

9

Representación de curvas y superficies

La clásica tetera, presentada en la figura 9.1, es quizás el icono más conocido de la graficación por computador. Desde que Martin Newell [CROW87] la modeló en 1975, ha sido utilizada por docenas de investigadores como una estructura para demostrar las técnicas más recientes en la producción de superficies y texturas realistas. Para modelar la elegante tetera se tuvo que especificar su forma como una colección de elementos superficiales suaves, conocidos como **parches bicúbicos**. En muchas aplicaciones de la graficación por computador es necesario generar curvas y superficies suaves. Gran número de los objetos reales son inherentemente suaves y gran parte de la graficación por computador está orientada al modelado del mundo real. El diseño asistido por computador (CAD), los tipos de letra de alta calidad, los gráficos de datos y los dibujos de artistas contienen curvas y superficies suaves. La trayectoria de una cámara o un objeto en una secuencia de animación casi siempre es suave; así mismo, una trayectoria



Figura 9.1 La famosa *tetera*, modelo que consiste en un conjunto de superficies curvas suaves.

a través de un espacio de intensidades o colores (Caps. 11 y 12) por lo general debe ser suave.

La necesidad de representar curvas y superficies surge en dos casos: al modelar objetos existentes (un automóvil, un rostro o una montaña) y al modelar *a partir de cero*, cuando no se representa un objeto físico previamente existente. En el primer caso es posible que no exista una descripción matemática del objeto. Por supuesto, podemos usar como modelo las coordenadas de la infinidad de puntos en el objeto, pero esta estrategia no es factible en un computador con capacidad de almacenamiento limitada. Es más usual aproximar el objeto con pedazos de planos, esferas y otras formas que sean fáciles de describir de manera matemática, lo cual requiere que los puntos en el modelo estén cerca de los puntos correspondientes en el objeto.

En el segundo caso, cuando no hay un modelo previo que modelar, el usuario crea el objeto durante el proceso de modelado; por consiguiente, el objeto equivale exactamente a su representación, ya que ésta es su única encarnación. Para crear el objeto, el usuario puede esculpirlo en forma interactiva, describirlo matemáticamente o proporcionar una descripción aproximada que será *completada* por un programa. En CAD, la representación por computador se usa posteriormente para generar realizaciones físicas del objeto diseñado en forma abstracta.

En este capítulo se presenta el área general del **modelado de superficies**. Se trata de un área muy extensa, y sólo presentaremos las tres representaciones más comunes de superficies tridimensionales: superficies de malla poligonal, superficies paramétricas y superficies cuádricas. También analizaremos las curvas paramétricas, ya que son un concepto interesante y porque las superficies paramétricas constituyen una generalización de las curvas.

El **modelado de sólidos**, que se presenta en el capítulo siguiente, es la representación de volúmenes completamente rodeados por superficies, como son un cubo, un aeroplano o un edificio. La representación de superficies que se analiza en este capítulo puede utilizarse en el modelado de sólidos para definir las superficies que acotan el volumen.

Una **malla poligonal** es un conjunto de superficies planas limitadas por polígonos conectados entre sí. Las cajas abiertas, los gabinetes y los exteriores de edificios se pueden representar fácil y naturalmente con mallas poligonales, lo mismo que los volúmenes acotados por superficies planas. Las mallas poligonales también se pueden utilizar, aunque con menor facilidad, para representar objetos con superficies curvas, como en la figura 9.2; sin embargo, esta representación sólo es aproximada. En la figura 9.3 se presenta la sección transversal de una forma curva y la malla poligonal que la representa. Podemos hacer que los errores de la representación sean arbitrariamente pequeños si usamos más polígonos para crear mejores aproximaciones lineales por trozos, pero esta estrategia aumenta los requisitos de espacio y el tiempo de ejecución de los algoritmos que procesan la representación. Así mismo, al ampliar la imagen vuelven a ser obvias las aristas rectas.

Las **curvas polinomiales paramétricas** definen puntos en una curva tridimensional usando tres polinomios en un parámetro t para x , y y z . Los coefi-

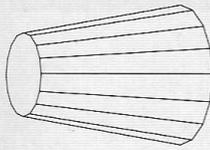


Figura 9.2
Objeto tridimensional representado con polígonos.

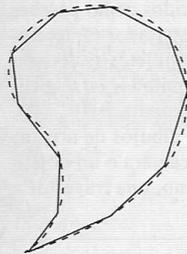


Figura 9.3
Sección transversal de una forma curva (línea punteada) y su representación poligonal (líneas sólidas).

cientes de los polinomios se seleccionan de manera que la curva siga la trayectoria deseada. Aunque se pueden emplear polinomios de diversos grados, sólo presentaremos el caso más común: los polinomios cúbicos (cuyos parámetros pueden tener hasta la tercera potencia). Con frecuencia utilizaremos el término **curva cúbica** para estas curvas.

Los **parches de superficie polinomial paramétrica de dos variables** definen las coordenadas de los puntos en una superficie curva usando tres polinomios de dos variables, que corresponden a x , y y z . Las fronteras de los parches son curvas polinomiales paramétricas. Se requieren mucho menos parches de superficie polinomial de dos variables que parches poligonales para aproximar una superficie curva con una precisión determinada. Sin embargo, los algoritmos para trabajar con los polinomios de dos variables son más complejos que los que se utilizan para polígonos. Como ocurre con las curvas, se pueden emplear polinomios de varios grados, pero aquí sólo analizaremos el caso común de los polinomios en los cuales ambos parámetros son cúbicos. Estas superficies se conocen como **superficies bicúbicas**.

Las **superficies cuádricas** son aquellas que se definen implícitamente con una ecuación $f(x, y, z) = 0$, donde f es un polinomio cuádrico en x , y y z . Las superficies cuádricas constituyen una representación práctica de la esfera, el elipsoide y el cilindro.

En el capítulo 10, acerca del modelado de sólidos, se incorporan estas representaciones a sistemas para representar no sólo superficies, sino también volúmenes acotados (sólidos). Las representaciones de superficies que se describen en este capítulo en ocasiones se usan combinadas con otras para acotar un volumen tridimensional.

9.1 Mallas poligonales

Una **mallá poligonal** es una colección de aristas, vértices y polígonos conectados de manera que cada arista esté compartida a lo sumo por dos polígonos. Una arista conecta dos vértices y un polígono es una secuencia cerrada de aristas. Una arista puede ser compartida por dos polígonos adyacentes, un vértice es compartido al menos por dos aristas y cada arista forma parte de *algún* polígono. Una mallá poligonal se puede representar en varias formas, cada una con sus ventajas y desventajas. La tarea del programador de la aplicación es elegir la representación más apropiada. En una sola aplicación se pueden emplear varias representaciones: una para el almacenamiento externo, otra para uso interno y otra más para que el usuario pueda crear la mallá en forma interactiva.

Para evaluar las representaciones se pueden emplear dos criterios básicos: espacio y tiempo. Las operaciones típicas con una mallá poligonal comprenden la determinación de todas las aristas incidentes a un vértice, detectar los polígonos que comparten una arista o un vértice, hallar los vértices conectados por

una arista, encontrar las aristas de un polígono, dibujar la malla e identificar errores en la representación (p. ej., aristas, vértices o polígonos faltantes). En términos generales, cuanto más explícitamente se representen las relaciones entre polígonos, vértices y aristas, más rápidas serán las operaciones y más espacio necesitará la representación. Woo [WOO85] ha analizado la complejidad temporal de nueve operaciones básicas de acceso y nueve operaciones básicas de actualización con estructuras de datos de malla poligonal.

En las secciones 9.1.1 y 9.1.2 se analizan varios temas relacionados con las mallas poligonales: la representación de mallas poligonales, asegurar que la representación sea correcta y calcular los coeficientes del plano de un polígono.

9.1.1 Representación de mallas poligonales

En esta sección analizaremos tres representaciones de mallas poligonales: explícitas, apuntadores a una lista de vértices y apuntadores a una lista de aristas. En la **representación explícita**, cada polígono se representa con una lista de coordenadas de vértices:

$$P = ((x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)).$$

Los vértices se almacenan en el orden en que los detectaríamos en un recorrido por el polígono. Hay aristas entre los vértices sucesivos en la lista, así como entre el primer y el último vértices. En el caso de un solo polígono, esta representación es eficiente en cuanto a espacio; sin embargo, en el caso de una malla poligonal se desperdicia mucho espacio ya que las coordenadas de los vértices compartidos se duplican. Un problema aún mayor es que no hay una representación explícita de las aristas y los vértices compartidos. Por ejemplo, para arrastrar interactivamente un vértice y todas sus aristas incidentes, es necesario hallar todos los polígonos que comparten el vértice. Para esta búsqueda se requiere la comparación de los triples de coordenadas de un polígono con los de los demás polígonos. La forma más eficiente de hacerlo sería ordenando los triples de coordenadas, pero este proceso requiere, en el mejor de los casos, $N \log_2 N$ pasos y no se elimina el peligro de que el mismo vértice pueda tener, debido al redondeo durante los cálculos, valores ligeramente distintos en las coordenadas para cada polígono, de tal forma que quizá nunca se detectaría una equivalencia exacta.

Con esta representación, para presentar la malla como polígonos rellenos o huecos es necesario transformar cada vértice y recortar cada arista de cada polígono. Si se dibujan las aristas, cada arista compartida se dibuja dos veces, lo cual ocasiona problemas en los graficadores de pluma, las grabadoras de película y las pantallas vectoriales, debido a la sobreescritura. También puede surgir un problema en las pantallas de barrido si las aristas se dibujan en direcciones opuestas, en cuyo caso pueden intensificarse pixeles adicionales.

Los polígonos definidos con **apuntadores a una lista de vértices**, el método que utiliza SPHIGS, almacenan sólo una vez cada vértice de la malla poligonal.

en la lista de vértices $V = ((x_1, y_1, z_1), \dots, (x_n, y_n, z_n))$. Un polígono se define con una lista de índices (o apuntadores) a la lista de vértices. Así, un polígono formado por los vértices 3, 5, 7 y 10 de la lista se representaría como $P = (3, 5, 7, 10)$.

En la figura 9.4 se presenta un ejemplo de esta representación, la cual tiene varias ventajas con respecto a la representación poligonal explícita. Como cada vértice se almacena una sola vez, se ahorra bastante espacio. Además, las coordenadas de un vértice se pueden modificar con facilidad. Por otra parte, sigue siendo difícil encontrar polígonos que compartan una arista y aristas compartidas que se dibujen dos veces al presentar todos los polígonos. Estos problemas se pueden eliminar representando explícitamente las aristas, como se hace en el método siguiente.

Al definir polígonos con **apuntadores a una lista de aristas**, una vez más tenemos la lista de vértices V , pero un polígono no se representa con una lista de apuntadores a la lista de vértices, sino a una lista de aristas, donde cada arista se incluye una sola vez. A su vez, cada arista de la lista apunta a los dos vértices de la lista de vértices que definen la arista y además a los polígonos (uno o dos) a los que pertenece la arista. De esta manera, un polígono se describe como $P = (E_1, \dots, E_n)$ y una arista como $E = (V_1, V_2, P_1, P_2)$. Cuando una arista sólo pertenece a un polígono, P_1 o P_2 será nulo. En la figura 9.5 se presenta un ejemplo de esta representación.

Los polígonos huecos se muestran presentando todas las aristas, en lugar de todos los polígonos; de esta manera se evitan recortes, transformaciones y discretizaciones redundantes. Los polígonos rellenos también son fáciles de presentar. En algunas situaciones, como en la descripción de una estructura de hoja metálica en forma de panal tridimensional, algunas de las aristas son compartidas por tres polígonos. En estos casos, las descripciones de las aristas se pueden extender para incluir un número arbitrario de polígonos: $E = (V_1, V_2, P_1, P_2, \dots, P_n)$.

En ninguna de estas tres representaciones (es decir, polígonos explícitos, apuntadores a vértices y apuntadores a una lista de aristas) es sencillo determinar cuáles son las aristas incidentes a un vértice: hay que inspeccionar todas las aristas. Por supuesto, se puede añadir información explícitamente para permitir la determinación de estas relaciones. Por ejemplo, la representación de arista alada que usa Baumgart [BAUM75] expande la descripción de las aristas para

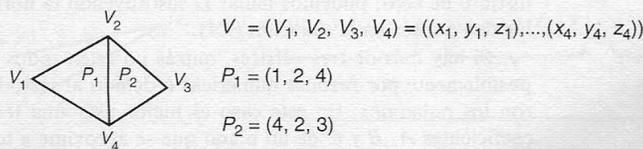


Figura 9.4 Malla poligonal definida con índices en una lista de vértices.

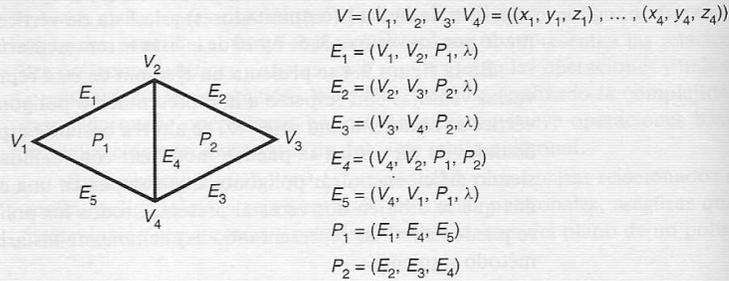


Figura 9.5 Malla poligonal definida con listas de aristas para cada polígono (λ representa un valor nulo).

incluir apuntables a los dos vértices adyacentes de cada polígono, mientras que la descripción de los vértices incluye un apuntable a una arista incidente (arbitraria) al vértice, con lo cual se obtiene más información acerca de los polígonos y los vértices.

9.1.2 Ecuaciones de planos

Al trabajar con polígonos o con mallas poligonales, muchas veces es necesario conocer la ecuación del plano donde se encuentra el polígono. Por supuesto, en algunos casos se conoce la ecuación en forma implícita, a través de los métodos de construcción interactivos que se usan para definir el polígono. Si no se conoce la ecuación, es posible utilizar las coordenadas de los tres vértices para encontrar el plano. Recuerde la ecuación de un plano:

$$Ax + By + Cz + D = 0. \quad (9.1)$$

Los coeficientes A , B y C definen la normal al plano, $[A \ B \ C]$. Dados los puntos P_1 , P_2 y P_3 en el plano, la normal al plano se puede calcular como el producto cruz de vectores $P_1P_2 \times P_1P_3$ (o $P_2P_3 \times P_2P_1$, etc.). Si el producto cruz es cero, los tres puntos son colineales y no definen un plano. En este caso se pueden usar otros vértices, si los hay. Una vez que se ha obtenido un producto cruz distinto de cero, podemos hallar D sustituyendo la normal a $[A \ B \ C]$ y uno de los tres puntos en la ecuación (9.1).

Si hay más de tres vértices, quizás no estén todos sobre un mismo plano, posiblemente por razones numéricas o debido al método por el cual se generaron los polígonos. En este caso es mejor usar una técnica que encuentre los coeficientes A , B y C de un plano que se aproxime a todos los vértices y no el método del producto cruz. Se puede demostrar que A , B y C son proporcionales a las áreas con signo de las proyecciones del polígono sobre los planos (y, z) , (z, x) y (x, y) , respectivamente. Por ejemplo, si el polígono es paralelo al plano

(x, y) , entonces $A = B = 0$, como era de esperar: las proyecciones del polígono sobre los planos (y, z) y (z, x) tienen área igual a cero. Este método es mejor que el de producto cruz porque las áreas de las proyecciones son función de las coordenadas de todos los vértices y por ende no pueden ser afectadas por la elección de unos cuantos vértices que quizás no sean coplanares con los demás vértices o con la mayoría de ellos, o que posiblemente sean colineales. Por ejemplo, el área (y por lo tanto el coeficiente) C del polígono proyectado sobre el plano (x, y) en la figura 9.6 es el área del trapecoide A_3 menos las áreas de A_1 y A_2 . Por lo general,

$$C = \frac{1}{2} \sum_{i=1}^n (y_i + y_{i \oplus 1})(x_{i \oplus 1} - x_i), \quad (9.2)$$

donde el operador \oplus es la suma usual excepto que $n \oplus 1 = 1$. Las áreas para A y B se expresan en forma similar, excepto que el área de B está negada (véase el Ej. 9.1).

La ecuación (9.2) proporciona la suma de las áreas de todos los trapecoides formados por aristas sucesivas de los polígonos. Si $x_{i \oplus 1} < x_i$, el área contribuye en forma negativa a la suma. El signo de la suma también es útil: si los vértices han sido numerados en el sentido del giro de las manecillas del reloj (según la proyección sobre el plano), el signo será positivo; en caso contrario, será negativo.

Una vez que determinamos la ecuación del plano usando todos los vértices, podemos estimar la no planaridad del polígono calculando la distancia perpendicular del plano a cada uno de los vértices. Esta distancia d para el vértice en (x, y, z) es

$$d = \frac{Ax + By + Cz + D}{\sqrt{A^2 + B^2 + C^2}}. \quad (9.3)$$

Esta distancia es positiva o negativa, dependiendo del lado del plano donde se encuentre el punto. Si el vértice está sobre el plano, $d = 0$. Por supuesto,

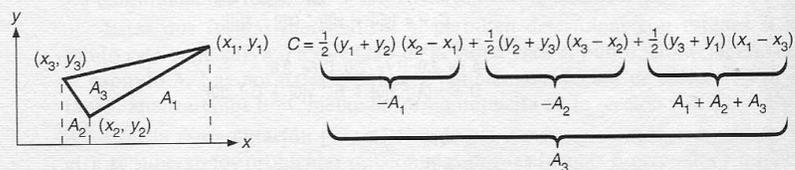


Figura 9.6 Cálculo del área C de un triángulo usando la ecuación (9.2).

para determinar en qué lado del plano está un punto, lo único importante es el signo de d , y la división entre la raíz cuadrada no es necesaria.

La ecuación del plano no es única; cualquier constante multiplicadora distinta de cero cambia la ecuación, pero no el plano. Muchas veces es conveniente almacenar los coeficientes del plano con una normal normalizada; esto se puede hacer escogiendo

$$k = \frac{1}{\sqrt{A^2 + B^2 + C^2}}, \quad (9.4)$$

que es el recíproco de la longitud de la normal. Es más fácil calcular las distancias con la ecuación (9.3), ya que el denominador es 1.

Ejemplo 9.1

Problema: Escriba una función que calcule los coeficientes de la ecuación del plano dados n vértices de un polígono que es aproximadamente plano. Suponga que los vértices del polígono se numeran en sentido contrario al giro de las manecillas, vistos hacia el plano desde su lado positivo. Los vértices y el número de vértices son argumentos que se pasan a la función.

Respuesta: Usando la ecuación (9.2) y ecuaciones similares para A y B , el programa es:

Esta función calcula los coeficientes de la ecuación del plano.

```
hallar_coeficientes_plano (float x[ ], float y[ ], float z[ ], int, 'núm_vértices,
                        float *a, float *b, float *c, float *d)
{
    float A, B, C, D;
    int i, j;

    A = B = C = 0.0;
    for (i = 0; i < núm_vértices; i++) {
        j = (i + 1) % núm_vértices;
        A += (z[i] + z[j]) * (y[j] - y[i]);
        B += -(x[i] + x[j]) * (z[j] - z[i]);
        C += (y[i] + y[j]) * (x[j] - x[i]);
    }
    A /= 2.0; B /= 2.0; C /= 2.0;
    D = -(A * x[0] + B * y[0] + C * z[0]);

    *a = A;
    *b = B;
    *c = C;
    *d = D;
}
```

9.2 Curvas cúbicas paramétricas

Las polilíneas y los polígonos son aproximaciones por trozos de primer grado de curvas y superficies, respectivamente. A menos que las curvas o superficies que se aproximan también sean lineales por trozos, hay que crear y almacenar gran cantidad de coordenadas de puntos extremos para lograr una precisión razonable. La manipulación interactiva de los datos para aproximar una curva es algo tediosa, ya que hay que ubicar con precisión gran cantidad de puntos.

En esta sección desarrollamos una representación más compacta y manejable de las curvas suaves por trozos; en la sección 9.3 se generaliza el desarrollo matemático a las superficies. El método general consiste en emplear funciones que sean de un grado mayor que el de las funciones lineales. Estas funciones por lo general sólo aproximan la forma deseada, pero requieren menos espacio de almacenamiento y ofrecen mayor facilidad de manipulación interactiva que las funciones lineales.

Las aproximaciones de mayor grado se pueden basar en uno de tres métodos. Primero, podemos expresar y y z como funciones *explícitas* de x , de manera que $y = f(x)$ y $z = g(x)$. Los problemas con este método son los siguientes: (1) es imposible obtener valores múltiples de y para una sola x , de manera que las curvas como los círculos y las elipses se tienen que representar con varios segmentos de curva; (2) esta definición no es rotacionalmente invariante (para describir una versión rotada de la curva se requiere mucho trabajo y por lo general hay que dividir un segmento de curva en varios segmentos más); y (3) la descripción de curvas con tangentes verticales es muy difícil, ya que resulta complicado representar una pendiente infinita.

Segundo, podemos modelar las curvas como soluciones a ecuaciones *implícitas* de la forma $f(x, y, z) = 0$; este método está plagado de peligros. Primero, la ecuación dada puede tener más soluciones que las que deseamos. Por ejemplo, al modelar un círculo podríamos usar $x^2 + y^2 = 1$, lo cual está bien. Sin embargo, ¿cómo modelamos medio círculo? Es necesario añadir restricciones como $x \geq 0$, que no pueden estar contenidas en la ecuación implícita. Así mismo, si se unen dos segmentos de curva definidos de manera implícita, puede ser difícil determinar si sus direcciones tangentes concuerdan en el punto de unión. La continuidad de tangentes es crítica en muchas aplicaciones.

Estas dos formas matemáticas sí permiten determinar con rapidez si un punto está en la curva o a qué lado de la curva yace el punto, como hicimos en el capítulo 3. Las normales a la curva también son fáciles de calcular. Por lo tanto, analizaremos brevemente la forma implícita en la sección 9.4.

Con la **representación paramétrica** de las curvas, $x = x(t)$, $y = y(t)$, $z = z(t)$, se superan los problemas ocasionados por las formas funcionales e implícitas y se ofrece una variedad de atractivos que serán evidentes en lo que resta de este capítulo. Las curvas paramétricas reemplazan la utilización de pendientes geométricas (que pueden ser infinitas) con vectores tangente paramétricos (que,

como veremos, nunca pueden ser infinitos). En este caso, una curva se aproxima con una **curva polinomial por trozos** en lugar de la curva lineal por trozos que usamos en la sección 9.1. Cada segmento Q de la curva global está indicado por tres funciones, x , y y z , que son polinomios cúbicos en el parámetro t .

Los polinomios cúbicos son los que se emplean con mayor frecuencia, ya que los polinomios de grado menor no ofrecen mucha flexibilidad para controlar la forma de la curva y los de mayor grado pueden introducir ondulaciones indeseadas y además requieren más cálculos. Ninguna representación de menor grado permite que un segmento de curva interpole (pase por) dos puntos extremos especificados con derivadas específicas en cada punto extremo. Dado un polinomio cúbico con sus cuatro coeficientes, se usan cuatro incógnitas para resolver los coeficientes desconocidos. Los cuatro valores conocidos podrán ser los dos puntos extremos y las derivadas en estos puntos. Así mismo, los dos coeficientes de un polinomio de primer orden (línea recta) se determinan con los dos puntos extremos. En el caso de una línea recta, las derivadas en cada extremo se determinan con la propia línea y no se pueden controlar en forma independiente. En el caso de polinomios cuadráticos (de segundo grado), que por lo tanto tienen tres coeficientes, se pueden especificar los dos puntos extremos y otra condición, como la pendiente o un punto adicional.

Así mismo, las cúbicas paramétricas son las curvas de menor grado que no son planas en tres dimensiones. Este hecho se puede observar sabiendo que los tres coeficientes de un polinomio de segundo orden se pueden especificar por completo con tres puntos y que tres puntos definen un plano donde se encuentra el polinomio.

Las curvas de mayor grado requieren más condiciones para determinar los coeficientes y pueden *ondular* en formas difíciles de controlar. A pesar de estas complejidades, las curvas de mayor grado se emplean en aplicaciones —como el diseño de automóviles y aviones— donde es necesario controlar derivadas de mayor grado para crear superficies aerodinámicamente eficientes. De hecho, el desarrollo matemático de las curvas y superficies paramétricas muchas veces se expresa en función de un grado arbitrario n . En este capítulo, mantenemos n fijo en 3.

9.2.1 Características básicas

Los polinomios cúbicos que definen un segmento de curva $Q(t) = [x(t) \ y(t) \ z(t)]^T$ tienen la forma

$$\begin{aligned} x(t) &= a_x t^3 + b_x t^2 + c_x t + d_x, \\ y(t) &= a_y t^3 + b_y t^2 + c_y t + d_y, \\ z(t) &= a_z t^3 + b_z t^2 + c_z t + d_z, \quad 0 \leq t \leq 1. \end{aligned} \quad (9.5)$$

Para tratar con segmentos finitos de la curva, limitamos el parámetro t , sin perder generalidad, al intervalo $[0, 1]$.

Con $T = [t^3 \ t^2 \ t \ 1]^T$ y definiendo la matriz de coeficientes de los tres polinomios como

$$C = \begin{bmatrix} a_x & b_x & c_x & d_x \\ a_y & b_y & c_y & d_y \\ a_z & b_z & c_z & d_z \end{bmatrix}, \quad (9.6)$$

podemos reescribir la ecuación (9.5) como

$$Q(t) = [x(t) \ y(t) \ z(t)]^T = C \cdot T. \quad (9.7)$$

Esta representación es una forma compacta de expresar la ecuación (9.5).

En la figura 9.7 se presentan dos segmentos unidos de curvas paramétricas y sus polinomios; también se ilustra la capacidad de las paramétricas para representar fácilmente varios valores de y para un solo valor de x en el caso de polinomios que en sí sólo tienen un valor. (Esta figura de una curva, como todas las demás en esta sección, presenta curvas bidimensionales representadas por $[x(t) \ y(t)]^T$.)

Continuidad entre segmentos de curva. La derivada de $Q(t)$ es el **vector tangente** paramétrico de la curva. Al aplicar esta definición a la ecuación (9.7) obtenemos

$$\begin{aligned} \frac{d}{dt} Q(t) = Q'(t) &= \begin{bmatrix} \frac{d}{dt} x(t) & \frac{d}{dt} y(t) & \frac{d}{dt} z(t) \end{bmatrix}^T = \frac{d}{dt} C \cdot T = C \cdot [3t^2 \ 2t \ 1 \ 0]^T \\ &= [3a_x t^2 + 2b_x t + c_x \quad 3a_y t^2 + 2b_y t + c_y \quad 3a_z t^2 + 2b_z t + c_z]^T. \end{aligned} \quad (9.8)$$

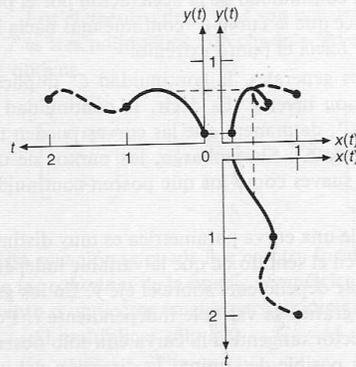


Figura 9.7 Dos segmentos unidos de curvas paramétricas bidimensionales y los polinomios que los definen. Las líneas punteadas entre el gráfico (x, y) y los gráficos $x(t)$ y $y(t)$ muestran la correspondencia entre los puntos en la curva (x, y) y los polinomios cúbicos de definición. Los gráficos $x(t)$ y $y(t)$ para el segundo segmento han sido trasladados para que inicien en $t = 1$, no en $t = 0$, para mostrar la continuidad de las curvas en su punto de unión.

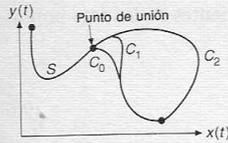


Figura 9.8

Segmento de curva S unido a los segmentos C_0 , C_1 y C_2 con grados 0, 1 y 2 respectivamente, de continuidad paramétrica. La diferencia visual entre C_1 y C_2 es muy pequeña cerca de la unión, pero notoria lejos de ella.

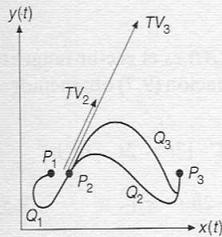


Figura 9.9

Los segmentos de curva Q_1 , Q_2 y Q_3 se unen en el punto P_2 y son idénticos excepto por sus vectores tangente en P_2 . Q_1 y Q_2 tienen vectores tangente idénticos y por lo tanto ambos tienen continuidad G^1 y C^1 en P_2 . Q_1 y Q_3 tienen vectores tangente en la misma dirección, pero Q_3 tiene el doble de magnitud, de manera que sólo tienen continuidad G^1 en P_2 . El vector tangente mayor de Q_3 significa que la curva es jalada más hacia la dirección del vector tangente antes de dirigirse hacia P_3 . El vector TV_2 es el vector tangente de Q_2 mientras que TV_3 es el de Q_3 .

Si dos segmentos de curva se unen, la curva tiene **continuidad geométrica** G^0 . Si las direcciones (aunque no necesariamente las magnitudes) de los vectores tangente de las dos curvas son iguales en el punto de unión, la curva tiene **continuidad geométrica** G^1 . En el diseño de objetos asistido por computador, con frecuencia se requiere continuidad G^1 entre segmentos de curva. La continuidad G^1 significa que las pendientes geométricas de los segmentos son iguales en el punto de unión. Para que dos vectores tangente TV_1 y TV_2 tengan la misma dirección, es necesario que uno sea múltiplo escalar del otro: $TV_1 = k \cdot TV_2$, donde $k > 0$ [BARS88].

Si los vectores tangente de dos segmentos de curva cúbica son iguales (es decir, si sus direcciones y sus magnitudes son iguales) en el punto de unión de los segmentos, la curva tiene **continuidad de primer grado** en el parámetro t , o **continuidad paramétrica**, y se dice que tiene **continuidad** C^1 . Si la dirección y la magnitud de $d^n/dt^n[Q(t)]$ hasta la n -ésima derivada son iguales en el punto de unión, se dice que la curva tiene **continuidad** C^n . En la figura 9.8 se presenta una curva con tres grados de continuidad diferentes. Observe que un segmento de curva paramétrica siempre es continuo en todos sus puntos; la continuidad que nos interesa aquí es la que existe en los puntos de unión.

El vector tangente $Q'(t)$ es la **velocidad** de un punto en la curva con respecto al parámetro t . En forma similar, la segunda derivada de $Q(t)$ es la **aceleración**. Suponga que se mueve una cámara por la curva cúbica paramétrica en incrementos de tiempo iguales y que se toma una fotografía después de cada incremento; el vector tangente indica la velocidad de la cámara a lo largo de la curva. Para evitar los movimientos irregulares en la secuencia de animación, la velocidad de la cámara y la aceleración en los puntos de unión deben ser continuas. Esta continuidad de la aceleración por el punto de unión de la figura 9.8 es la que hace que la curva C^2 continúe más hacia la derecha que la curva C^1 antes de doblar hacia el punto extremo.

En términos generales, la continuidad C^1 implica G^1 , pero no siempre se aplica la situación inversa. Es decir, la continuidad G^1 usualmente es **menos restrictiva** que C^1 , de manera que las curvas pueden tener continuidad G^1 pero no necesariamente C^1 . Sin embargo, los puntos de unión con continuidad G^1 aparecerán tan suaves como los que poseen continuidad C^1 , como puede verse en la figura 9.9.

El gráfico de una curva paramétrica es muy distinto del gráfico de una función ordinaria, en el sentido de que la variable independiente se grafica sobre el eje x y la variable dependiente sobre el eje y . En los gráficos de curvas paramétricas nunca se grafica la variable independiente t . Por lo tanto, no es posible determinar el vector tangente a la curva con sólo observar el gráfico de la curva paramétrica. Es posible determinar la dirección del vector, pero no su magnitud. Se podrá comprender por qué ocurre esto si se considera de la siguiente manera: si $\gamma(t)$, $0 \leq t \leq 1$, es una curva paramétrica, su vector tangente en el instante 0 es $\gamma'(0)$. Si $\eta(t) = \gamma(2t)$, $0 \leq t \leq 1/2$, los gráficos paramétricos de γ y η son idénticos. Por otra parte, $\eta'(0) = 2\gamma'(0)$. Por lo tanto, dos curvas con gráficos idénticos pueden tener vectores tangente diferentes. Este hecho es la

base de la definición de la **continuidad geométrica**: para que dos curvas se unan suavemente, sólo se requiere que concuerden las direcciones de sus vectores tangente; no es necesario que concuerden sus magnitudes.

Relación con las restricciones. Un segmento de curva $Q(t)$ se define con restricciones aplicables a los puntos extremos, vectores tangente y continuidad entre segmentos de curva. Cada polinomio cúbico de la ecuación (9.5) tiene cuatro coeficientes, de manera que se requieren cuatro restricciones, lo que nos permite formular cuatro ecuaciones para las cuatro incógnitas y resolver éstas. Los tres tipos principales de curvas que se analizan en esta sección son las de **Hermite** o **hermitianas**, definidas por dos puntos extremos y dos vectores tangente de los puntos extremos; las de **Bézier**, definidas por dos puntos extremos y otros dos puntos que controlan los vectores tangente de los puntos extremos; y varios tipos de **splines**, cada una de ellas definida por cuatro puntos de control. Las **splines** tienen continuidad C^1 y C^2 en los puntos de unión y se acercan a sus puntos de control, pero generalmente no los interpolan. Los tipos de **splines** son las **B-splines** uniformes y las **B-splines** no uniformes.

Para ver la forma en que los coeficientes de la ecuación (9.5) pueden depender de las cuatro restricciones, recordemos que una curva cúbica paramétrica se define con $Q(t) = C \cdot T$. Reescribimos la matriz de coeficientes como $C = G \cdot M$, donde M es una **matriz base** 4×4 y G es una matriz de cuatro elementos de restricciones geométricas, llamada **matriz geométrica**. Las restricciones geométricas son las condiciones, como los puntos extremos o los vectores tangente, que definen la curva. Usaremos G_x para referirnos al vector fila de los componentes x de la matriz geométrica; G_y y G_z tienen significado similar. G o M , o bien tanto G como M , difieren para cada tipo de curva.

Los elementos de G y M son constantes, de manera que el producto $G \cdot M \cdot T$ es tres polinomios cúbicos en t . Al expandir el producto $Q(t) = G \cdot M \cdot T$ se obtiene

$$Q(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = [G_1 \quad G_2 \quad G_3 \quad G_4] \begin{bmatrix} m_{11} & m_{21} & m_{31} & m_{41} \\ m_{12} & m_{22} & m_{32} & m_{42} \\ m_{13} & m_{23} & m_{33} & m_{43} \\ m_{14} & m_{24} & m_{34} & m_{44} \end{bmatrix} \begin{bmatrix} t^3 \\ t^2 \\ t \\ 1 \end{bmatrix}. \quad (9.9)$$

Esta ecuación se puede leer en otra forma: el punto $Q(t)$ es una suma ponderada de las **columnas** de la matriz geométrica G , cada una de las cuales representa un punto o un vector en el espacio tridimensional.

Al multiplicar sólo $x(t) = G_x \cdot M \cdot T$ se obtiene

$$x(t) = (t^3 m_{11} + t^2 m_{21} + t m_{31} + m_{41}) g_{1x} + (t^3 m_{12} + t^2 m_{22} + t m_{32} + m_{42}) g_{2x} \\ + (t^3 m_{13} + t^2 m_{23} + t m_{33} + m_{43}) g_{3x} + (t^3 m_{14} + t^2 m_{24} + t m_{34} + m_{44}) g_{4x}. \quad (9.10)$$

La ecuación (9.10) hace énfasis en que la curva es una suma ponderada de los elementos en la matriz geométrica. Las ponderaciones son polinomios cúbicos

de t y se denominan **funciones de mezcla**. Las funciones de mezcla B se obtienen de $B = M \cdot T$. Observe la similitud con la aproximación lineal por trozos, donde sólo se requieren dos restricciones geométricas (los puntos extremos de la línea), de manera que cada segmento de curva es una línea recta definida por los puntos extremos G_1 y G_2 :

$$\begin{aligned}x(t) &= g_{1x}(1-t) + g_{2x}(t), \\y(t) &= g_{1y}(1-t) + g_{2y}(t), \\z(t) &= g_{1z}(1-t) + g_{2z}(t),\end{aligned}\tag{9.11}$$

Las cúbicas paramétricas en realidad son sólo una generalización de las aproximaciones de líneas rectas. La curva cúbica $Q(t)$ es una combinación de las cuatro columnas de la matriz geométrica, así como un segmento de línea es una combinación de dos vectores columna.

Para ver cómo se calcula la matriz base M , pasamos ahora a formas específicas de curvas cúbicas paramétricas.

9.2.2 Curvas de Hermite

La forma de Hermite o hermitiana (llamada así en honor del matemático) de un segmento de curva polinomial cúbica está determinada por las restricciones en los puntos extremos P_1 y P_4 y los vectores tangente en los puntos extremos R_1 y R_4 . (Se usan los índices 1 y 4, en lugar de 1 y 2, por cuestiones de consistencia con secciones subsecuentes, donde se usarán los puntos intermedios P_2 y P_3 , en lugar de vectores tangente, para definir la curva).

Para hallar la **matriz base hermitiana** M_H , que relaciona el **vector geométrico hermitiano** G_H con los coeficientes de los polinomios, escribimos cuatro ecuaciones, una para cada restricción, en los cuatro coeficientes polinomiales desconocidos y resolvemos las incógnitas.

Si definimos G_H , el componente x de la matriz geométrica hermitiana, como

$$G_H = [P_1, P_4, R_1, R_4],\tag{9.12}$$

y reescribimos $x(t)$ de las ecuaciones (9.5) y (9.9) como

$$x(t) = a_x t^3 + b_x t^2 + c_x t + d_x = C_x \cdot T = G_H \cdot M_H \cdot T = G_H \cdot M_H [t^3 \ t^2 \ t \ 1]^T,\tag{9.13}$$

las restricciones de $x(0)$ y $x(1)$ se pueden encontrar por sustitución directa en la ecuación (9.13):

$$x(0) = P_1 = G_H \cdot M_H [0 \ 0 \ 0 \ 1]^T,\tag{9.14}$$

$$x(1) = P_4 = G_H \cdot M_H [1 \ 1 \ 1 \ 1]^T.\tag{9.15}$$

Así como en el caso general diferenciamos la ecuación (9.7) para hallar la ecuación (9.8), ahora diferenciamos la ecuación (9.13) para encontrar $x'(t) = G_{H_x} \cdot M_H [3t^2 \ 2t \ 1 \ 0]^T$. Por lo tanto, las ecuaciones de restricción de vectores tangente se pueden escribir como

$$x'(0) = R_{1_x} = G_{H_x} \cdot M_H [0 \ 0 \ 1 \ 0]^T, \tag{9.16}$$

$$x'(1) = R_{4_x} = G_{H_x} \cdot M_H [3 \ 2 \ 1 \ 0]^T. \tag{9.17}$$

Las cuatro restricciones de las ecuaciones (9.14)-(9.17) se pueden reescribir en forma matricial:

$$[P_1, \ P_4, \ R_1, \ R_4] = G_{H_x} = G_{H_x} \cdot M_H \cdot \begin{bmatrix} 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}. \tag{9.18}$$

Para que se satisfaga esta ecuación (y las expresiones correspondientes de y y z), M_H debe ser la inversa de la matriz de 4×4 en la ecuación (9.18):

$$M_H = \begin{bmatrix} 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & -3 & 0 & 1 \\ -2 & 3 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}. \tag{9.19}$$

Ahora podemos usar M_H en $x(t) = G_{H_x} \cdot M_H \cdot T$ para hallar $x(t)$ con base en el vector de geometría G_{H_x} . En forma similar, $y(t) = G_{H_y} \cdot M_H \cdot T$ y $z(t) = G_{H_z} \cdot M_H \cdot T$, por lo que podemos escribir

$$Q(t) = [x(t) \ y(t) \ z(t)]^T = G_H \cdot M_H \cdot T, \tag{9.20}$$

donde G_H es la matriz

$$[P_1 \ P_4 \ R_1 \ R_4].$$

Al expandir el producto $M_H \cdot T$ en $Q(t) = G_H \cdot M_H \cdot T$ se obtienen las **funciones de mezcla de Hermite** B_H como polinomios que ponderan cada elemento de la matriz de geometría:

$$Q(t) = G_H \cdot M_H \cdot T = G_H \cdot B_H \\ = (2t^3 - 3t^2 + 1)P_1 + (-2t^3 + 3t^2)P_4 + (t^3 - 2t^2 + t)R_1 + (t^3 - t^2)R_4. \tag{9.21}$$

En la figura 9.10 se muestran las cuatro funciones de mezcla. Observe que, en $t = 0$, sólo la función P_1 es distinta de cero: únicamente P_1 afecta la curva en

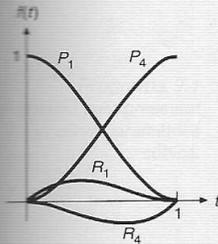


Figura 9.10 Las funciones de mezcla de Hermite, rotuladas con los elementos del vector de geometría que ponderan.