T-Splines: Interpolación de Curvas C^2 Contínua

[1]Andrea Silvetti, [2]Claudio Delrieux y [1]Silvia Castro
[1]Departamento de Ciencias de la Computación
[2]Departamento de Ingeniería Eléctrica
ICIC — Instituto de Ciencias e Ingeniería de Computación
Universidad Nacional del Sur
Alem 1253 — (8000) Bahía Blanca — ARGENTINA

Resumen:

La interpolación es el más intuitivo de los esquemas de representación geométrica o algebraica de curvas a partir de puntos de control. Al mismo tiempo, en muchos problemas de ingeniería o arquitectura es necesario garantizar que las curvas resultantes pasen por localizaciones específicas, probablemente también con una tangente determinada. Sin embargo, los métodos de interpolación tienen una desventaja fatal, dado que la interpolación con trozos de curvas polinomiales solamente garantizan un orden de continuidad C^1, es decir, curvas contínuas y derivables, pero discontínuas en su segunda derivada. En este trabajo proponemos un método de interpolación, denominado T-Splines, que supera dicha deficiencia, garantizando curvas cuya segunda derivada geométrica es contínua. El mismo se basa en encontrar un grafo de control auxiliar, sobre el que se realiza una parametrización no uniforme de la curva, de modo que se satisfagan las restricciones geométricas que garantizen que el B-Spline no uniforme que aproxima a dicho grafo auxiliar interpola al mismo tiempo los puntos de control con las derivadas especificadas.

1 Introducción

En el diseño geométrico asistido por computadora (CAD) existen dos maneras diferentes de describir los datos necesarios para la especificación de una curva [1, 2]. El primer esquema, denominado *interpolación*, se caracteriza por describir la curva a partir de un conjunto destacado de puntos que deben pertenecer a la curva especificada [7]. El segundo esquema, denominado *aproximación*, también utiliza puntos de control para especificar, pero normalmente dichos puntos no pertenecen a la curva sino que controlan su forma de una manera menos directa.

La interpolación es indudablemente el más intuitivo de los esquemas. Sobre la base de un método geométrico o algebraico de interpolación de curvas a partir de puntos de control, es posible el desarrollo de sistemas interactivos de CAD en los cuales el usuario pueda modificar de una manera directa la forma y posición de la curva. Al mismo tiempo, en muchos problemas de ingeniería o arquitectura es necesario garantizar que las curvas resultantes pasen por localizaciones específicas, probablemente también con una tangente determinada [4, 5, 6, 9]. En los métodos de aproximación, como por ejemplo en el método de Bézier o utilizando B-Splines, este tipo de resultados son de mayor dificultad. A partir de la ubicación de los puntos de control es posible predecir intuitivamente la forma de la curva resultante, pero no siempre es directo e intuitivo el proceso de modificar dichos puntos para obtener un cambio específico. Por otra parte, si bien la descripción matemática de la curva queda determinada a partir de los datos, no es posible determinar en general si dicha curva pasa exactamente por un determinado lugar, o con qué tangente lo hace.

Sin embargo, los métodos de interpolación tienen una desventaja fatal, que determinó que para la mayor parte de las aplicaciones se prefieran los métodos de aproximación. En efecto, la interpolación polinomial de Lagrange produce curvas paramétricas para las cuales hay que evaluar polinomios de grado muy elevado, las cuales exhiben, además, una inestabilidad numérica considerable y una tendencia a oscilar alrededor de los puntos de control, lo cual es una propiedad indeseable. El esquema de interpolación de Hermite y otros métodos de interpolación con trozos de curvas polinomiales, por su parte, solamente garantizan un orden de continuidad C^1, es decir, curvas contínuas y derivables, pero discontínuas en su segunda derivada.

En este trabajo proponemos un método de interpolación, denominado T-Splines, que supera dicha deficiencia, garantizando curvas cuya segunda derivada geométrica es contínua. El mismo se basa en encontrar un grafo de control auxiliar, sobre el que se realiza una parametrización no uniforme de la curva, de modo que se satisfagan las restricciones geométricas que garantizen que el B-Spline no uniforme que aproxima a dicho grafo auxiliar interpola al mismo tiempo los puntos de control con las derivadas especificadas.

Este trabajo queda organizado del siguiente modo: en la siguiente sección se efectuará un breve repaso de los conceptos básicos de la teoría de aproximación e interpolación de curvas con polinomios paramétricos, discutiéndose las ventajas y limitaciones de los diferentes métodos. En la sección 3 se presentará en detalle el método propuesto, describiéndose la fundamentación teórica y algunos detalles de implementación del mismo. En la sección 4 se mostrarán algunas aplicaciones del método, y por último en la sección 5 se presentarán las conclusiones y las posibilidades de trabajo futuro.

2 Aproximación e interpolación de curvas

Esta sección está destinda a presentar brevemente los conceptos indispensables relacionados a la aproximación e interpolación de curvas, para dar completitud al trabajo y unidad a la notación. Aquellos lectores familiarizados con el tema pueden omitir la lectura de la misma. Los métodos de aproximación e interpolación de curvas y superficies son de importancia fundamental en la Computación Gráfica y el Diseño Asistido por Computadora, dado que proveen un marco para representar objetos de complejidad y diversidad geométrica arbitraria de una manera flexible y computacionalmente económica. En la representación por medio de curvas y superficies, se busca una expresión analítica en función de puntos de control, tal que se cumpla una serie de propiedades. Entre ellas podemos destacar la invariancia de la representación frente a transformaciones afines y la analiticidad de la representación en términos de garantizar un orden de continuidad C^n dado. El estudio de los métodos de aproximación de superficies se derivan de extender dimensionalmente la aproximación de curvas (normalmente mediante productos tensoriales), por lo que el foco de estudio se puede concentrar específicamente en los métodos bidimensionales para obtener modelos de curvas.

Normalmente este proceso de modelado comienza como un problema de ajuste de datos, donde los mismos provienen de ubicaciones geométricas (puntos de control) que restringen la forma de la curva. Los puntos de control unidos en secuencia por una poligonal constituyen el *grafo de control*, el cual es siempre el punto de partida para la síntesis de una curva (ver figura 1). La interpolación puede considerarse como un caso particular de aproximación, en el cual se agrega la restricción adicional de que la curva pase por cada punto del grafo de control. La idea clave para encontrar una función es considerar que la misma no puede ser cartesiana (de la forma $y = f(x)$) dado que esto no permite representar funciones multivaluadas. Luego, la representación de la función en términos de un parámetro u independiente surge como la alternativa más natural. De ese modo la forma de la función aproximadora es

$$\boldsymbol{C}(u) = \langle x(u), y(u) \rangle.$$

Por lo tanto se representa la curva como dos funciones paramétricas $x = x(u)$ e $y = y(u)$.

Uno de los esquemas de interpolación más directos consiste en unir entre sí distintos polinomios de bajo grado, de modo que en los puntos de control no solamente se interpola la posición sino también la derivada. De esa manera, la curva queda representada con un interpolante polinomial a trozos o segmentos. En la interpolación de Hermite, por ejemplo, se establece la condición de que el valor del polinomio al comienzo y al final de cada segmento debe coincidir con la posición de dos puntos de control sucesivos, haciéndose lo propio con la derivada del polinomio. De esa manera, con cuatro datos es posible despejar un polinomio cúbico $x(u) = au^3 + bu^2 + cu + d$ con $u \in [0, 1]$, que satisface las condiciones. Por lo tanto, además de los puntos de control, es necesario conocer o estimar las tangentes a la curva esperadas en dichos puntos. Además, como es fácil ver, no puede garantizarse continuidad en la segunda derivada. Estos polinomios son evaluados para una serie de valores de u entre 0 y 1, y se grafica la poligonal resultante.

Una forma geométricamente intuitiva de acercarse al problema es la siguiente. Dada una secuencia de puntos de control $\boldsymbol{P}_0, \boldsymbol{P}_1, \cdots, \boldsymbol{P}_n$, con $\boldsymbol{P}_i = \langle x_i, y_i \rangle$, la expresión $\boldsymbol{P}_i^1(u) = (1 - u)\boldsymbol{P}_i(u) + u\boldsymbol{P}_{i+1}$ representa el punto obtenido por la aplicación de una interpolación lineal entre dos puntos de control sucesivos $\boldsymbol{P}_i$ y $\boldsymbol{P}_{i+1}$ para un valor dado del parámetro $0 \leq u \leq 1$. La recurrencia

$$\boldsymbol{C}(u) = \boldsymbol{P}_i^r(u) = (1-u)\boldsymbol{P}_i^{r-1}(u) + u\boldsymbol{P}_{i+1}^{r-1}(u)$$

expresa la curva de grado n obtenida por la aplicación recursiva de la interpolación lineal entre todos los puntos sucesivos del grafo de control, luego a las interpolaciones de primer orden, luego a las de segundo, y así sucesivamente hasta alcanzar un único punto en el orden n para un valor dado del parámetro u. Este acercamiento geométrico al problema se conoce como **algoritmo de De Casteljau**.

Sin embargo, su expresión no da origen a una implementación computacionalmente eficiente, dado que su complejidad es de orden n^2. La idea, entonces, es encontrar una familia de polinomios que nos permita expresar a la curva como suma de polinomios, cuya expresión sea equivalente a la obtenida con el algoritmo de De Casteljau, pero de complejidad lineal en n. Esto permite encontrar una **curva de Bézier** definida como

$$\boldsymbol{C}(u) = \sum_{i=0}^{n} \boldsymbol{B}_i^n(u)\boldsymbol{P}_i,$$

donde la familia de polinomios $\boldsymbol{B}_i^n(u)$ es la base de Bernstein de orden n

$$\boldsymbol{B}_i^n(u) = \left(\begin{array}{c} n \\ i \end{array}\right) u^n(1-u)^{i-n},$$

con

$$\left(\begin{array}{c} n \\ i \end{array}\right) = \begin{cases} \frac{n!}{i!(n-i)!} & \text{si } 0 \leq i \leq n \\ 0 & \text{en todo otro caso} \end{cases}.$$

Estas curvas aproximan a todo el grafo de control con un único polinomio, y por lo tanto tienen un grado de continuidad infinito (al ser los polinomios infinitamente derivables) [2, 3]. Sin embargo, el grado de los mismos es muy elevado, requiriendo un alto costo computacional, y producen resultados que "suavizan" en exceso las variaciones geométricas en el grafo de control.

Tratando de combinar las ventajas de ambos métodos, es posible diseñar un esquema de aproximación en el cual la curva se aproxime con segmentos polinomiales de grado relativamente bajo, pero garantizando un orden de continuidad arbitrario. De esa manera surge el uso de bases funcionales para la aproximación de curvas con segmentos polinomiales, obteniéndose las bases de Splines o B-Splines [1, 2, 8]. Dichas bases se definen como integrales sucesivas de la función escalón. De esa manera, la interpolación lineal vista más arriba puede pensarse como un B-Spline de grado 1 (lineal) el cual es una función rampa, es decir, la integral de grado 1 de la función escalón, y que es continua pero no derivable (es decir, C^0). Generalizando, para obtener una curva C^n se requiere un B-Spline de grado n, para lo que hay que integrar recursivamente la base de grado anterior, obteniendo una función base de grado $n+1$. Dicha función base puede evaluarse con la expresión recursiva

$$\boldsymbol{N}_i^n(u) = \frac{u - u_{i-1}}{u_{i+N-1} - u_{i-1}} \boldsymbol{N}_i^{n-1}(u) + \frac{u_{i+N} - u}{u_{i+N} - u_i} \boldsymbol{N}_{i+1}^{n-1}(u),$$

donde los u_i son los *nudos* de la parametrización, y la curva queda expresada de una manera similar a la curva de Bézier:

$$\boldsymbol{C}(u) = \sum_{i=0}^{n} \boldsymbol{N}_i^n(u) \boldsymbol{P}_i,$$

La desventaja de este método es que la curva resultante, unión de segmentos polinomiales de grado $n+1$, normalmente no pasa por los puntos de control, y, como ya mencionáramos, en el análisis y diseño de curvas, los esquemas de interpolación polinomial a trozos son más adecuados y de mayor versatilidad que los esquemas de aproximación, esencialmente porque es posible localizar en el modelo los datos necesarios de una manera intuitiva. Por dicha razón, en la siguiente sección presentaremos un método de interpolación que intenta retener las ventajas geométricas de la aproximación con B-Splines y las ventajas interactivas de los esquemas de interpolación.

3 Interpolación con T-Splines

El punto de partida de este trabajo consiste en remarcar que todos los esquemas de aproximación e interpolación con segmentos de polinomios son reducibles a la aplicación sucesiva de curvas de Bézier, es decir, cada uno de los segmentos polinomiales de grado n es una curva de Bézier de grado n. El método de Hermite, por ejemplo, ubica puntos de control $\boldsymbol{P}_i$ y tangentes $\boldsymbol{m}_i$ sobre el modelo. El interpolante es un polinomio cúbico, y por lo tanto las curvas de Bézier cúbicas requieren cuatro puntos para ser definidas. De esa manera, cada punto de control $\boldsymbol{P}_i$ se transforma en un punto de unión entre dos curvas de Bézier (y por lo tanto llamado *punto de Bézier de unión* $\boldsymbol{b}_{3i}$. Los puntos faltantes, denominados *puntos interiores* $\boldsymbol{b}_{3i-1}$ y $\boldsymbol{b}_{3i+1}$ se encuentran desplazando $\frac{\boldsymbol{m}_i}{3}$ en forma negativa y positiva, respectivamente, al punto de control $\boldsymbol{P}_i$ [2]. Un interpolante entre dos puntos de control sucesivos $\boldsymbol{P}_i$ y $\boldsymbol{P}_{i+1}$ se encuentra como la curva de Bézier entre los puntos $\boldsymbol{b}_{3i}$, $\boldsymbol{b}_{3i+1}$, $\boldsymbol{b}_{3i+2}$ y $\boldsymbol{b}_{3(i+1)}$. Al estar unidas en los puntos de Bézier de unión (que son los puntos de control), se garantiza que la interpolación es contínua, y al ser el segmento que va de $\boldsymbol{b}_{3i-1}$ a $\boldsymbol{b}_{3i}$ igual al segmento que va de $\boldsymbol{b}_{3i}$ a $\boldsymbol{b}_{3i+1}$, entonces las derivadas de las curvas de Bézier son idénticas en los puntos de unión, garantizándose entonces continuidad C^1.

En el caso de los B-Spline uniformes[1], los puntos de Bézier interiores $\boldsymbol{b}_{3i+1}$ y $\boldsymbol{b}_{3i+2}$ se encuentran a un tercio y dos tercios (respectivamente) del segmento de recta que va de $\boldsymbol{P}_i$ a $\boldsymbol{P}_{i+1}$, y los puntos de Bézier de unión $\boldsymbol{b}_{3i}$ se encuentran a mitad de camino del segmento que va del punto interior $\boldsymbol{b}_{3i-1}$ al punto interior $\boldsymbol{b}_{3i+1}$. Esta construcción garantiza además que la segunda derivada es contínua porque el segmento que va de $\boldsymbol{P}_i$ a $\boldsymbol{b}_{3i+1}$, el que va de $\boldsymbol{b}_{3i+1}$ a $\boldsymbol{b}_{3i+2}$, y el que va de $\boldsymbol{b}_{3i+2}$ a $\boldsymbol{P}_{i+1}$ son iguales.

En este trabajo proponemos un esquema de interpolación en el cual los puntos de Bézier de unión $\boldsymbol{b}_{3i}$ son nuevamente los puntos de control $\boldsymbol{P}_i$, y los puntos interiores $\boldsymbol{b}_{3i+1}$ y $\boldsymbol{b}_{3i+2}$ se encuentran a partir de una estimación de las tangentes $\boldsymbol{m}_i$ en los puntos de control, pero respetando las condiciones geométricas expresadas para los B-Splines, de modo de garantizar continuidad C^2. Geométricamente, el problema de encontrar una interpolación polinomial C^2 contínua es equivalente al de encontrar puntos auxiliares $\boldsymbol{d}_i$ tales que

$$\boldsymbol{b}_{3i-1} = \frac{(1-\Delta_i)\boldsymbol{b}_{3i-2} + \Delta_i \boldsymbol{d}}{\Delta_{i+1} + \Delta_i},$$

[1]Es posible generalizar estos resultados para B-Spline no uniformes, pero los detalles no son necesarios para este trabajo

$$\boldsymbol{b}_{3\mathrm{i}+1} = \frac{(1-\Delta_i)\boldsymbol{d}Delta_i\boldsymbol{b}_{3\mathrm{i}+2}}{\Delta_{i+1}+\Delta_i}.$$

$$\Delta_i = u_{i+1} - u_i,$$

(i.e., Δ es la distancia paramétrica entre dos nudos sucesivos), De esa forma es posible despejar una parametrización de la curva en función de la subdivisión de la poligonal que une a los $\boldsymbol{d}_{\mathrm{i}}$ (ver figura 1).

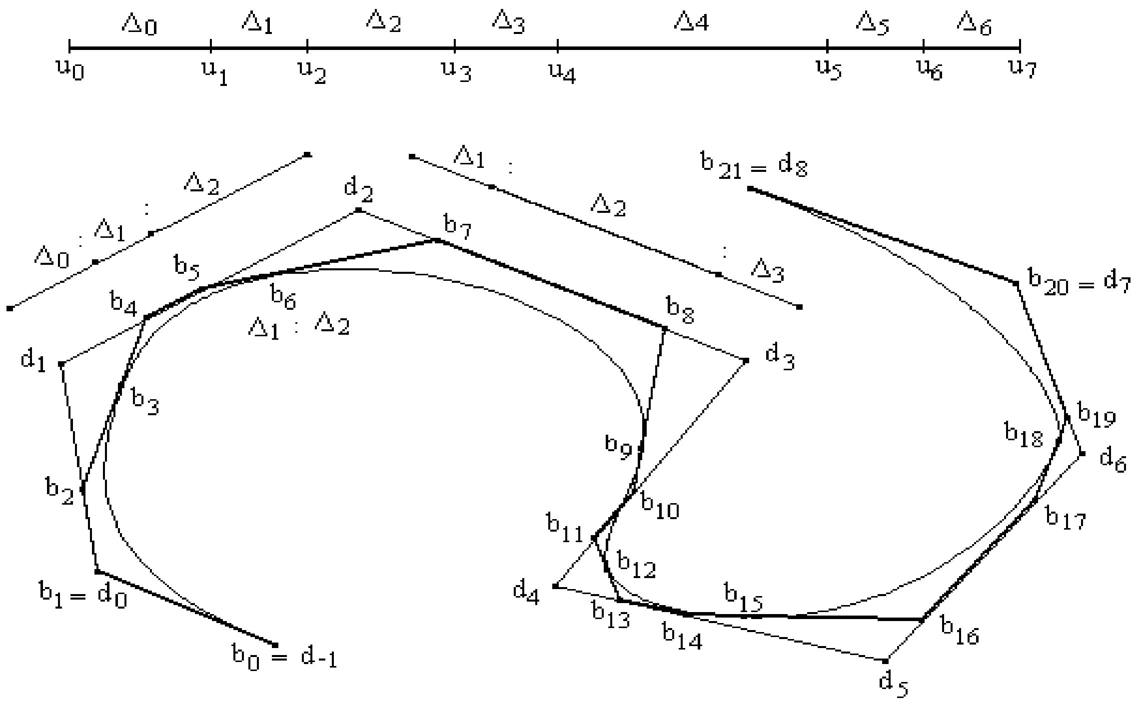

Figura 1: La estructura de control auxiliar y su parametrización.

Sin embargo, no es posible garantizar que dicha parametrización sea uniforme, por lo que es necesario evaluar las curvas cúbicas como B-Splines no uniformes entre secuencias de cuatro puntos auxiliares $\boldsymbol{d}_i$ sucesivos para garantizar C^2. Es decir, el problema de interpolar los puntos de control se transforma en el problema de aproximar los puntos auxiliares. Si, en cambio, se utilizan los puntos de Bézier de unión e interiores para obtener curvas de Bézier cúbicas, se garantiza continuidad G^2 pero no C^2 al no ser uniforme en general la asignación de valores de parámetro a los nudos. Es importante destacar que no siempre es posible encontrar una solución al problema geométrico planteado En aquellos casos en los que el segmento polinomial debe tener dos extremos entre puntos de control sucesivos, es fácil ver que no existe una secuencia de puntos auxiliares que permita representar el problema de interpolación como uno de aproximación.

4 Algunos resultados

En esta sección mostraremos muy brevemente las ventajas de la implementación interactiva de este método para el análisis y diseño de curvas. En la figura 2 es posible observar el resultado de ingresar varios puntos y sus derivadas asociadas, para encontrar interactivamente la curva interpolante. Tanto los puntos de union como las derivadas pueden ser modificadas interactivamente, hasta alcanzar el resultado deseado.

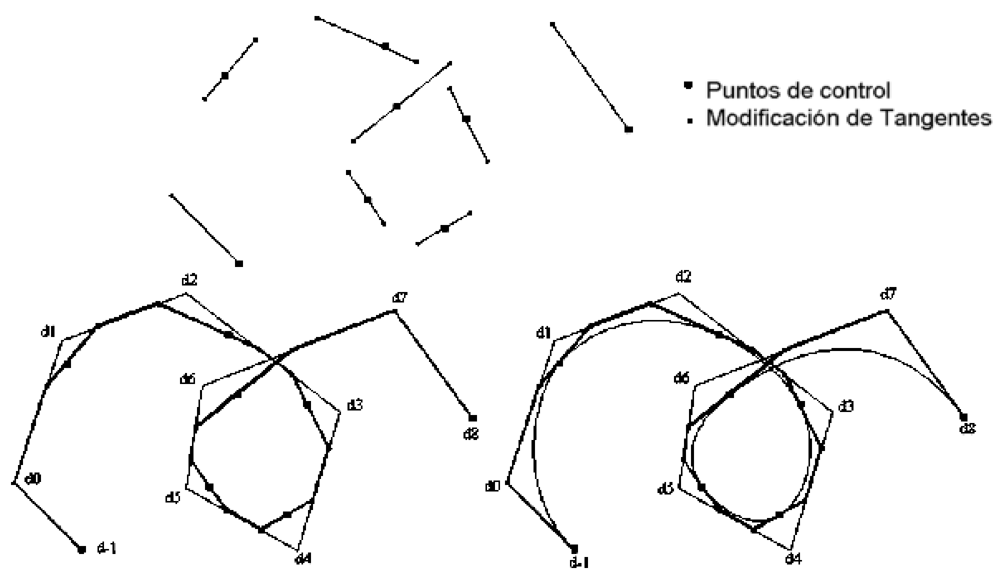

Figure 2: Modificación interactiva de curvas.

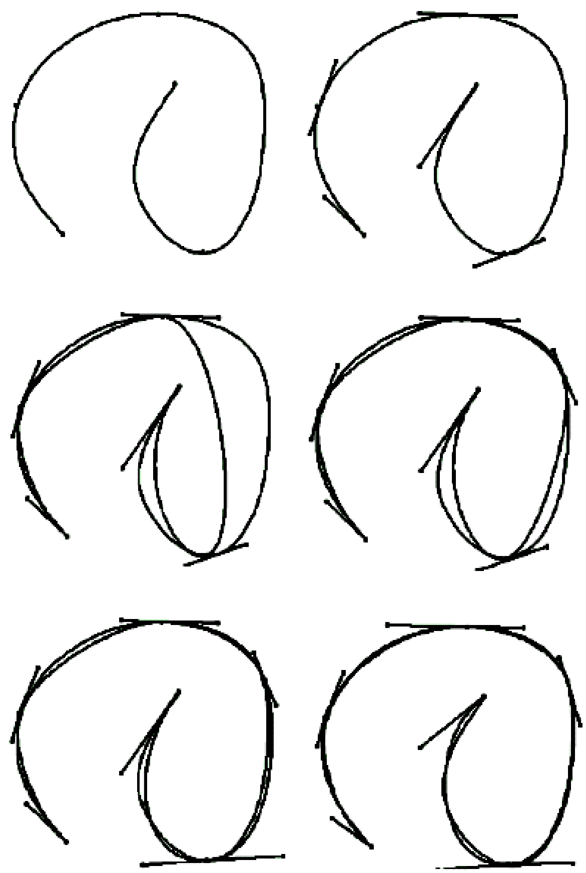

Figura 3: Interpolación automática de curvas.

Este proceso puede ser automatizado con bastante eficiencia, dado que es posible encontrar un conjunto de heurísticas que guíen al programa en el proceso de localizar los puntos de control adecuados y estimar las derivadas en dichos puntos. Una heurística de gran eficiencia consiste en localizar dichos puntos en los extremos del modelo que se busca analizar, y utilizar como estimación de las derivadas el vector que va de un punto n veces anterior al punto considerado a otro n veces posterior (ver figura 3). Entre el modelo y la curva resultante existirán diferencias, entre las cuales es posible aplicar recursivamente el método. Este proceso puede continuar hasta llegar a una representación de una complejidad dada o que se aparte del modelo por debajo de una cota de aproximación dada.

5 Conclusiones y trabajo futuro

El análisis o diseño de curvas por interpolación es indudablemente el más intuitivo de los esquemas. Además, en muchos problemas de ingeniería o arquitectura es necesario garantizar que las curvas resultantes pasen por localizaciones específicas, probablemente también con una tangente determinada. Sobre la base de un método de interpolación de curvas a partir de puntos de control, es posible el desarrollo de sistemas interactivos donde el usuario pueda modificar de una manera directa e intuitiva la forma y posición de la curva, o es posible encontrar una interpolación automática. Sin embargo, los métodos de interpolación tienen desventajas relacionadas con el orden de continuidad esperable. El esquema de interpolación de Hermite y otros métodos de interpolación con trozos de curvas polinomiales, por su parte, solamente garantizan un orden de continuidad C^1, es decir, curvas contínuas y derivables, pero discontínuas en su segunda derivada.

En este trabajo propusimos un método de interpolación, denominado T-Splines, que supera dicha deficiencia, garantizando curvas cuya segunda derivada geométrica es contínua. El mismo se basa en encontrar un grafo de control auxiliar, sobre el que se realiza una parametrización no uniforme de la curva, de modo que se satisfagan las restricciones geométricas que garantizen que el B-Spline no uniforme que aproxima a dicho grafo auxiliar interpola al mismo tiempo los puntos de control con las derivadas especificadas. Se desarrollaron algunos ejemplos y se describió un programa interactivo que permite el análisis y diseño de curvas con el esquema descripto. Por último, de discutieron algunos detalles de un procedimiento que permite elaborar una interpolación de curvas en forma automática y con un grado de precisión arbitrario.

Referencias

[1] R. Bartels, J. Beatty, and B. Barsky. *An Introduction to Splines for Use in Computer Graphics and Geometric Modelling.* Springer-Verlag, New York, 1987.

[2] Gerald Farin. *Curves and Surfaces for Computer Aided Geometric Design.* Academic Press, New York, 1988.

[3] J. C. Fiorot and P. Jeannin. *Rational Curves and Surfaces: Applications to CAD.* Wiley, New York, 1989.

[4] J. Foley, A. Van Dam, S. Feiner, and J. Hughes. *Computer Graphics. Principles and Practice.* Addison-Wesley, Reading, Massachussetts, second edition, 1990.

[5] W. K. Giloi. *Interactive Computer Graphics - Data Structures, Algorithms, Languages.* Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[6] W. Newman and R. Sproull. *Principles of Interactive Computer Graphics.* McGraw-Hill, New York, 1973.

[7] Les Piegl. Infinite Control Points: A Method for Representing Surfaces using Boundary Data. *IEEE Computer Graphics and Applications*, 7(3):45–55, 1987.

[8] L. Ramshaw. Blossoming: a conect-the-dots approach to splines. Technical Report TR-19, DIGITAL Systems Research Center, June 1987.

[9] Alan Watt and Mark Watt. *Advanced Animation and Rendering Techniques.* Addison-Wesley, London, 1992.