Le curve di Bézier

Giulio C. Barozzi
Università di Bologna

Molti strumenti che la tecnologia mette oggi a nostra disposizione (telefoni cellulari, lettori di CD audio e di DVD, ecc.) devono il loro funzionamento all'utilizzo di strumenti matematici e fisici sofisticati, che restano sconosciuti alla stragrande maggioranza dei loro utilizzatori. Lo stesso vale per la maggior parte degli strumenti diagnostici di cui dispone la moderna medicina: ecografia, TAC, NMR ecc.

In questa relazione vogliamo concentrare la nostra attenzione su una possibilità offerta della moderna tecnologia di stampa, che si fonda sull'uso di strumenti matematici relativamente semplici. Alludiamo al fatto che le moderne stampanti, siano esse di tipo laser oppure a getto d'inchiostro, consentono di ridurre ed ingrandire la pagina senza che questo alteri la nitidezza dei caratteri tipografici. Ciò significa che i caratteri stessi non sono memorizzati come immagini definite per punti (*bitmapped*), ma piuttosto sono memorizzate le equazioni delle curve che delimitano i caratteri stessi. Dunque la definizione del carattere dipende dalla qualità della stampante usata, indipendentemente dal fattore di ingrandimento o riduzione della pagina.

La matematica che sta dietro a tutto ciò è relativamente semplice; la sua comprensione richiede la conoscenza dei polinomi di grado 3, ivi compresa la derivazione degli stessi polinomi, ed anche la conoscenza dei rudimenti del calcolo vettoriale nel piano. Tutti concetti che non erano sconosciuti ai pionieri del calcolo infinitesimale.

È dunque abbastanza sorprendente scoprire che le curve che ci apprestiamo a descrivere traggono il loro nome da Pierre Bézier (1910-1999), un ingegnere meccanico francese che lavorava per la fabbrica Renault. I problemi con cui Bézier aveva a che fare non erano di natura tipografica, ma riguardavano piuttosto la gestione delle macchine a controllo numerico che tranciavano i pezzi di lamiera con cui fare le carrozzerie delle auto. Va detto che Bézier stesso, molto onestamente, ammise che alle stesse curve era giunto anche il suo collega della Citroën P. De Casteljau, ma i vincoli di segretezza imposti da questa fabbrica hanno fatto sì che tali curve restassero legate al suo nome.

Di che cosa si tratta dunque? Ogni curva aperta o chiusa nel piano si può ottenere raccordando un certo numero di archi che congiungano coppie di punti in modo "dolce". Per essere più precisi, il problema di base si pone nel modo seguente:

Dati nel piano tre punti A, B e C non allineati, congiungere A con C mediante una curva regolare, che sia tangente in A alla retta AB e tangente in C alla retta BC.

Vedremo nel seguito che il problema, in termini esposti, è un po' troppo vago, ma lo preciseremo strada facendo.

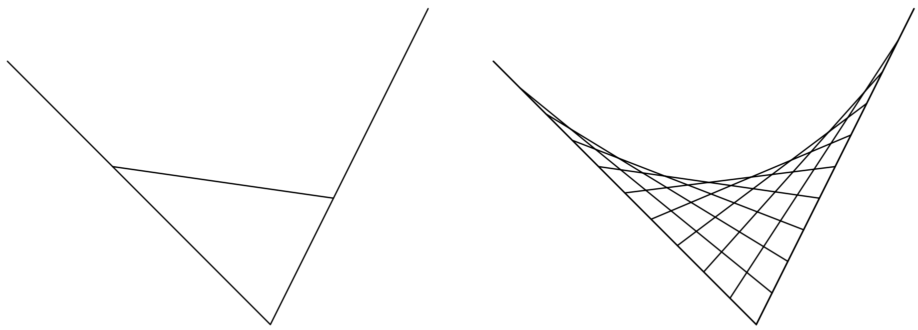
Siamo interessati ad una soluzione che sia semplice, cioè ammetta una descrizione matematica semplice. Aiutiamoci con una terminologia presa in prestito dalla Cinematica. Immaginiamo il punto X_1 che descrive il segmento orientato AB muovendosi di moto rettilineo uniforme ed impiegando un'unità di tempo (diciamo un secondo); la legge oraria che descrive tale moto sarà

$$X_1 - A = t(B - A) \quad \iff \quad X_1 = A + t(B - A) = (1 - t)\,A + t\,B, \quad 0 \le t \le 1.$$

Contemporaneamente un punto X_2 descriva, allo stesso modo, il segmento orientato BC; avremo analogamente

$$X_2 = B + t(C - B) = (1 - t)\,B + t\,C, \quad 0 \le t \le 1.$$

Una soluzione semplice al problema posto è fornita dall'inviluppo delle rette X_1X_2, al variare di t da 0 a 1.



Se X è il punto di tangenza della retta X_1X_2 con tale inviluppo, avremo l'uguaglianza dei rapporti

$$\frac{\overline{X_1X}}{\overline{X_1X_2}} = \frac{\overline{AX_1}}{\overline{AB}} = \frac{\overline{BX_2}}{\overline{BC}} = t$$

(abbiamo indicato col soprassegno le lunghezze dei segmenti). Dunque il punto X è dato dalla formula (analoga alle precedenti)

$$X = (1 - t)\,X_1 + t\,X_2.$$

Sostituendo al posto di X_1 e X_2 le espressioni precedenti, in funzione di A, B e C, si ottiene, con qualche passaggio

$$X = X(t) = (1 - t)^2 A + 2t(1 - t)B + t^2 C. \tag{1}$$

Si verifica subito che $X(0) = A$, $X(1) = B$; derivando si trova poi

$$X'(t) = -2(1 - t)A + 2(1 - 2t)B + 2tC, \tag{1}$$

quindi $X'(0) = -2A + 2B = 2(B - A)$, $X'(1) = -2B + 2C = 2(C - B)$.

In altri termini: i vettori tangenti alla curva nei punti iniziale e finale, a meno del fattore di proporzionalità 2, sono gli stessi segmenti orientati $B - A$ e $C - B$. È chiaro che se percorressima la curva trovata a velocità dimezzata (formalmeente: se scrivessimo $t/2$ al posto di t, facendo andare t da 0 a 2) otterremo una parametrizzazione della curva stessa in cui i vettori tangenti sarebbero esattamente $B - A$ e $C - B$.

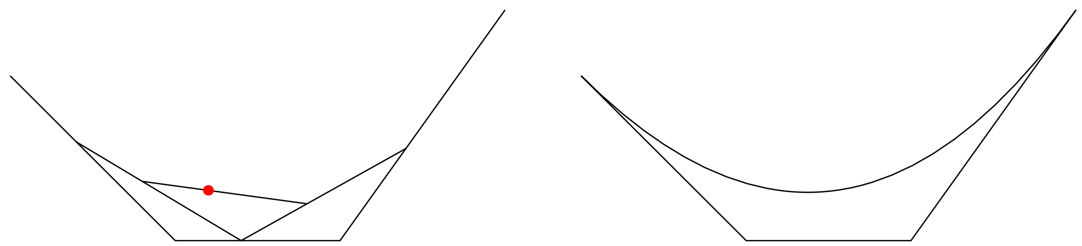
La curva che abbiamo ottenuto, parametrizzata dalla (1), può essere chiamata *curva di Bézier del secondo ordine*, in quanto l'ascissa e l'ordinata del punto X sono date da due polinomi di secondo grado in t.

La soluzione che abbiamo ottenuto è un po' troppo "rigida", nel senso che, dati gli estremi A e C della curva che vogliamo tracciare, il solo punto B (detto *punto di controllo*) determina i vettori tangenti in A e in C non solo quanto a direzione, ma anche quanto a lunghezza: abbiamo visto che tali lunghezze sono il doppio di quelle dei vettori $B - A$ e $C - B$. È desiderabile poter controllare separatamente e completamente i vettori tangenti negli estremi della curva.

Ciò può essere ottenuto introducendo non un solo punto di controllo ma due punti. Per congiungere A con D consideriamo due punti di controllo B e C. Vogliamo che le tangenti in A e in D siano proporzionali ad vettori $B - A$ e $D - C$ rispettivamente.

Immaginiamo di costruire due curve di Bézier del secondo ordine, una sulla terna A, B, C e la seconda sulla terna B, C, D. La prima curva sarà descritta dal punto X dato dell'equazione (1), mentre la seconda curva di Bézier sarà descritta da un punto Y dato dalla parametrizzazione

$$Y = Y(t) = (1-t)^2 B + 2t(1-t)C + t^2 D. \tag{2}$$



Finalmente consideriamo l'inviluppo delle rette XY. Il punto su tale inviluppo sarà dato dall'equazione

$$Z = (1-t)\,X + t\,Y;$$

sostituendo le espressioni di X e Y, si ottiene con qualche passaggio, la *curva di Bézier del terzo ordine* (o semplicemente *curva di Bézier*)

$$Z = Z(t) = (1-t)^3 A + 3(1-t)^2 tB + 3(1-t)t^2 C + t^3 D. \tag{3}$$

Si verifica subito che $Z(0) = A$, $Z(1) = D$; si trova poi

$$Z'(t) = -3(1-t)^2 A + 3(1-t)(1-3t)B + 3t(2-3t)C + 3t^2 D,$$

da cui $Z'(0) = 3(B - A)$, $Z'(1) = 3(D - C)$.

Possiamo osservare che i polinomi

$$(1-t)^3, \quad 3(1-t)^2 t, \quad 3(1-t)t^2, \quad t^3$$

sono precisamente quelli che si ottengono sviluppando l'identità $1 = 1^3 = [(1-t) + t]^3$ con la formula del binomio. Diamo loro un nome

$$H_1(t) := (1-t)^3, \quad H_2(t) := 3(1-t)^2 t, \quad H_3(t) := 3(1-t)t^2, \quad H_4(t) := t^3,$$

possiamo osservare che per $t = 0$ e $t = 1$ essi verificano le seguenti condizioni:

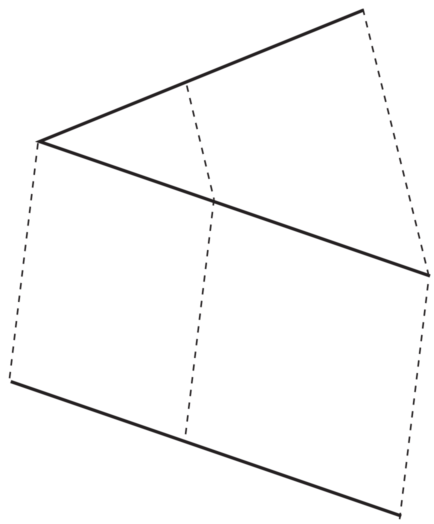
k	1	1	3	4
$H_k(0)$	1	0	0	0
$H'_k(0)$	-3	3	0	0
$H_k(1)$	0	0	0	1
$H'_k(1)$	0	0	-3	3

Si tratta dunque di polinomi che risolvono un problema di interpolazione nel senso di Hermite; tale problema consiste nel determinare il polinomio di grado $2n - 1$ che in n punti assegnati assume determinati valori, e lo stesso per la derivata prima del polinomio negli stessi punti.

Da un punto di vista analitico il problema è risolto. Dal punto di vista sintetico è interessante costruire le curve di Bézier con un programma di geometria dinamica, come Cabri Gèomètre. La cosa non è difficile. Occorre predisporre una macro che, dato un segmento AC e un punto X_1 su di esso, consente di costruire su un qualsivoglia altro segmento CD il punto X_2 che lo divide nello stesso rapporto in cui X_1 divide AB, vale a dire

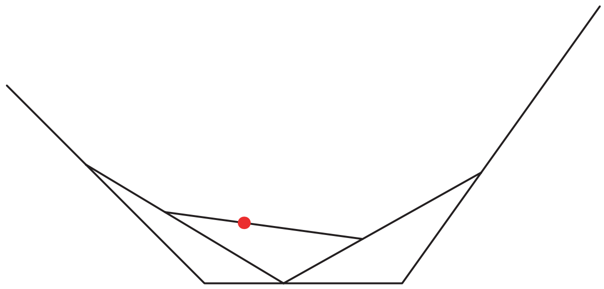
$$\frac{\overline{AX_1}}{\overline{AB}} = \frac{\overline{CX_2}}{\overline{CD}}.$$

Tale costruzione presenta qualche insidia, in quanto essa deve essere abbastanza "robusta" da funzionare correttamente quale che sia la posizione di CD rispetto ad AB, in particolare deve poter funzionare correttamente anche se C appartiene alla retta AB.



Per costruire una curva di Bézier del secondo ordine, si procede nel modo seguente: scelti i tre punti A, B e C, si costruisce il segmento AB e si sceglie un punto X_1 su di esso. Una prima applicazione della macro sopra descritta (a partire dai segmenti AB, BC e dal punto X_1) fornisce il punto X_2 sul segmento BC; si costruisce il segmento X_1X_2, e finalmente, una terza applicazione della nostra macro, fornisce il punto X sul tale segmento. La curva di Bézier è il luogo descritto dal punto X al variare di X_1 sul segmento AB.

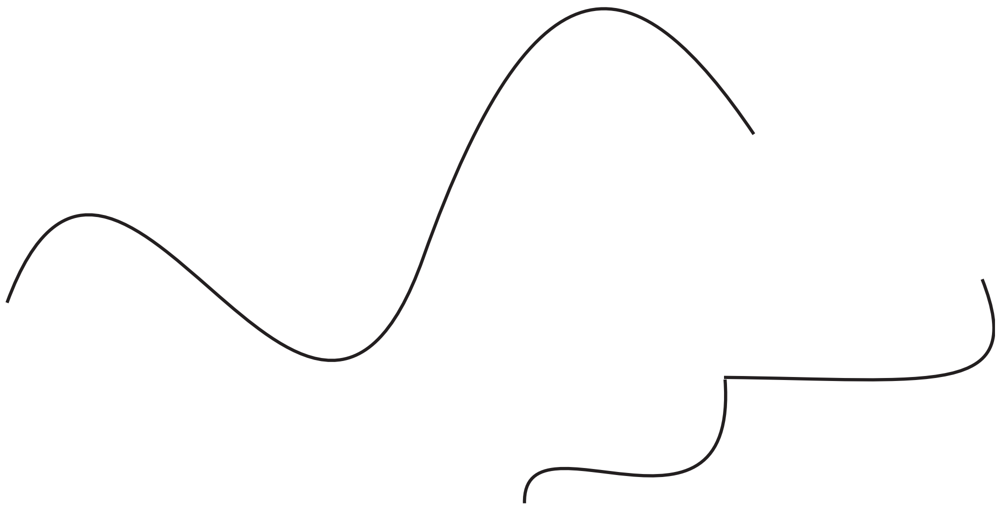
La costruzione di una curva di Bézier vera e propria (del terzo ordine) è un po' più complessa, ma sostanzialmente semplice; possiamo riprendere una delle figure precedenti.



Come si vede, la macro viene utilizzata 5 volte, a partire dal segmento AB e dal punto X_1, per generare in sequenza i punti X_2 (su BC), X_3 (su CD), X (su X_1X_2), Y (su X_2X_3) e Z (su XY). Il luogo descritto da quest'ultimo punto al variare di X_1 è la curva di Bézier generata dalla quaterna A, B, C, D.

I sistemi di disegno attualmente disponibili, come Adobe Illustrator, dispongono dierettamente delle curve di Bézier. Si abbassa il mouse in un determinato punto (il punto A) e lo si trascina fino a sollevarlo in un secondo punto B. Si sposta poi il mouse nel punto finale D e lo si trascina, mantenendolo abbassato, fino al punto C' simmetrico di C rispetto a D.

Il tracciamento della curva di Bézier è (apparentemente) istantaneo. Una volta tracciata tale curva, la si può ritoccare spostando separatamente i 4 punti che la determinano. Si possono poi concatenare due o più curve di Bézier in modo da ottenere tracciati comunque complessi. Questo può essere fatto sia in modo da raccordare due archi contigui e le tangenti nell'estremo comune, oppure soltanto raccordare gli archi ma non le tangenti, di modo che nell'estremo comune possa formarsi un punto angoloso. Nel primo caso le due curve di Bézier sono raccordate tramite un "punto morbido", nel secondo caso tramite un "punto ad angolo".



La curva di che fatto viene tracciato sullo schermo è inevitabilmente una poligonale con lati abbastanza piccoli da ingannare l'occhio. Di fatto un numero di punti dell'ordine di qualche decina viene calcolato per ogni arco di Bézier. Se si tiene presente che ogni punto richiede il calcolo di due polinomi cubici per un assegnato valore del parametro t, è evidente che è opportuno il ricorso allo schema di Ruffini-Horner, che consente il calcolo di un polinomio cubico con 3 addizioni e altrettante moltiplicazioni.

In conclusione, le curve di Bézier mi sembrano un argomento interessante per alcuni motivi:

– hanno larga applicazione nei sistemi di grafica al calcolatore e nei sistemi di stampa;
– hanno una spiegazione relativamente semplice;
– consentono una trattazione sia analitica, dunque orientata ad un linguaggio di programmazione procedurale, che sintetica dunque fruibile su un sistema di geometria dinamica;
– utilizzano nozioni di calcolo vettoriale che sono utili anche nello studio della Fisica.

Bibliografia

[1] Farin G., *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press (1988);

[2] Ferrard J.M., Lemberg H., *Mathématiques concrètes, illustrées par la TI-92 et la TI-89*, Springer Verlag-France (1998);

[3] Fruendi G., *Una soluzione moderna per un problema antico: costruzione di una curva che interpola un insieme finito di punti*, L'Insegnamento della Matematica e delle Scienze Integrate, vol. 23B (2000);

[4] Scheu G., *An Approach to the Bézier Curves with DERIVE*, DERIVE Newsletter # 19 (1995);

[5] Schlöglhofer F., *Bézier Cuves in School*, DERIVE Newsletter # 52 (2003).