7
B-Spline Approximation

B-spline methods for curves and surfaces were first proposed in the 1940s but were seriously developed only in the 1970s, by several researchers, most notably R. Riesenfeld. They have been studied extensively, have been considerably extended since the 1970s, and much is currently known about them. The designation "B" stands for Basis, so the full name of this approach to curve and surface design is the basis spline. This chapter discusses the important types of B-spline curves and surfaces, including the most versatile one, the nonuniform rational B-spline (NURBS, Section 7.14).

The B-spline curve overcomes the main disadvantages of the Bézier curve which are (1) the degree of the Bézier curve depends on the number of control points, (2) it offers only global control, and (3) individual segments are easy to connect with C^1 continuity, but C^2 is difficult to obtain. The B-spline curve features local control and any desired degree of continuity. To obtain C^n continuity, the individual spline segments have to be polynomials of degree n. The B-spline curve is an approximating curve and is therefore defined by control points. However, in addition to the control points, the user has to specify the values of certain quantities called "knots." They are real numbers that offer additional control over the shape of the curve. The basic approach taken in the first part of this chapter ignores the knots, but they are introduced in Section 7.8 and their effect on the curve is explored.

There are several types of B-splines. In the *uniform* (also called periodic) B-spline (Sections 7.1 and 7.2), the knot values are uniformly spaced and all the weight functions have the same shape and are shifted with respect to each other. In the *nonuniform* B-spline (Section 7.11), the knots are specified by the user and the weight functions are generally different. There is also an *open uniform* B-spline (Section 7.10), where the knots are not uniform but are specified in a simple way. In a *rational* B-spline (Section 7.14), the weight functions are in the form of a ratio of two polynomials. In a *nonrational* B-spline, they are polynomials in t. The B-spline is an approximating curve based on control points, but there is also an *interpolating* version that passes through the points (Section 7.7). Section 7.4 shows how tension can be added to the B-spline.

B-splines are mathematically more sophisticated than other types of splines, so we start with a gentle introduction. We first use basic assumptions to derive the expressions for the quadratic and cubic uniform B-splines directly and without mentioning knots. We then show how to extend the derivations to uniform B-splines of any order. Following this, we discuss a different, recursive formulation of the weight functions of the uniform, open uniform, and nonuniform B-splines.

7.1 The Quadratic Uniform B-Spline

We start with the quadratic uniform B-spline. We assume that $n+1$ control points, $\mathbf{P}_0$, $\mathbf{P}_1,\ldots,$ $\mathbf{P}_n$, are given and we want to construct a spline curve where each segment $\mathbf{P}_i(t)$ is a quadratic parametric polynomial based on three points, $\mathbf{P}_{i-1}$, $\mathbf{P}_i$, and $\mathbf{P}_{i+1}$. We require that the segments connect with C^1 continuity (only cubic and higher-degree polynomial segments can have C^2 or higher continuities) and that the entire curve has local control. To achieve all this, we have to give up something and we elect to give up the requirement that a segment will pass through its first and last control points. We denote the start and end points of segment $\mathbf{P}_i(t)$ by $\mathbf{K}_i$ and $\mathbf{K}_{i+1}$, respectively and we call them *joint points*, or just *joints*. These points are still unknown and will have to be determined. Figure 7.1a shows two quadratic segments $\mathbf{P}_1(t)$ and $\mathbf{P}_2(t)$ defined by the four control points $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$. The first segment goes from joint $\mathbf{K}_1$ to joint $\mathbf{K}_2$ and the second segment goes from joint $\mathbf{K}_2$ to joint $\mathbf{K}_3$, where the joints are drawn tentatively and will have to be determined and redrawn. Note that each segment is defined by three control points, so its control polygon has two edges. The first spline segment is defined only by $\mathbf{P}_0$, $\mathbf{P}_1$, and $\mathbf{P}_2$, so any changes in $\mathbf{P}_3$ will not affect it. This is how local control is achieved in a B-spline.

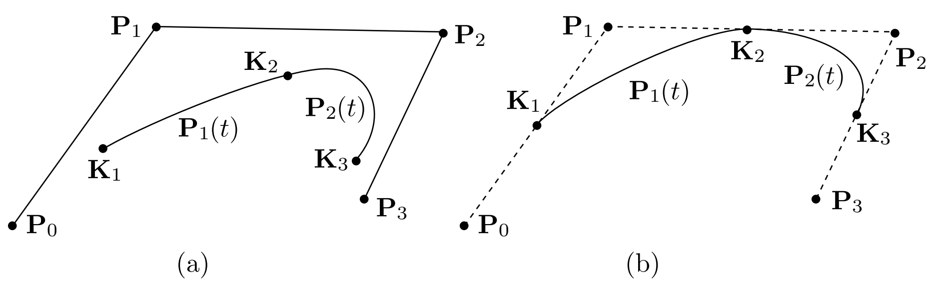

Figure 7.1: The Quadratic Uniform B-Spline.

We use the usual notation for the two segments

$$\mathbf{P}_i(t) = (t^2, t, 1)\mathbf{M}\begin{pmatrix}\mathbf{P}_{i-1}\\ \mathbf{P}_i\\ \mathbf{P}_{i+1}\end{pmatrix}, \quad i = 1, 2, \tag{7.1}$$