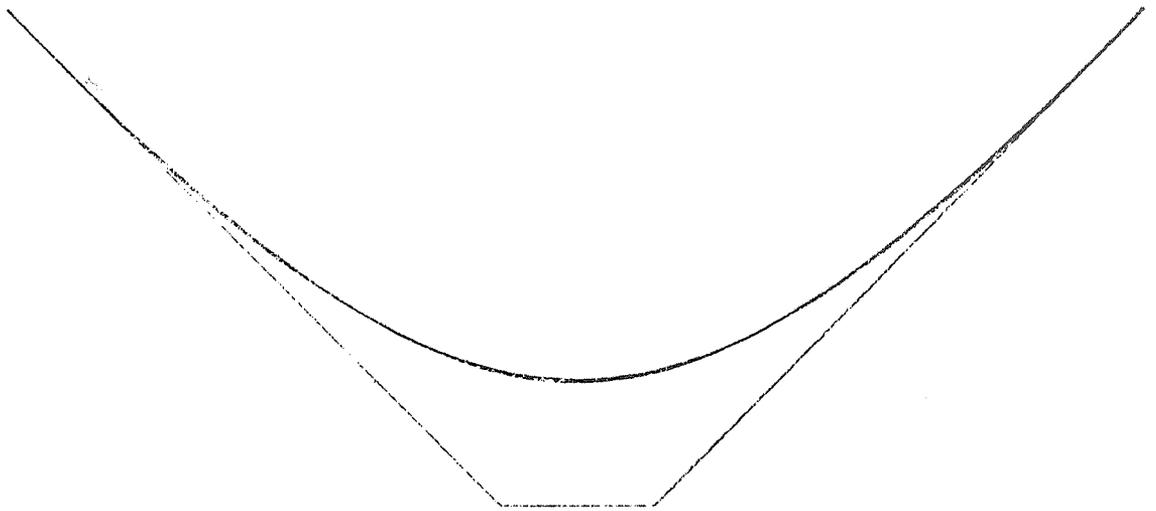
Relaxed



5 punti

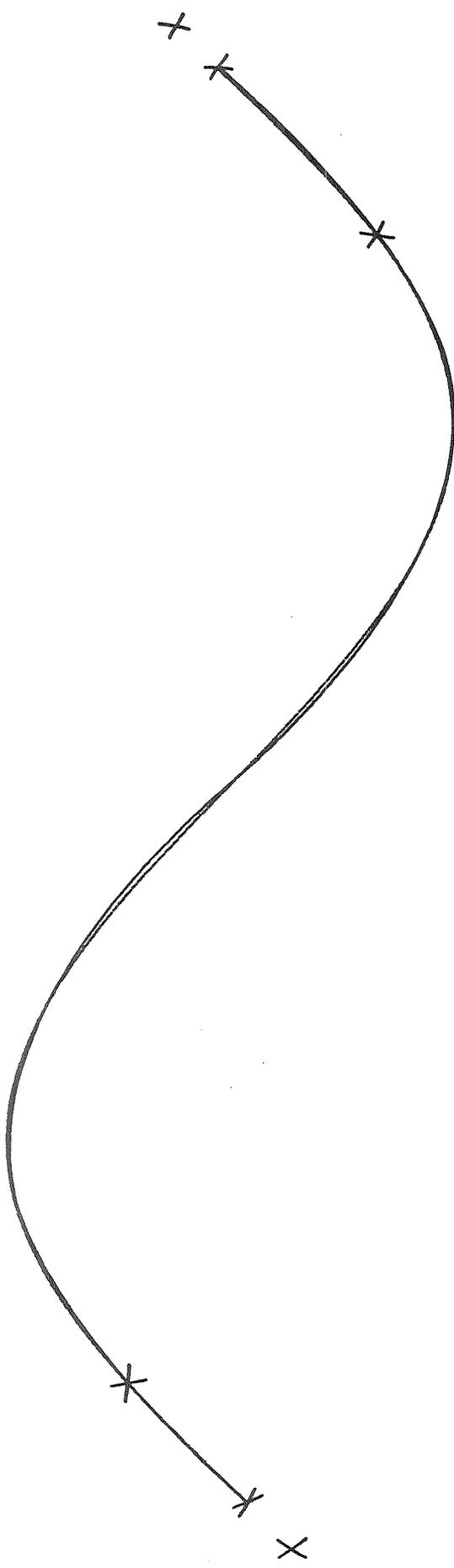
B-spline

ord. = 4



4 punti

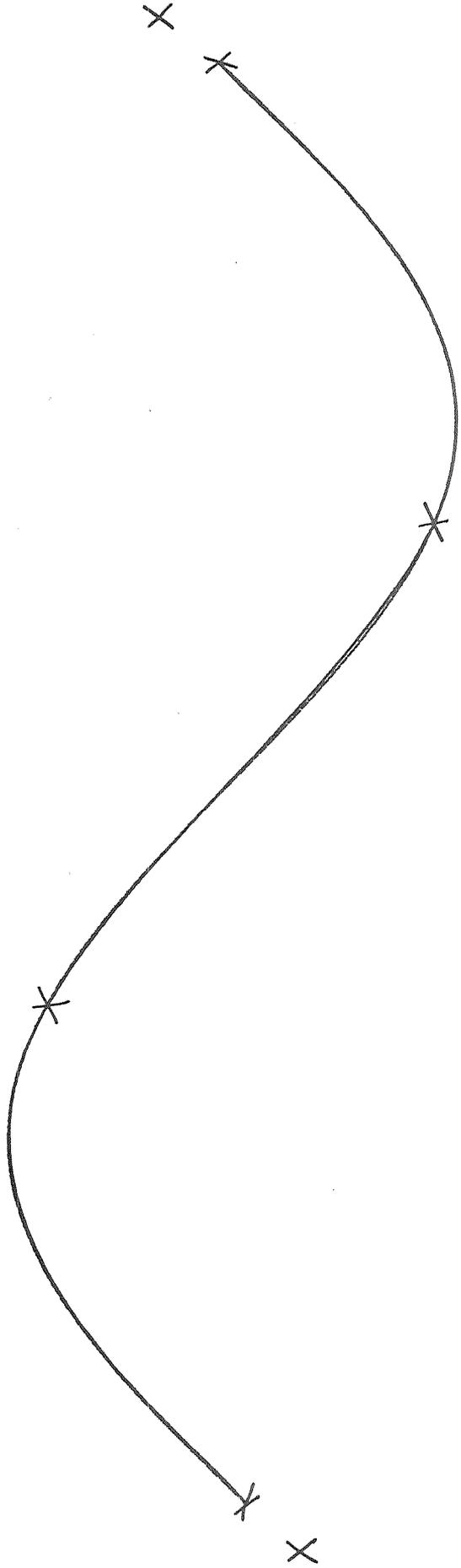
angular



6 punti

$$a = 0.55$$

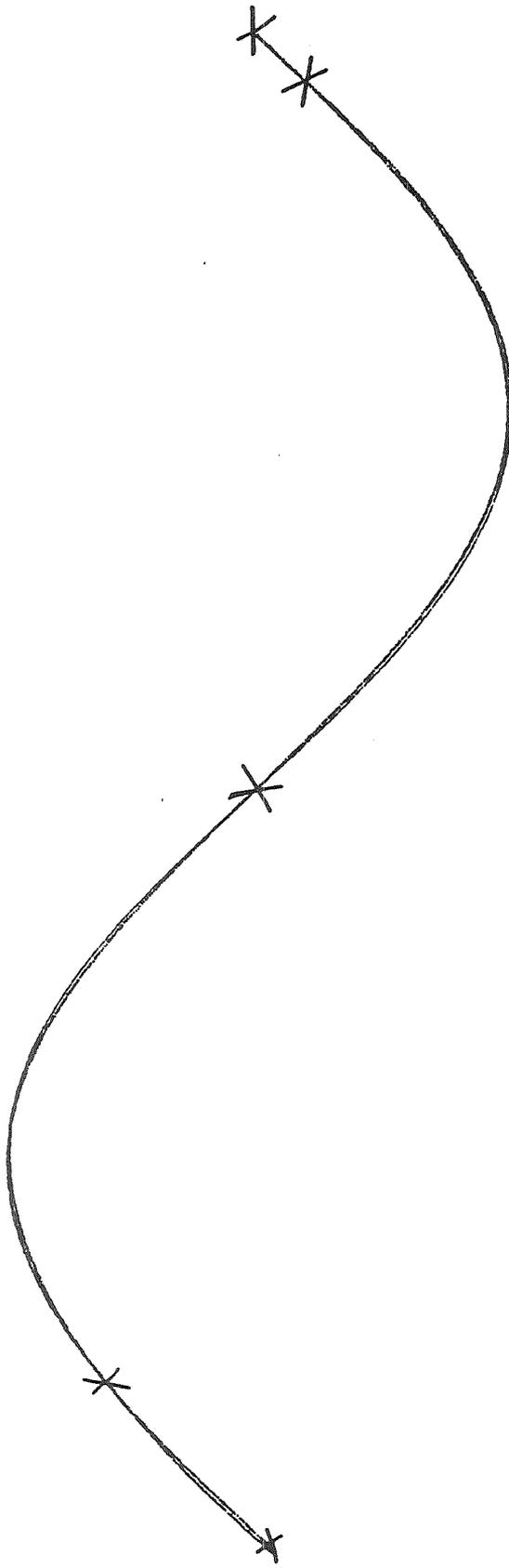
cardinal



6 punti

accolica

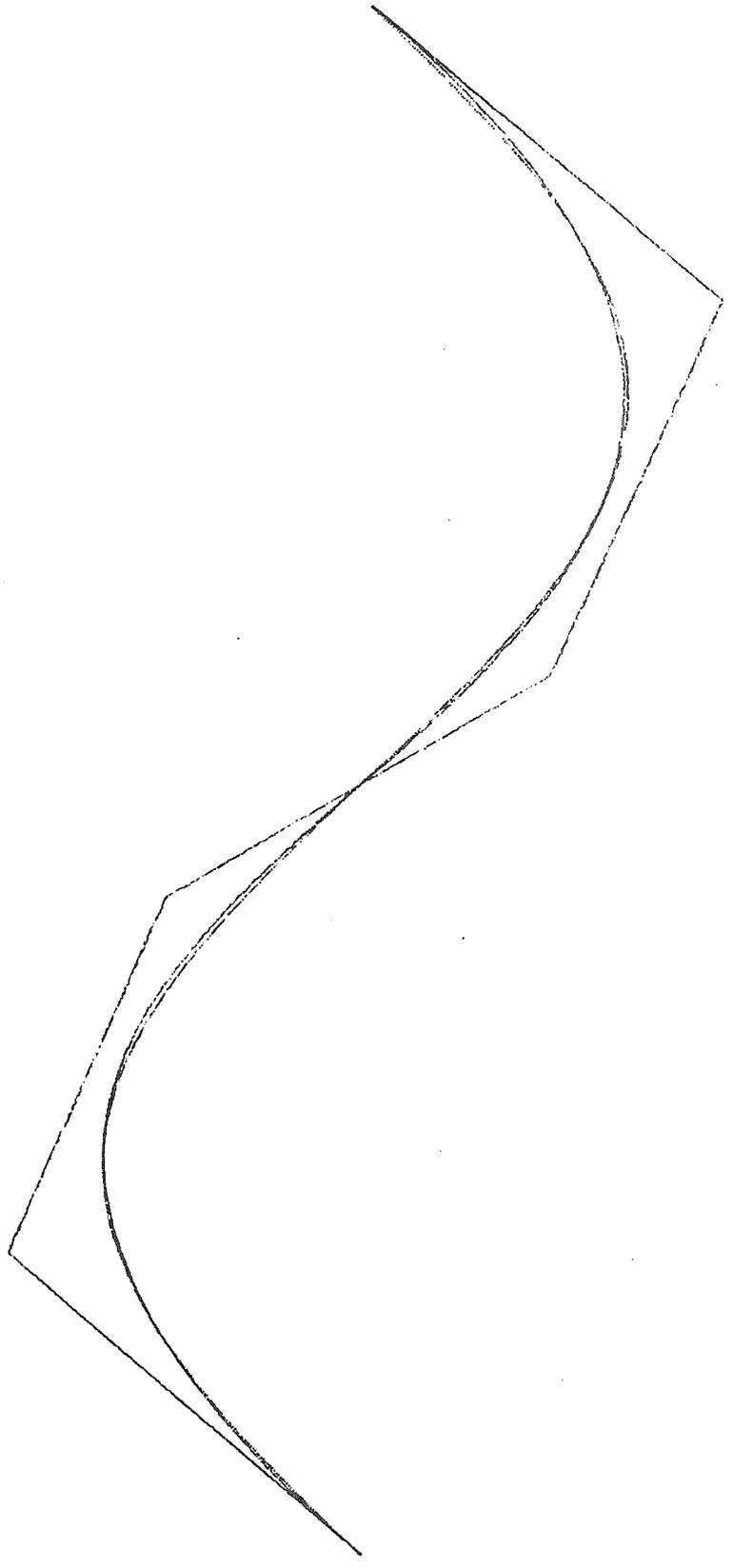
cubica



5 punti

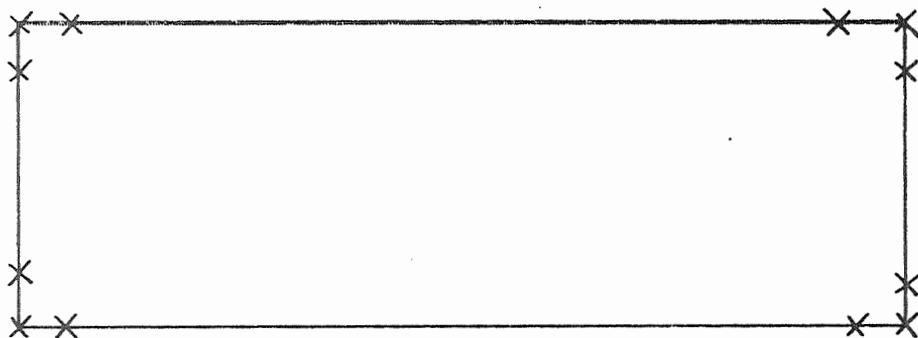
ord. = 4

B-spline



6 punti

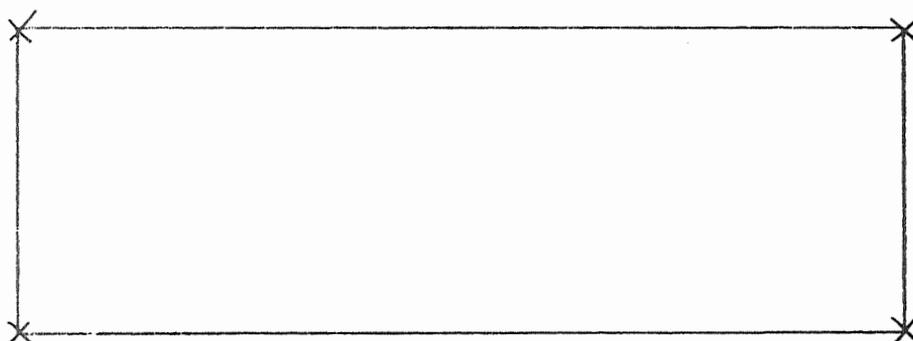
angular



15 punti

cardinal

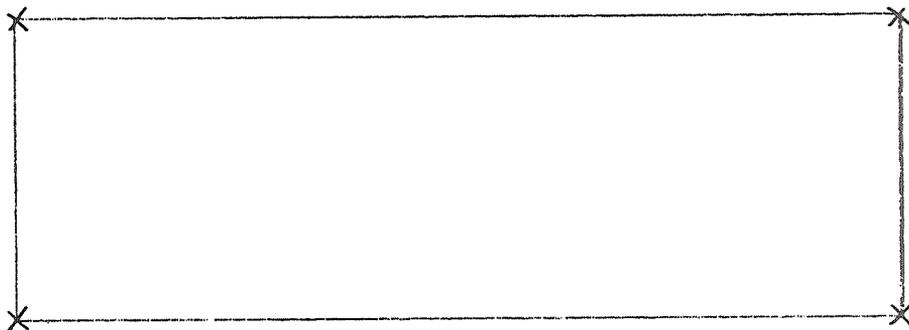
$a = 0.001$



7 punti

B-spline

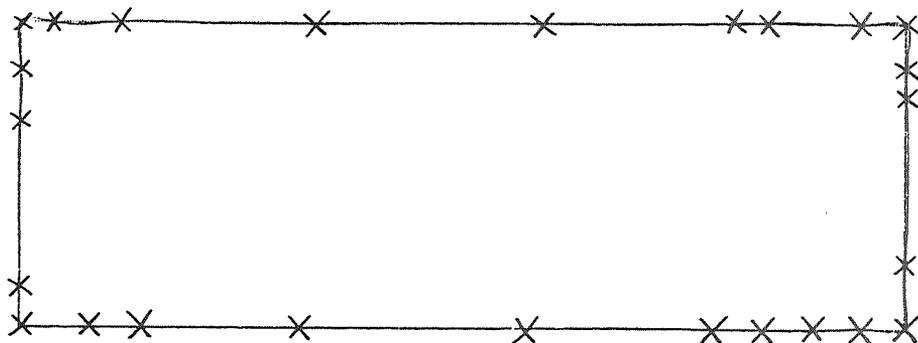
ord. = 2



5 punti

cubica

Relaxed



26 punti

conclusioni generali sul comportamento dei vari metodi nella riproduzione di figure elementari

Considerando ciascuna figura singolarmente, si puo' osservare come tutti metodi, salvo rare eccezioni, abbiano saputo riprodurre con buoni risultati i profili proposti. E' quindi partendo da tale "buon" risultato comune che tenteremo di formulare un giudizio prendendo in considerazione sia il numero di punti di interpolazione necessari al conseguimento dei risultati sia la facilita' di scelta dei punti stessi e degli eventuali altri parametri. Vediamo di riassumere in dettaglio i pregi e i difetti dei singoli metodi.

angular: si dimostra il miglior metodo per i motivi seguenti:

- richiede mediamente il minor numero di punti di interpolazione
- salvo rari casi, la scelta di tali punti e' agevole anche per l'utente non particolarmente esperto
- non richiede la definizione di ulteriori parametri di controllo. Vale la pena di ricordare che questo fatto e' particolarmente importante quando si pensi alla riproduzione di figure complesse nelle quali la necessita' di specificare parametri diversi per diversi tratti elementari porterebbe a indubbe complicazioni.

cardinal: risultati assolutamente non omogenei. Si sono riscontrate le seguenti difficolta':

- selezione del parametro onerosa; risulta difficile, anche dopo un certo uso del metodo, prevedere le variazioni della curva al variare del parametro. Cio' comporta un impiego di tempo non indifferente. Tuttavia e' possibile individuare un criterio grossolano osservando che, per valori molto bassi del parametro, l'interpolazione tende a coincidere con l'interpolazione lineare; all'aumentare del valore del parametro la curva si gonfia volgendo la concavita' verso l'interno della figura
- per la scelta dei punti possono essere fatte considerazioni analoghe alle precedenti: risulta generalmente difficoltosa e imprevedibile

per contro il numero dei punti richiesti si mantiene sempre su valori molto bassi

cubica: presenta nella stessa misura pregi e difetti.

- necessita di una specifica di tipo (equivalente alla specifica di un parametro) ma, poiche' e' la versione relaxed che quasi sempre si dimostra la migliore, la specifica precedente risulta implicita.
- e' il metodo che mediamente richiede il numero maggiore di punti di interpolazione, ma fornisce spesso eccellente qualita' di riproduzione
- risulta totalmente inefficace nel riprodurre forme poligonali, ma riproduce in maniera eccellente situazioni di flesso

B-spline: si tratta di un metodo per molti versi non omogeneo con i precedenti; infatti genera una curva di interpolazione che passa solo per il primo e l'ultimo dei punti dell'insieme di interpolazione. La letteratura pero' lo annovera fra i metodi di interpolazione anche perche' non lo si puo' considerare un metodo di approssimazione: i punti di interpolazione "guidano" l'andamento della curva. Se lo scopo che ci si prefigge e' quello di generare figure, piuttosto che riprodurre figure passanti per determinati punti, si tratta indubbiamente di un buon metodo. Nessuna delle figure in esame ha presentato particolari difficolta' di riproduzione: la cosa piu' difficile da imparare e' stato il "vedere" il poligono di definizione che "pilota" la curva nel suo svilupparsi. Anche la scelta del parametro (ordine della curva) e' stata sempre molto meno gravosa che negli altri metodi e soprattutto si e' presentata come facilmente intuibile.

specificita' della scelta dei punti di interpolazione: un'ulteriore indagine

Con riferimento ai metodi che interpolano precisamente i punti dati, e' stata svolta una verifica della effettiva specificita' degli insiemi dei punti di interpolazione su cui operano i singoli metodi. In altri termini, un insieme di punti $P_{i,j}$, ottimale per il metodo M_j , e' stato usato come insieme di interpolazione per un differente metodo M_k . E' immediato osservare che la

situazione ideale dal punto di vista operativo sarebbe quella della "non specificita'" degli insiemi di punti; metodi diversi applicati ad un identico insieme porterebbero in tal caso a differenze secondarie tra le curve risultanti. Come gia' e' intuitivo dalle figure fin qui mostrate, la situazione reale e' ben diversa e, se non vengono fatte scelte di punti particolarmente appropriate, si rischia di ottenere forme addirittura assurde come risultato dell'interpolazione. E' sembrato interessante documentare con esempi queste osservazioni, anche per verificare reali affinita' di comportamento tra i vari metodi. Giova precisare che il test e' stato effettuato in maniera esaustiva considerando, per ogni figura F_i ed ogni metodo M_j il rispettivo insieme ottimale di punti P_{ij} e quindi applicando a P_{ij} metodi differenti M_k . Le figure 6, 7, 8 mostrano alcuni degli esempi piu' significativi mentre le considerazioni che seguono si riferiscono all'insieme completo dei test effettuati.

Vediamo in maggior dettaglio.

- punti ottimali per il metodo angular: interpolando con il metodo cardinal si hanno gli stessi risultati con le parabole e le iperboli e risultati abbastanza vicini negli altri casi (il cerchio, Fig.6, costituisce un caso a se'); interpolando con il metodo cubica i risultati sono invece tutt'altro che soddisfacenti (solo per la sinusoida si hanno risultati accettabili)
- punti ottimali per il metodo cardinal: interpolando con il metodo angular si ottengono ottimi risultati tranne che nel caso del rettangolo e della sinusoida; interpolando con il metodo cubica si hanno, analogamente a quanto succedeva nel caso precedente, risultati decisamente mediocri
- punti ottimali per il metodo cubica: interpolando con il metodo angular si hanno buoni risultati (eccettuata la sinusoida); interpolando con il metodo cardinal le cose peggiorano soprattutto per la presenza di oscillazioni indesiderate

Riproduzione di caratteri tipografici

La riproduzione di caratteri tipografici, oltre a rappresentare un'interessante ed utile applicazione dei metodi di interpolazione, e' stata un valido banco di

prova per i metodi presi in esame, poiche' i profili in questione non hanno piu' la semplicita' delle figure esaminate in precedenza . Come esempio specifico e' stata scelta la lettera "a" in varie versioni (figure 9-20).

Come si puo' notare, le varie versioni possiedono profili dall'andamento estremamente vario: flessi, cuspidi, chiocciole, segmenti, spigoli sono spesso coesistenti nella stessa figura.

Esaminiamo brevemente, a commento delle figure, il comportamento dei tre metodi "omogenei" angular, cardinal e cubica nella riproduzione. Tali figure riportano i profili originari ed i punti di interpolazione (a), le curve risultanti sovrapposte agli originali (b) e infine i risultati da soli per consentire una migliore valutazione (c).

Metodo Angular

Come prima osservazione si puo' dire che il metodo Angular e' risultato decisamente il piu' facile da usare ed ha mediamente richiesto il minor numero di punti di interpolazione. E' facile verificare come i risultati siano stati sempre eccellenti.

Metodo Cardinal

Un discorso totalmente diverso vale per la Cardinal : come era prevedibile questo metodo ha presentato nella riproduzione notevoli difficolta' ed ha richiesto un numero di punti di interpolazione non indifferente. Salvo rare eccezioni, i risultati sono stati i peggiori in senso assoluto.

Metodo Cubica

L'interpolazione con la Cubica ha fornito risultati decorosi con un numero di punti di interpolazione non eccessivo e senza troppe difficolta' operative.

A commento finale si possono trarre le conclusioni seguenti: il metodo Angular si e' dimostrato superiore come risultati e come facilita' d'uso rispetto agli altri metodi; il metodo Cardinal ha trovato non poche difficolta' mentre il metodo cubica non e' certamente emerso.

Riproduzione di un profilo geografico .

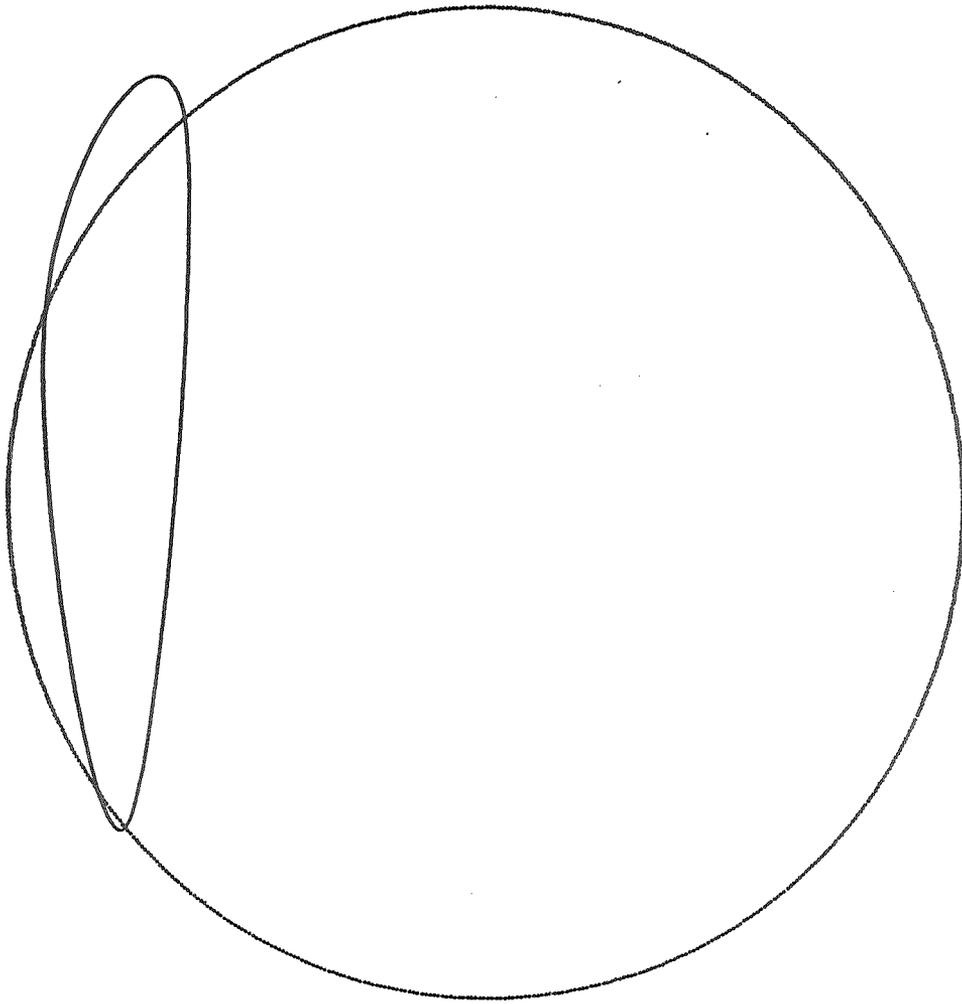
Come esempio conclusivo, verra' presentata la riproduzione di un profilo geografico (costa occidentale della Sicilia) .

La riproduzione di profili costieri e' tutt'oggi motivo di studi e ricerche: in particolare sarebbe desiderabile la capacita' di riprodurre questi profili con la piu' alta precisione possibile, senza peraltro dover spendere tempo eccessivo per l'acquisizione dei punti .

L'acquisizione automatica dei dati (ovvero la possibilita' di selezionare automaticamente i punti di interpolazione ottimali per un certo metodo) risolverebbe gran parte del problema : purtroppo in questo campo siamo ancora agli inizi e, come e' facile intuire, il problema non e' affatto di semplice soluzione .

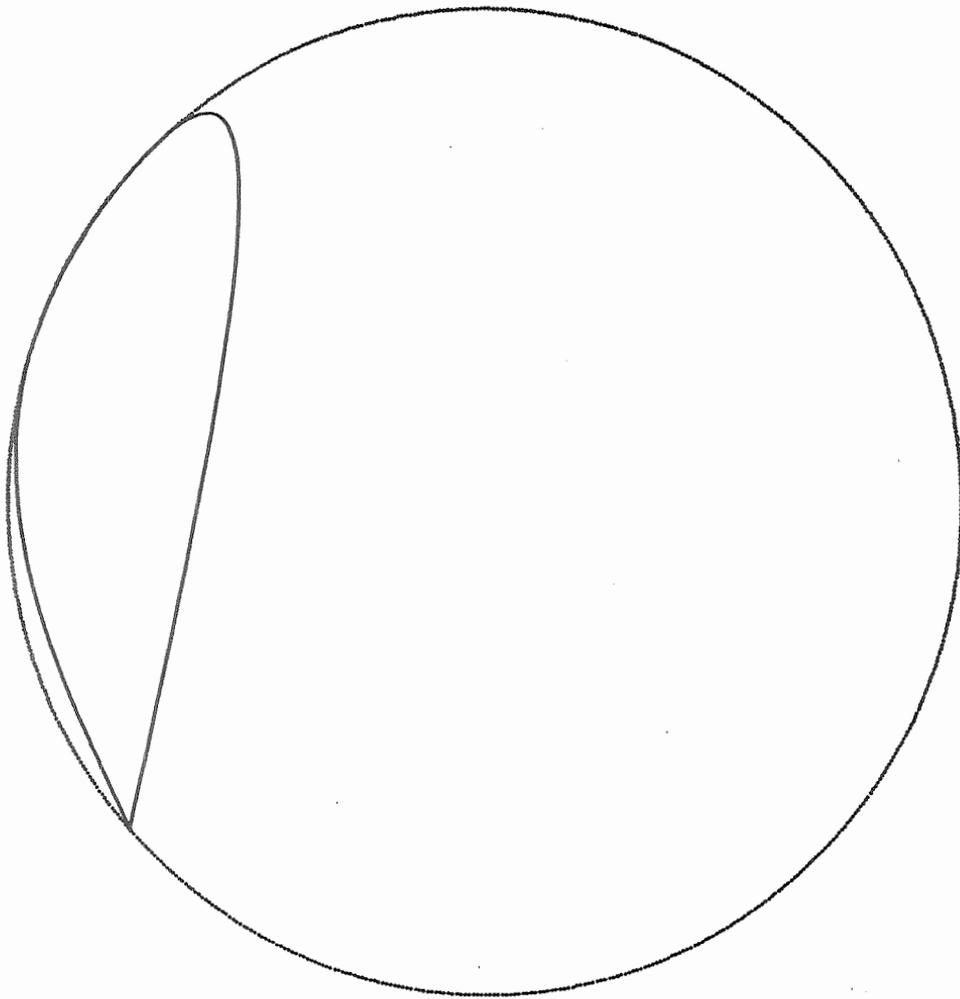
Tornando all'esempio presentato, va detto che il profilo della Sicilia e' stato acquisito attraverso una tavola digitalizzatrice (per un totale di circa 2500 punti).

Come primo tentativo, e' stata eseguita una prova abbastanza grossolana : l'insieme di punti da interpolare e' stato generato prelevando un punto ogni "n" punti dell'originale . In questa sede riporteremo solo i risultati ottenuti con $n=50$. Come era da attendersi, nessun metodo di interpolazione, inclusa la semplice interpolazione lineare, mostra una qualsiasi superiorita'. E' evidente infatti come sia impossibile trovare una qualsiasi "sintonia" fra profili casuali e metodi di interpolazione "morbida".



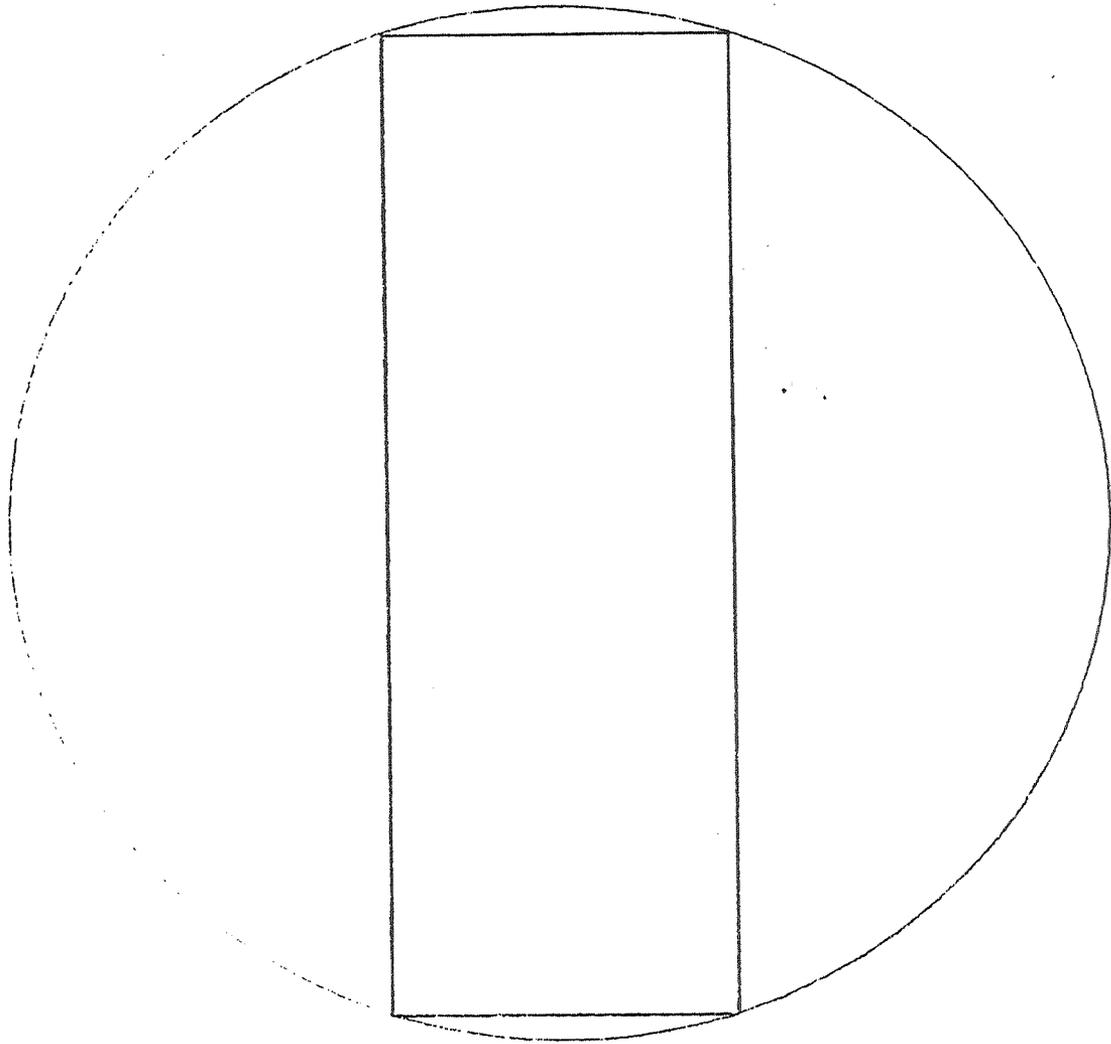
Punti ottimali per metodo : Angular
interpolati con il metodo : Cardinal ($a = 0.9$)

Fig. 6-a



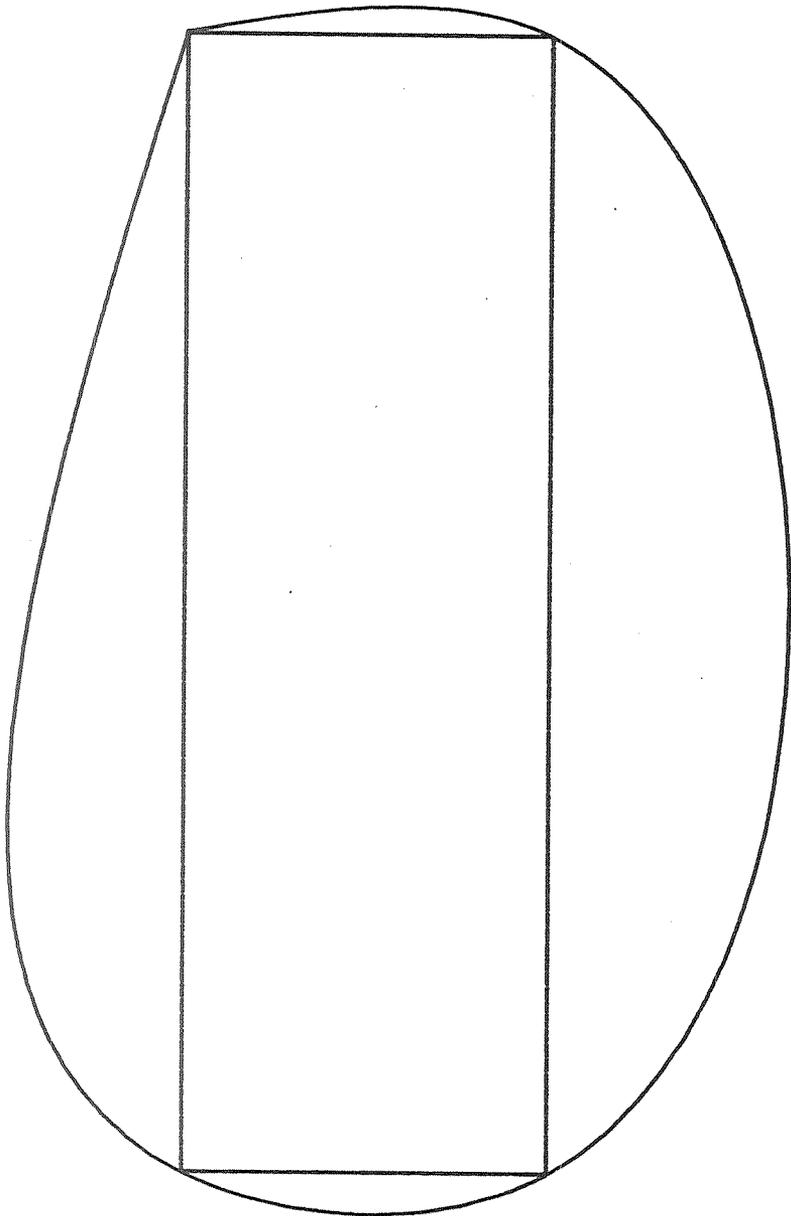
Punti ottimali per metodo : Angular
interpolati con il metodo : Cubica Relaxed

Fig. 6 - b



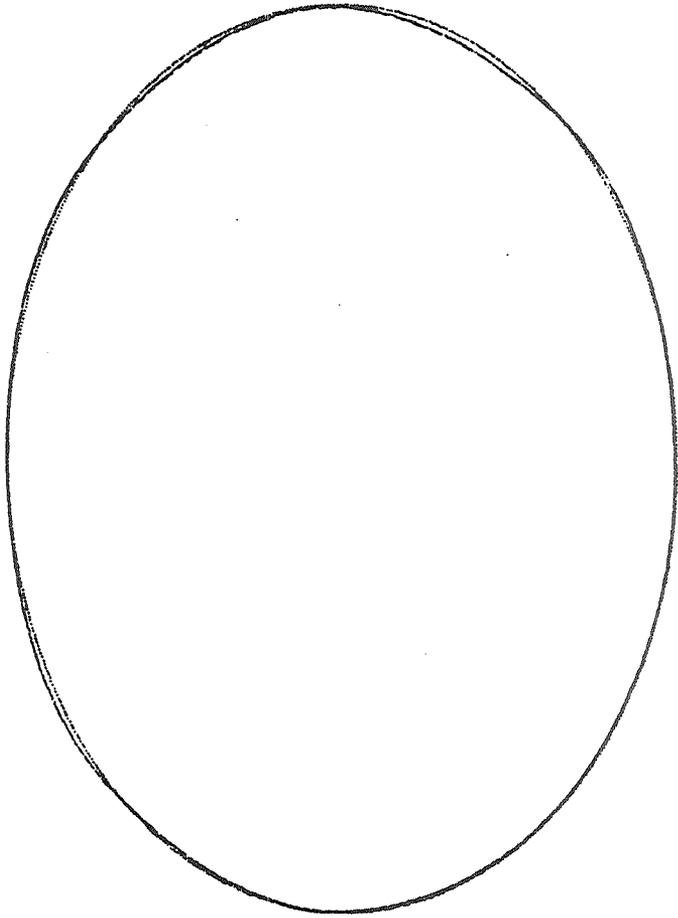
Punti ottimali per metodo : Cardinal
interpolati con il metodo: Angular

Fig. 7 - a



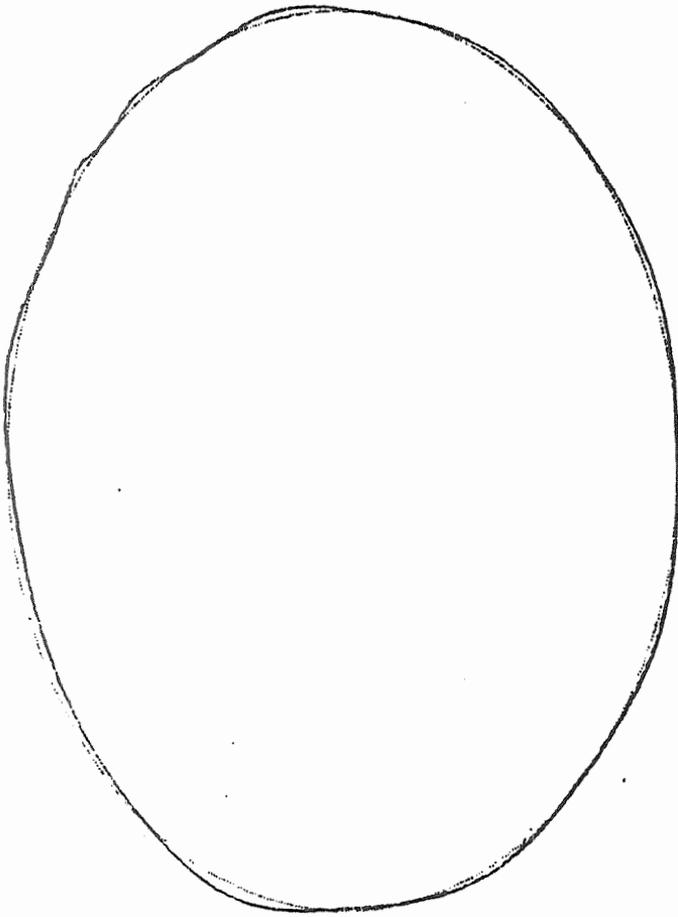
Punti ottimali per metodo : Cardinal
interpolati con il metodo : Cubica Relaxed

Fig. 7 - b



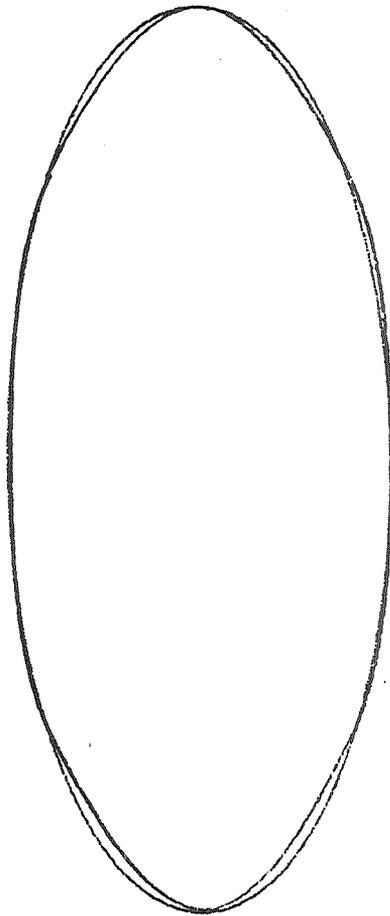
Punti ottimali per metodo : Cubica Relaxed
interpolati con il metodo : Angular

Fig. 8.- a



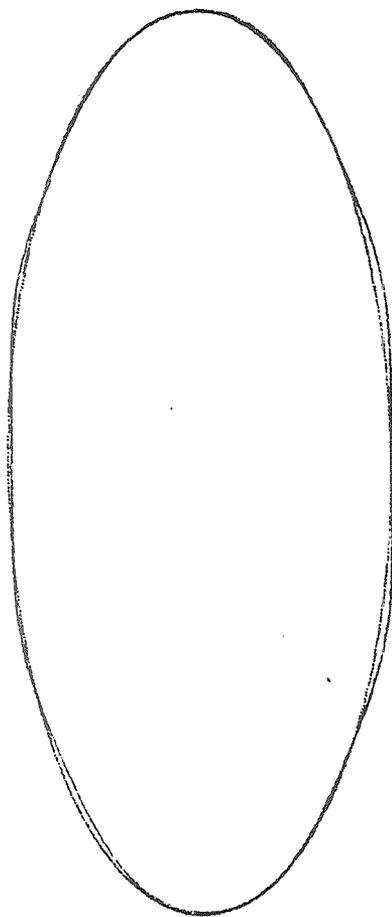
Punti ottimali per metodo : Cubica Relaxed
interpolati con il metodo : Cardinal ($a = 0.7$)

Fig. 8 - b



Punti ottimali per metodo : Cubica Relaxed
interpolati con il metodo : Angular

Fig. 8 - c



Punti ottimali per metodo : Cubica Relaxed
interpolati con il metodo : Cardinal ($a = 0.7$)

Fig. 8 - d

Metodo : Angular

Punti : 19

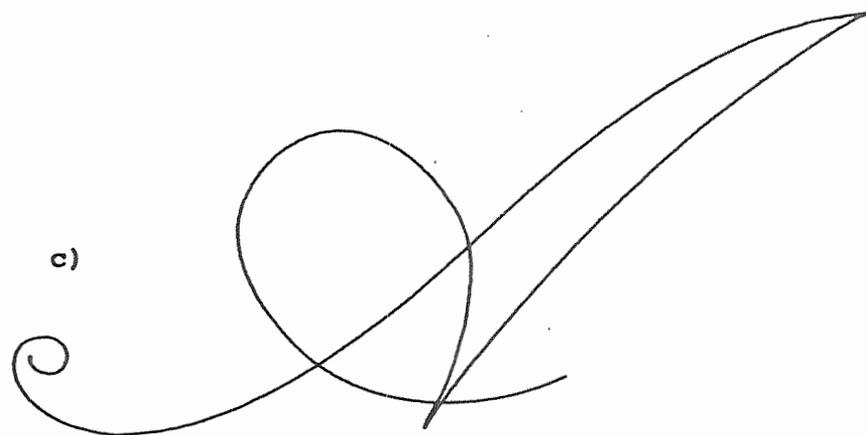
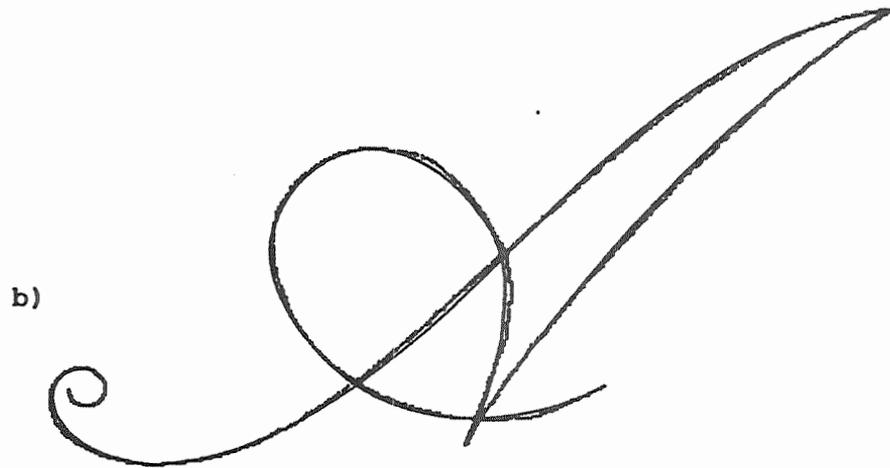
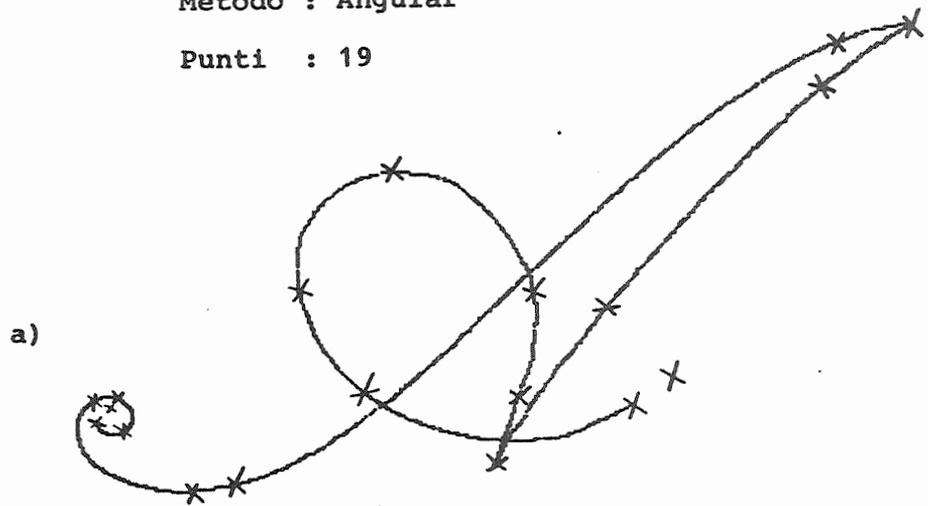


Fig. 9

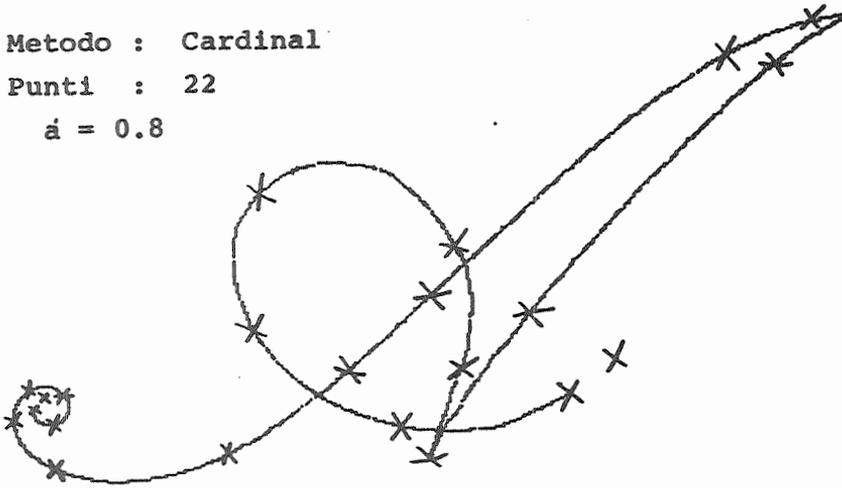
- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

Metodo : Cardinal

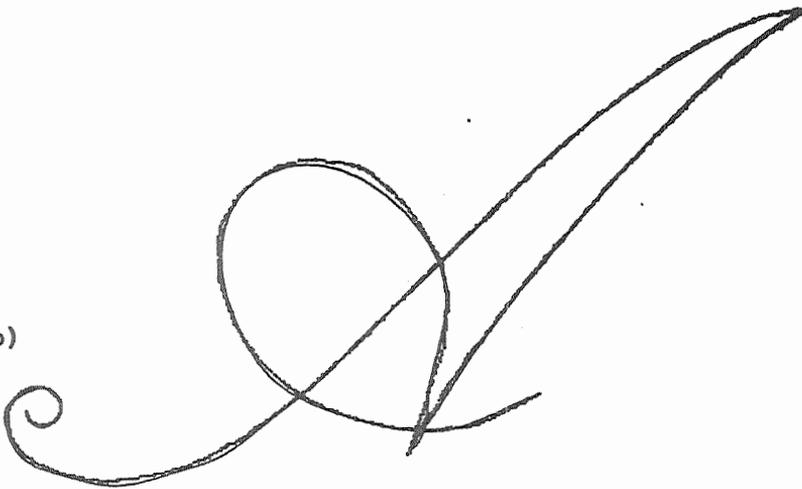
Punti : 22

$\alpha = 0.8$

a)



b)



c)

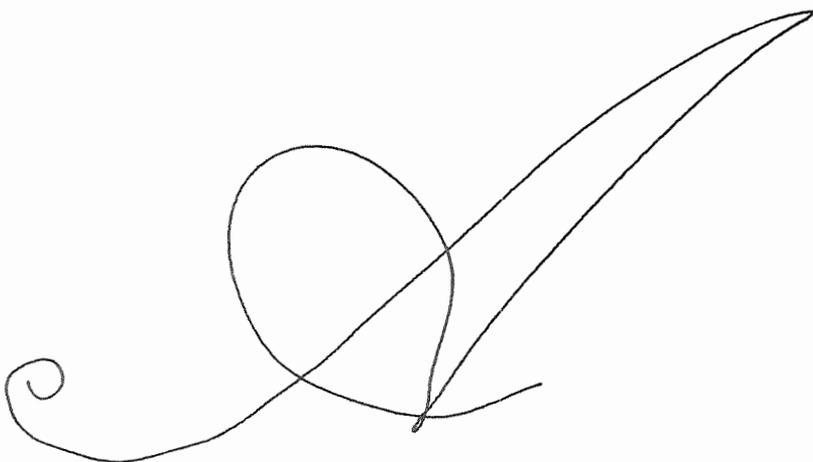


Fig. 10

a) punteggiata originale

b) curva interpolata sovrapposta all'originale

c) solo curva interpolata

Metodo : Cubica Relaxed

Punti : 23

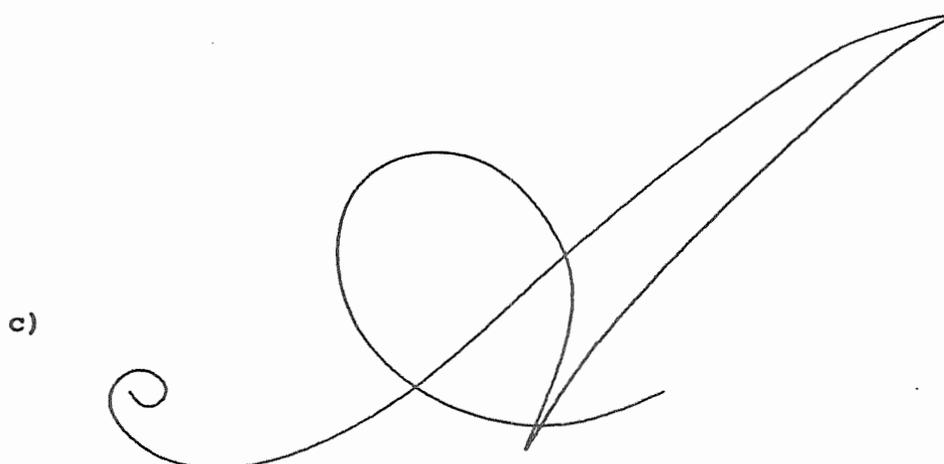
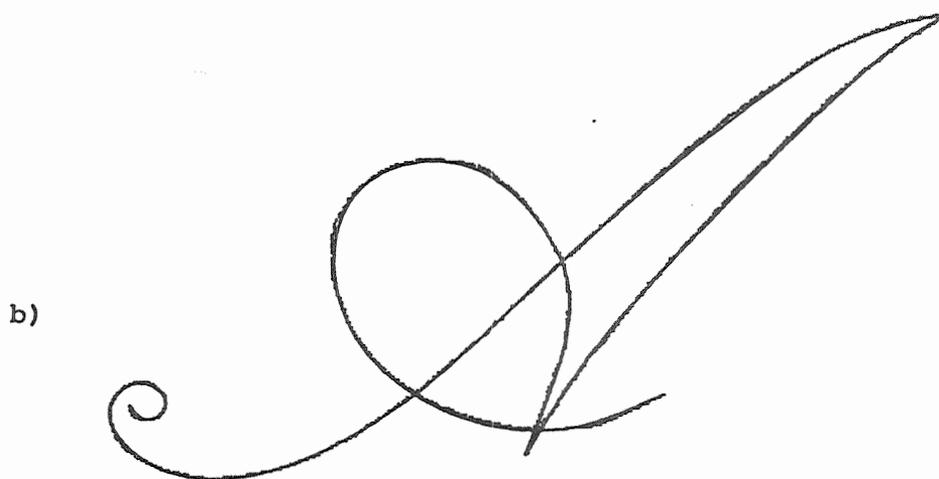
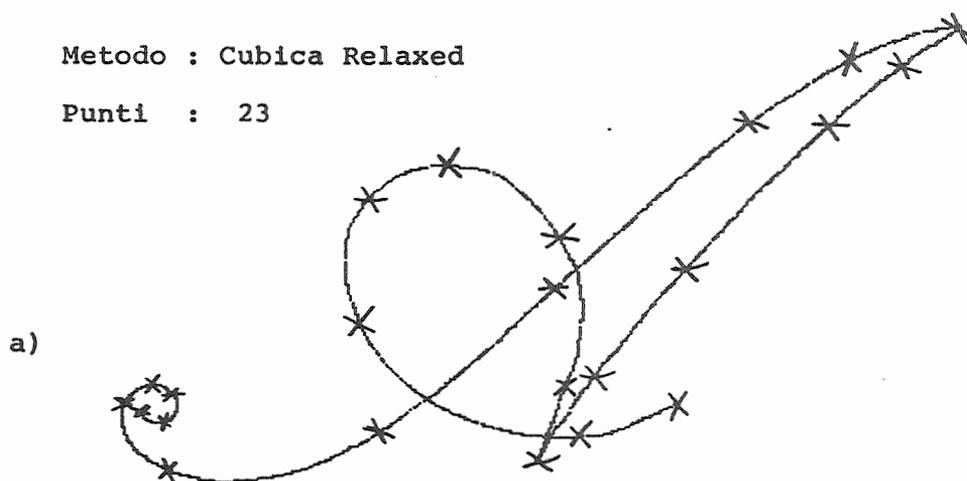


Fig. 11

- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

Metodo : Angular

Punti : 11

a)



b)



c)



Fig. 12

a) punteggiata originale

b) curva interpolata sovrapposta all'originale

c) solo curva interpolata

Metodo : Cardinal

Punti : 12

$a = 0.8$

a)



b)



c)



Fig. 13

a) punteggiata originale

b) curva interpolata sovrapposta all'originale

c) solo curva interpolata

Metodo : Cubica Relaxed

Punti : 12

a)



b)



c)



Fig. 14

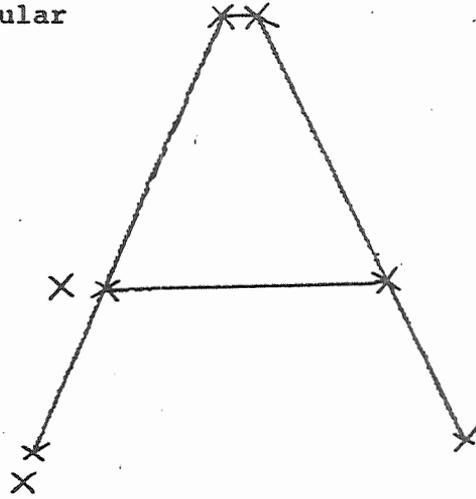
a) punteggiata originale

b) curva interpolata sovrapposta all'originale

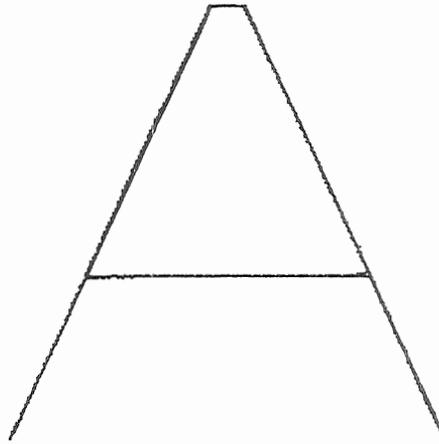
c) solo curva interpolata

Metodo : Angular
Punti : 8

a)



b)



c)

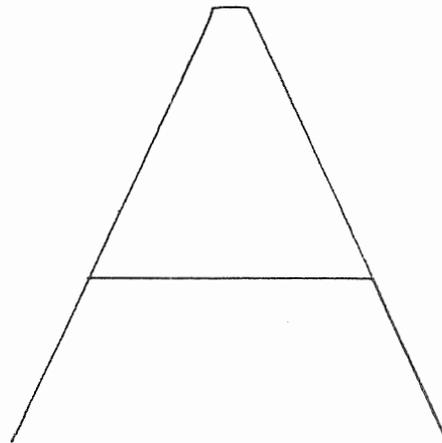


Fig. 15

- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

Metodo : Cardinal
Punti : 8
a = 0.0001

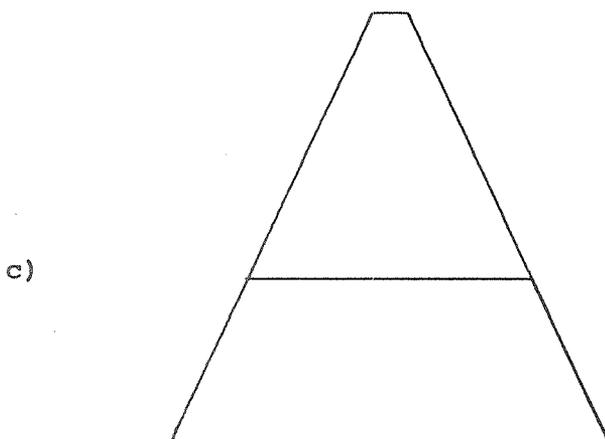
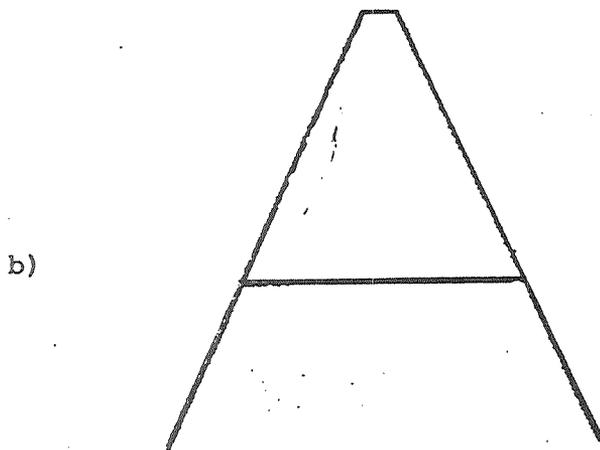
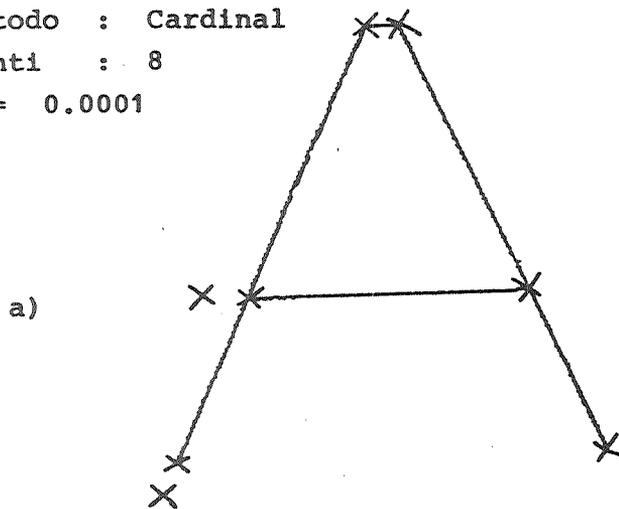


Fig. 16

- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

Metodo : Cubica R.
Punti : 17

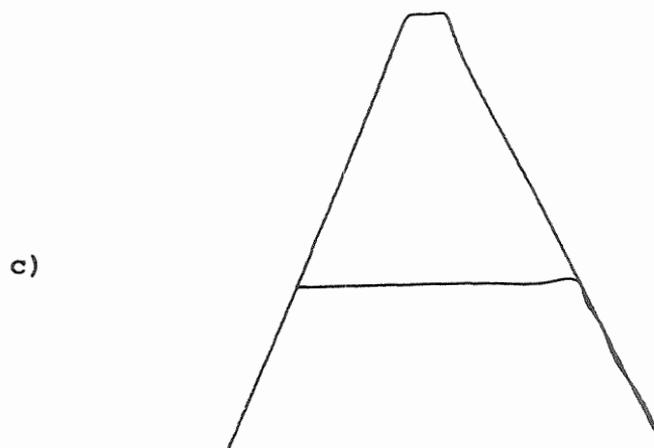
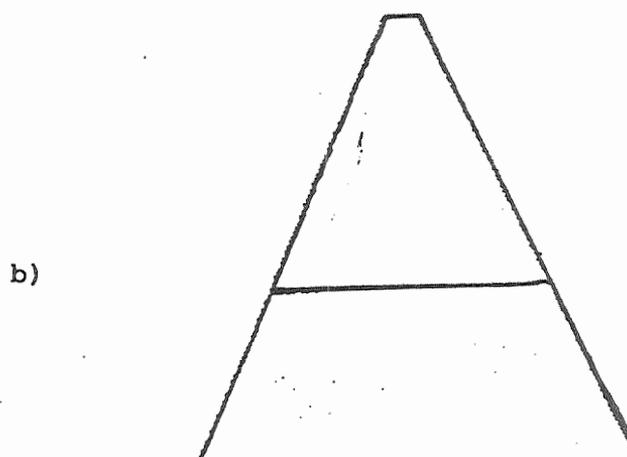
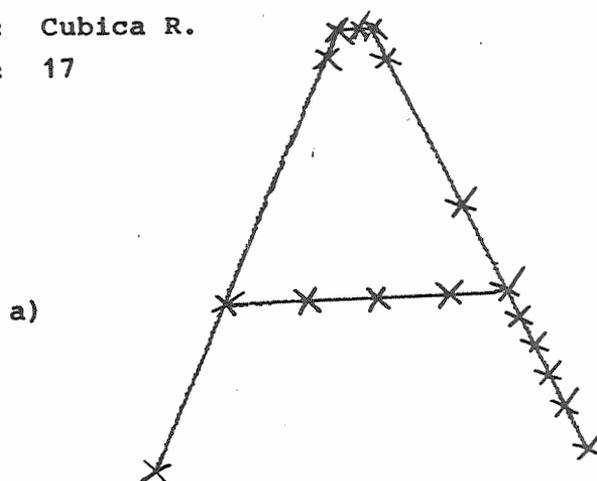


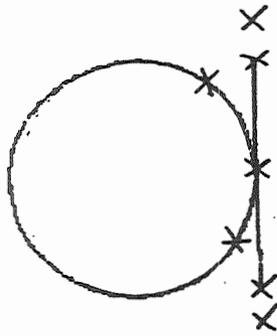
Fig. 17

- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

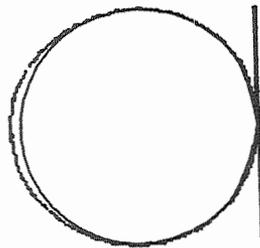
Metodo : Angular

Punti : 8

a)



b)



c)

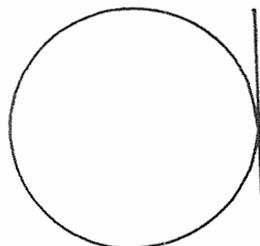


Fig. 18

a) punteggiata originale

b) curva interpolata sovrapposta all'originale

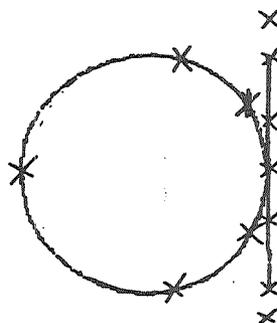
c) solo curva interpolata

Metodo : Cardinal

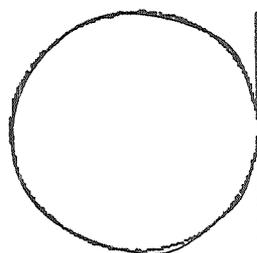
Punti : 13

$a = 0.7$

a)



b)



c)

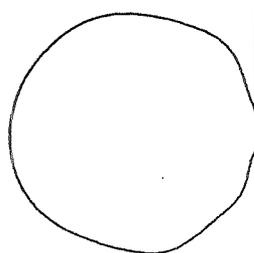


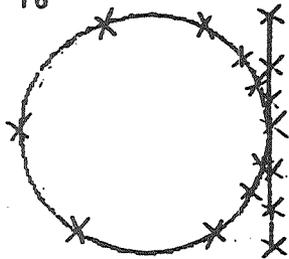
Fig. 19

- a) punteggiata originale
- b) curva interpolata sovrapposta all'originale
- c) solo curva interpolata

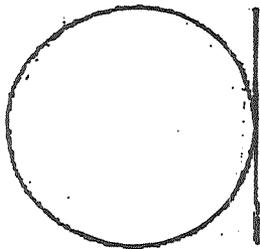
Metodo : Cubica Relaxed

Punti : 16

a)



b)



c)

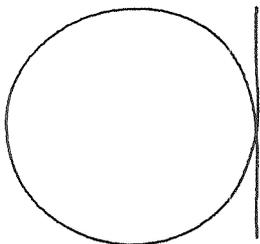
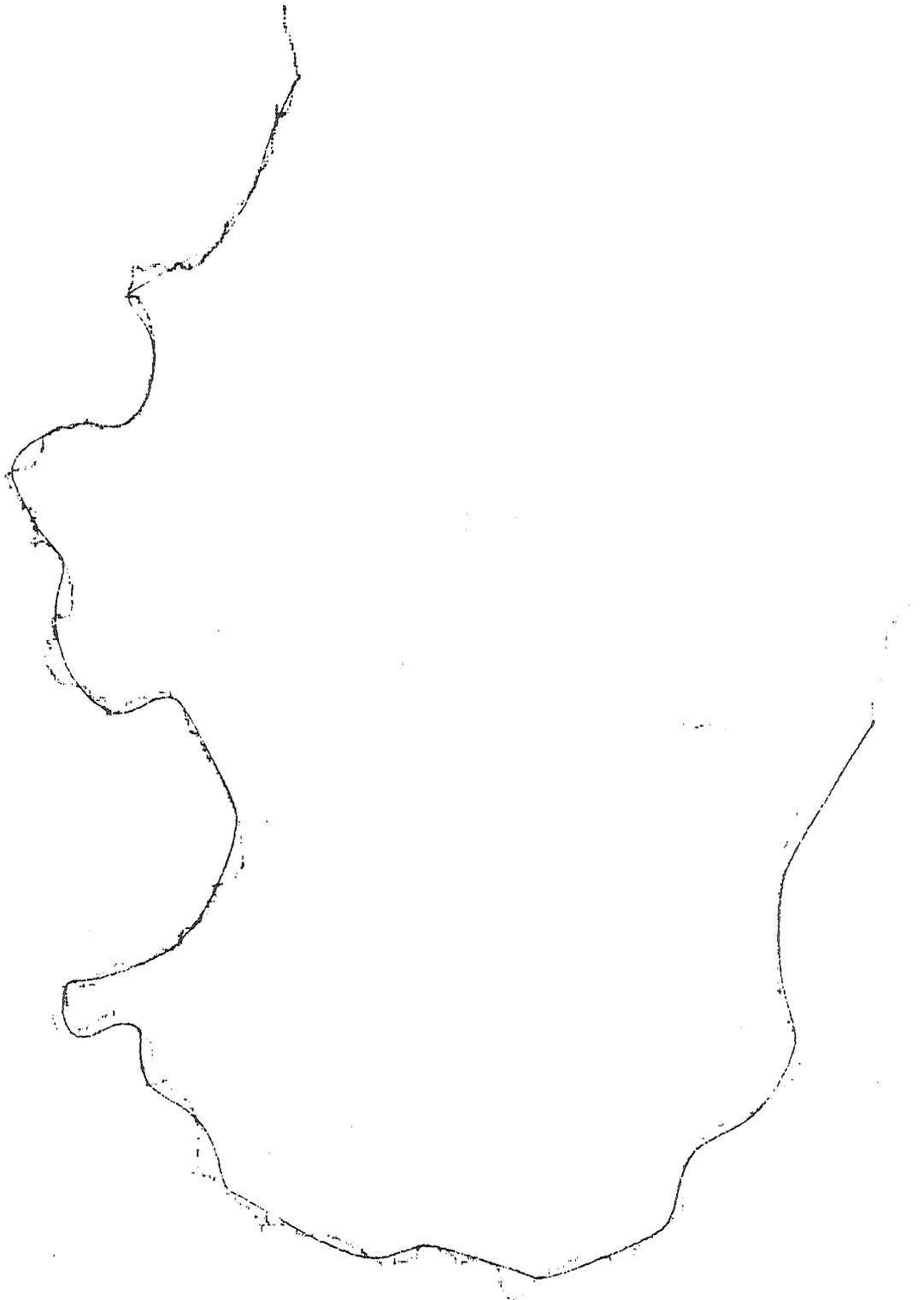


Fig. 20

a) punteggiata originale

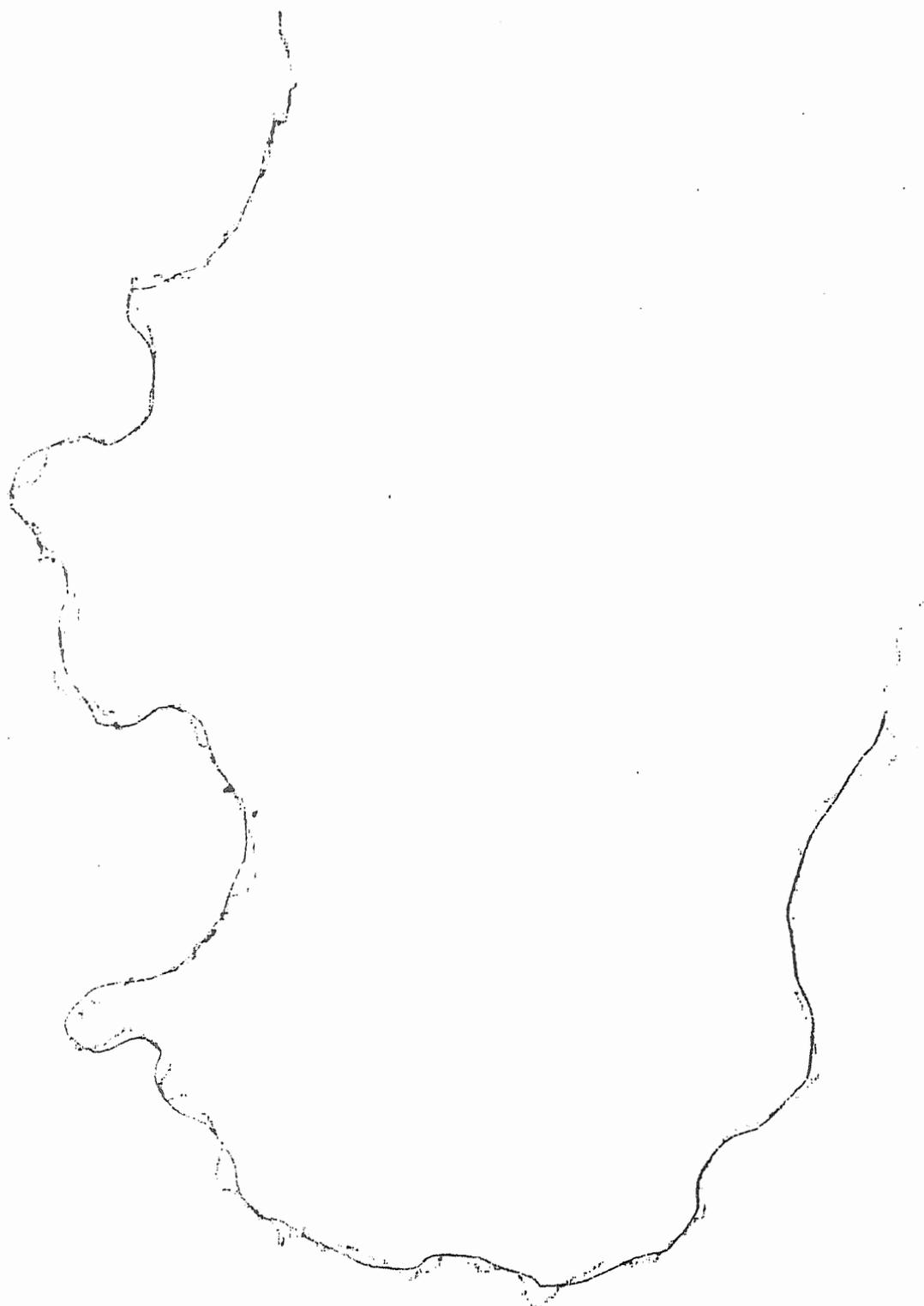
b) curva interpolata sovrapposta all'originale

c) solo curva interpolata

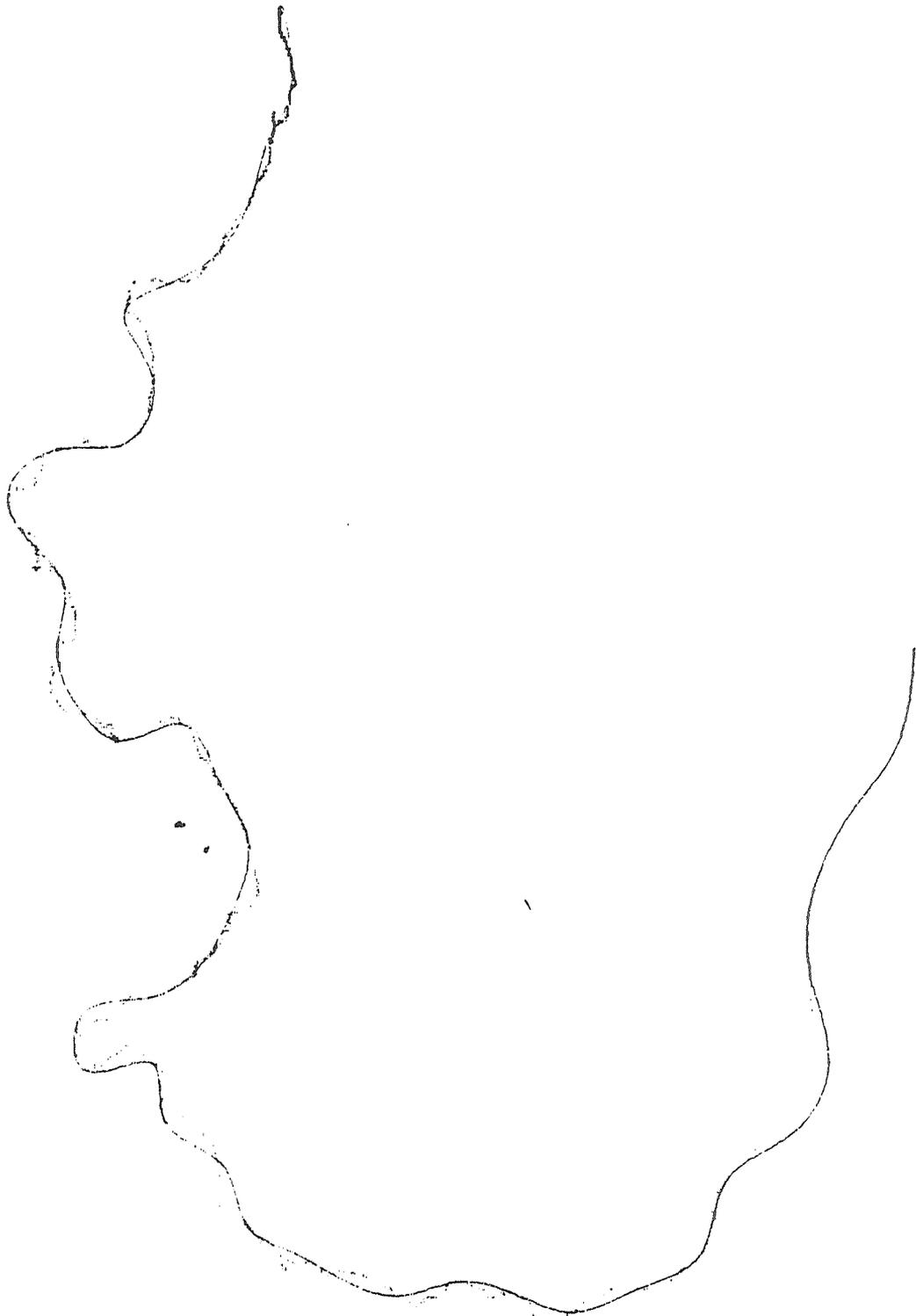


Metodo : Angular

Punti utilizzati : \approx 50 (1 ogni 50 dell'originale)

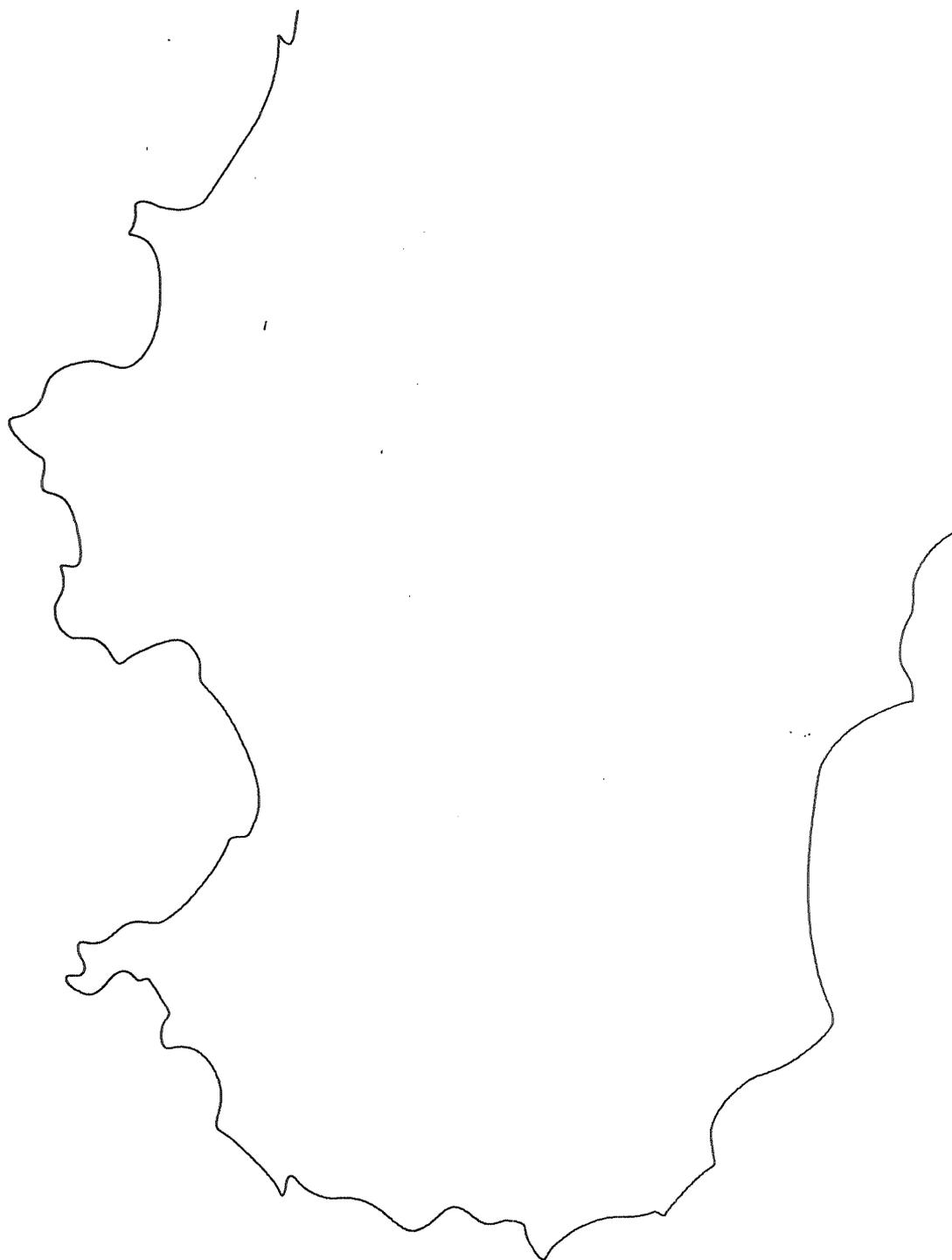


Metodo : Cardinal
Punti utilizzati : ≈ 50 (1 ogni 50 dell'originale)
Valore coefficiente : 1



Metodo : Cubica Ciclica

Punti utilizzati : \approx 50 (1 ogni 50 dell'originale)

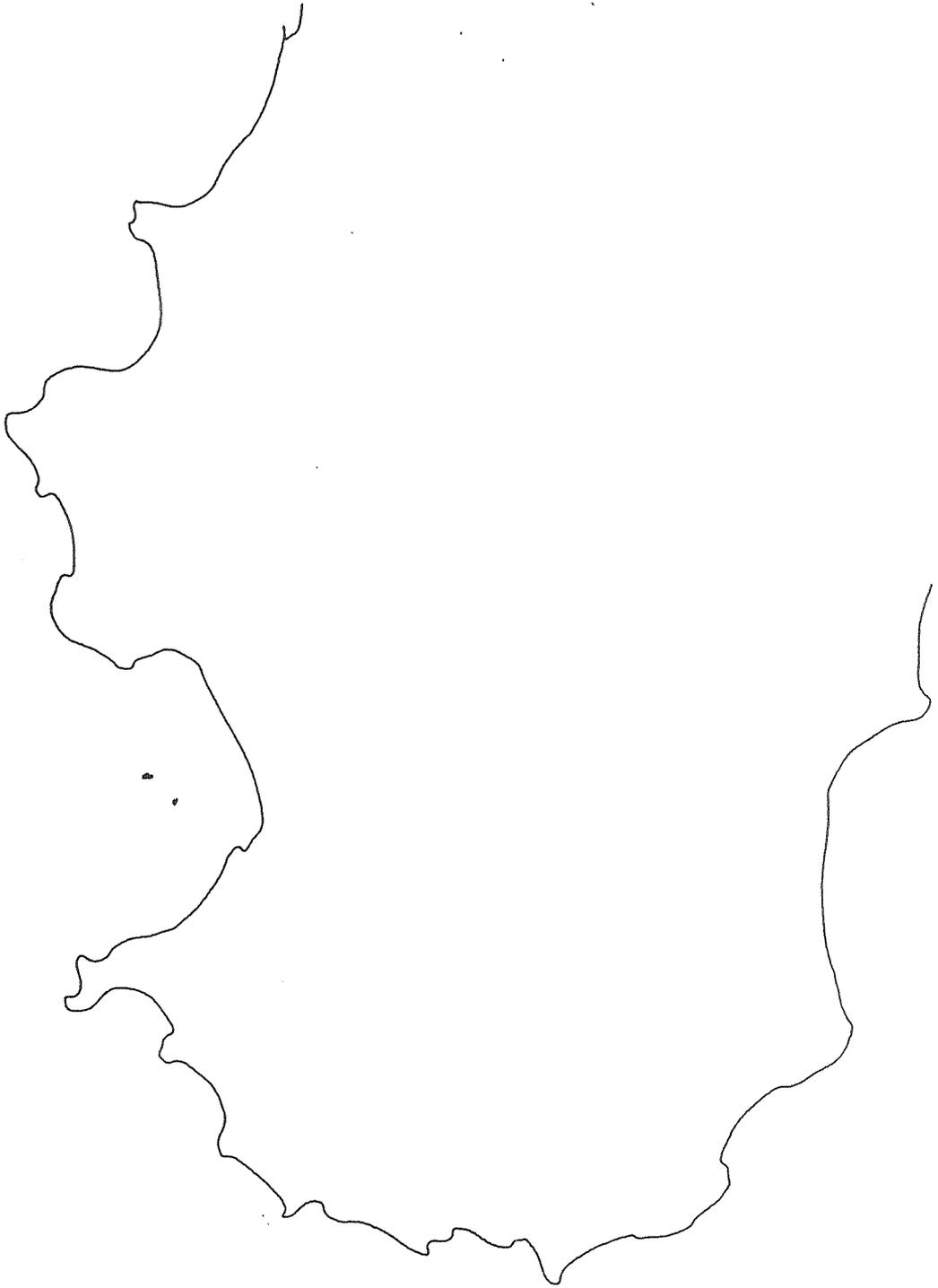


Metodo : Angular
Punti utilizzati : 92

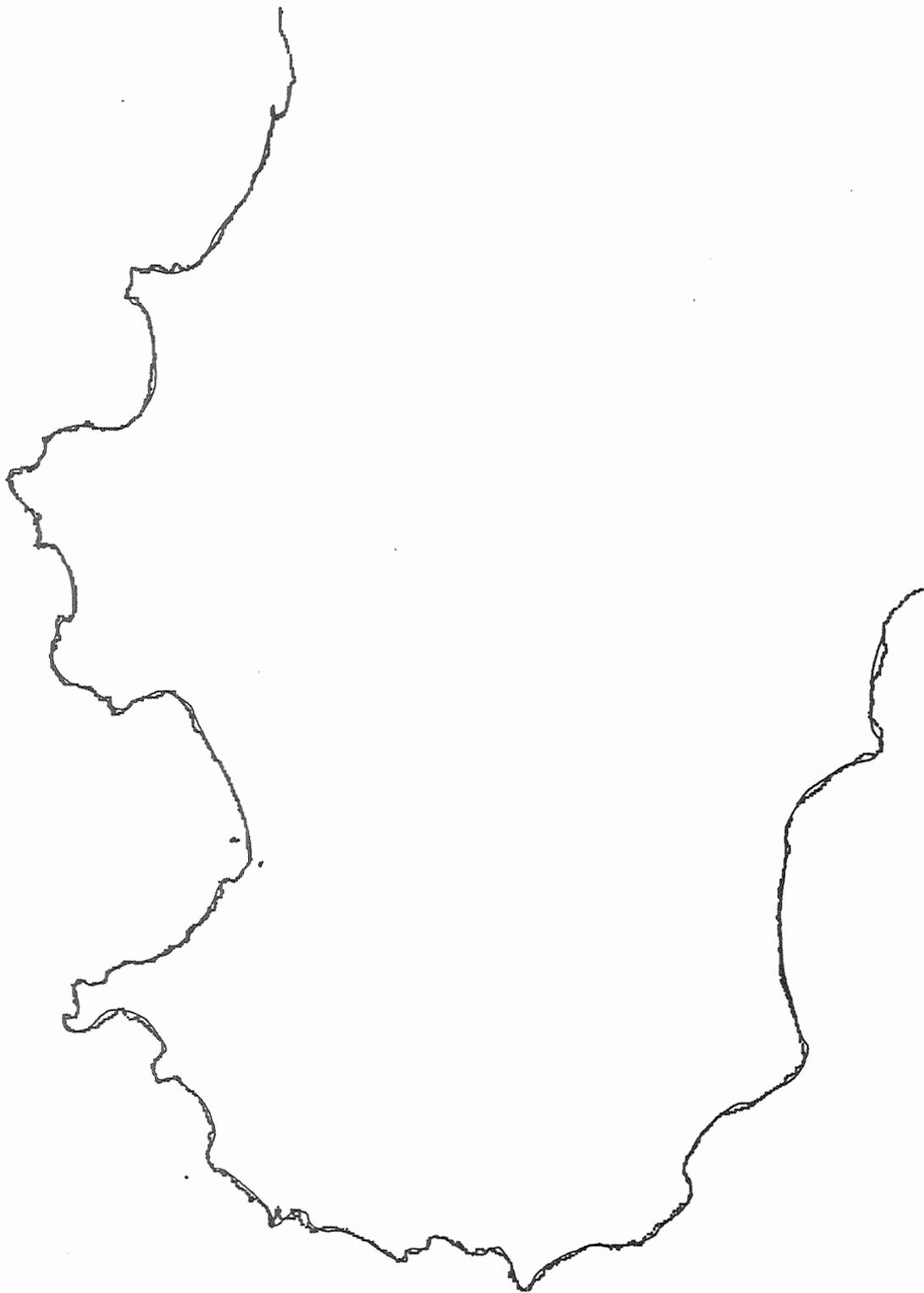


Metodo : Angular

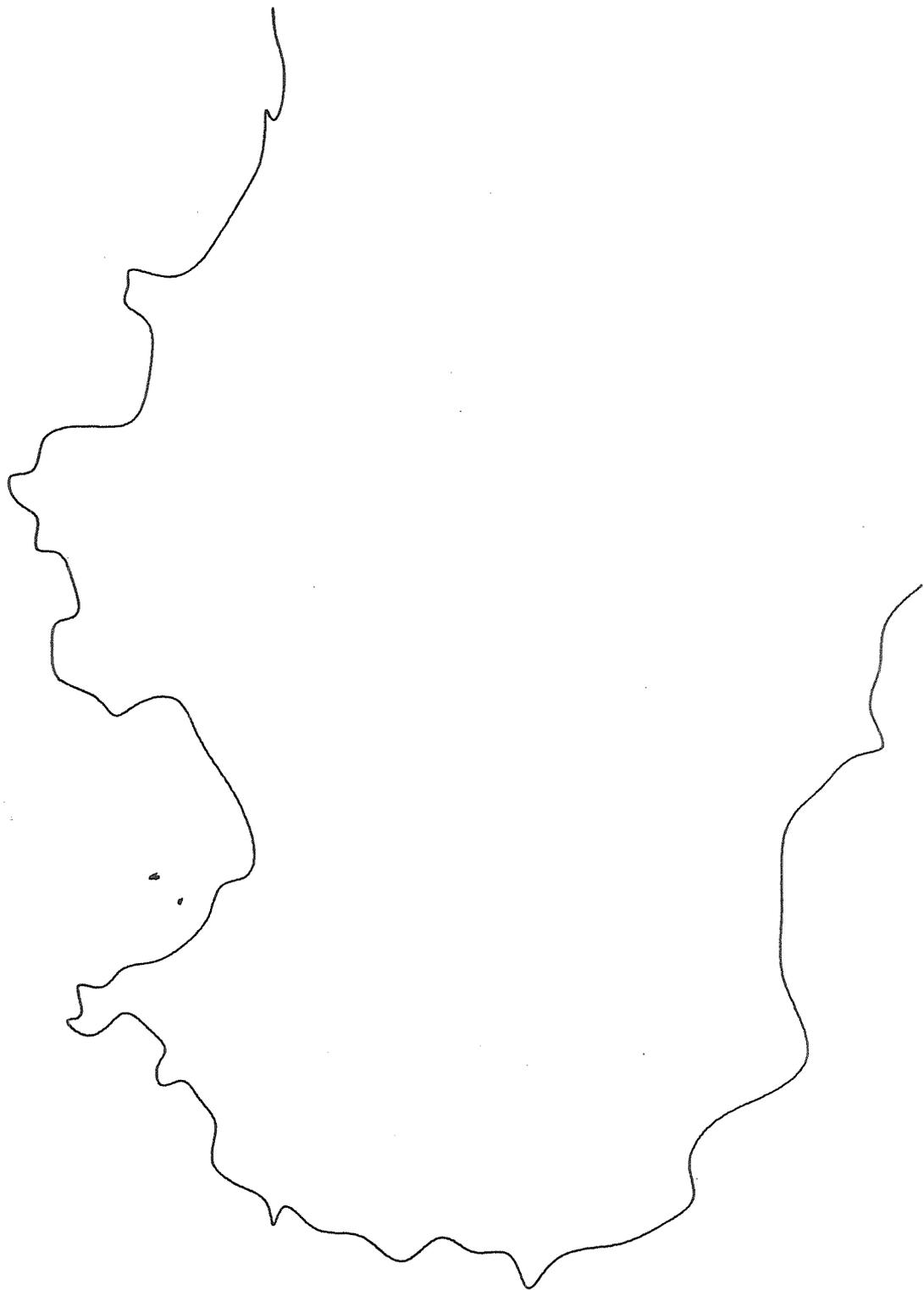
Punti utilizzati : 92



Metodo : Cardinal
Punti utilizzati : 81
Valore coefficiente : 0.9

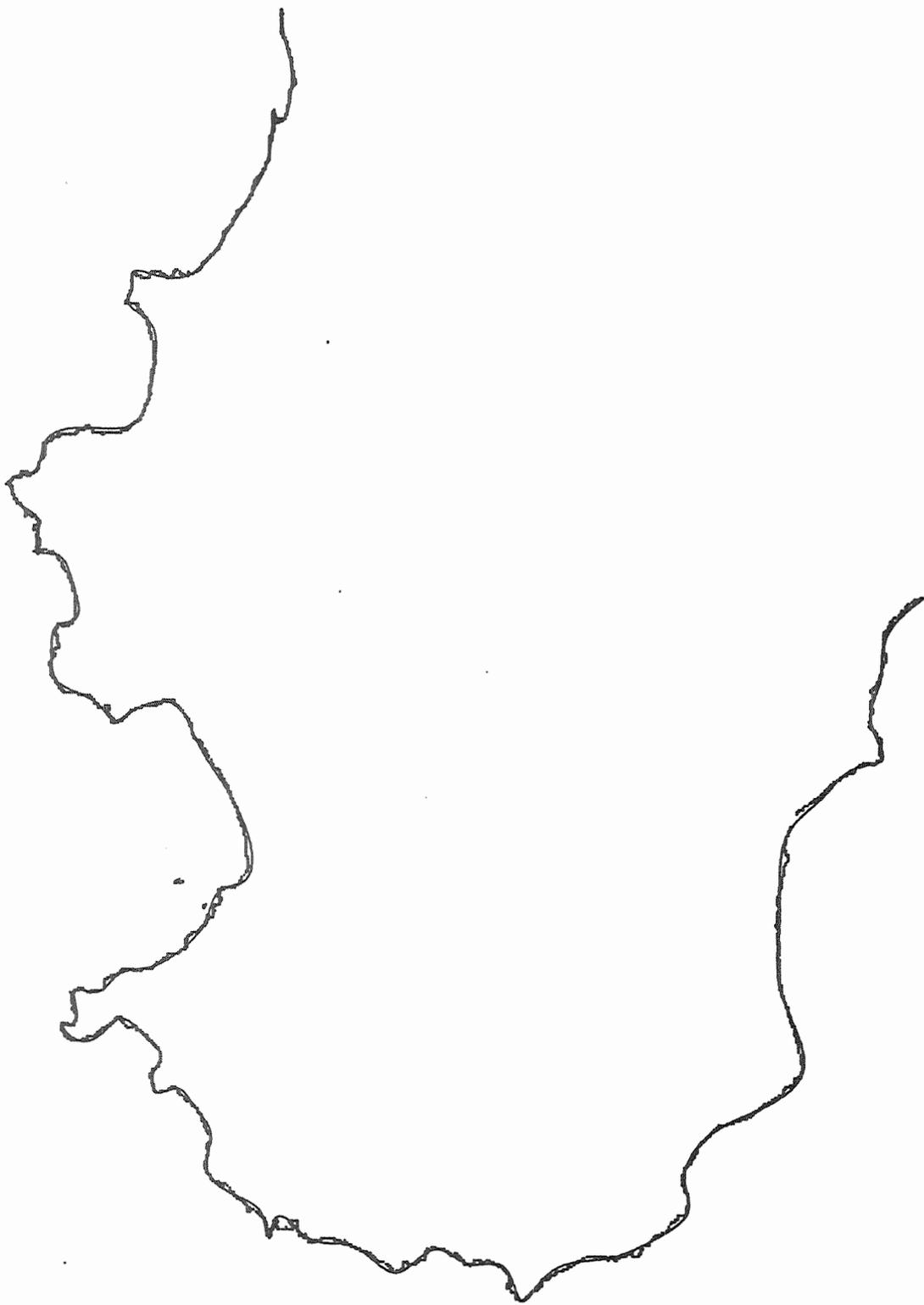


Metodo : Cardinal
Punti utilizzati : 81
Valore coefficiente : 0.9



Metodo : Cubica Relaxed

Punti utilizzati : 84



Metodo : Cubica Relaxed

Punti utilizzati : 84

NOTE CONCLUSIVE

Premesso che dal presente lavoro non ci si attendevano certo risultati eclatanti o definitivi, ma solo valutazioni orientative, pure il bilancio finale appare soddisfacente .

Limitandosi infatti all'aspetto grafico del problema dell'interpolazione, in cui appaiono essenziali in un metodo il suo facile uso, un numero necessario di punti quanto piu' piccolo possibile ed una notevole flessibilita' nel riprodurre le forme piu' svariate, si puo' decisamente affermare che non tutte le tecniche esaminate si sono comportate in modo equivalente.

Alcune di esse, infatti, trovano notevole difficolta' nel riprodurre forme poligonali (Cubiche), altre non gradiscono profili mistilinei (Cardinal) ; solo una, e cioe' la Angular, possiede poi proprieta' locale (anche se tutte, nell'uso pratico, risentono sensibilmente dello spostamento di un punto solo nell'intorno di questo); infine tutte meno una richiedono la specifica di ulteriori parametri oltre ai punti da interpolare (il coefficiente "a" nella Cardinal, il "tipo" nella Cubica, l'ordine nelle B-spline).

Non esistono invece, salvo profili particolari, grosse differenze nel numero di punti necessari per una corretta riproduzione; in leggero vantaggio in questo senso appaiono solo Angular e Cardinal .

Per quanto riguarda infine la facilita' d'impiego, sia Angular che Cubica occupano i primi posti .

Una nota particolare va riservata, a parere di chi scrive , al problema dell'acquisizione automatica dei punti da interpolare; problema aperto, come piu' volte osservato e tale, allo stato attuale delle cose, da non consentire precise valutazioni quantitative.

Bibliografia

- 1) D. F. Rogers & J. A. Adams , "Matematical Element for Computer Graphics" , Mc Graw-Hill , New York (1976) .
- 2) C. de Boor , "A Practical Guide to Splines" , Springer Berlin (1972) .
- 3) G. Alia , F. Barsi , E. Martinelli , N. Tani , "A New Approach to the Interpolation Problem in Computer Graphics" , I.E.I. Internal Report , R-84003/P/E , Serie Speciale Convenzione Selenia-C.N.R. , Pisa , Sept. 1984
- 4) G. Alia , F. Barsi , E. Martinelli , N. Tani , "Table look-up Graphic Interpolation" , Proceeding of EUROGRAPHICS '85 , pp. 279-293 , Sept. 1985 , Nice , France .
- 5) J. H. Clark , "Parametric Curves , Surfaces and Volumes in Computer Graphics and Computer Aided Geometric Design" , Technical Report N. 221 , Computer Systems Laboratory , Dept. Electrical Eng. and Computer Science , Stanford University , Stanford , California , USA , November 1981 .
- 6) N. Tani , "Sperimentazione di un nuovo metodo di interpolazione grafica mediante un sistema interattivo realizzato su minicalcolatore" , Rapporto Interno I.E.I. R-85002/P , Serie Speciale Convenzione Selenia-C.N.R. , Pisa , Febbraio 1985 .
- 7) W. Bohm , G. Farin , J. Kahmann , "A Survey of Curve and Surfaces Methods in CAGD" , Computer Aided Geometric Design , Vol. 1 , (1984) .
- 8) J. A. Adams , "Cubic Spline Curve Fitting with Controlled End Conditions" , Computer Aided Design , (1974) .
- 9) C. de Boor , "On Calculating with B-splines" , Journal Approximation Theory , (1976) .
- 10) A. R. Forrest , "Interactive Interpolation and Approximation by Bezier Polynomials" , Computer Journal , (1972)
- 11) T. Pavdilis , "Curve Fitting with Conic Splines" , ACM Transaction on Graphics , (1973)

12) G.Alia, F.Barsi, E.Martinelli, N.Tani, "Angular Spline: a new approach to the Interpolation Problem in Computer Graphics", in corso di stampa su Computer Vision, Graphics and Image Processing, 1986