3.5 Bezier Curves and B-Spline Curves

In addition to the splines we have studied in the previous section, there are others that are important. In particular, Bezier curves and B-splines are widely used in computer graphics and computer-aided design. B-splines are often used to numerically integrate and differentiate functions that are defined only through a set of data points. These two types of curves are not really interpolating splines, because the curves do not normally pass through all of the points. In this respect they show some similarity to least-squares curves, which are discussed in a later section. However, both Bezier curves and B-splines have the important property of staying within the polygon determined by the given points. We will be more explicit about this property later. In addition, these two new spline curves have a nice geometric property in that in changing one of the points we change only one portion of the curve, a "local" effect. For the cubic spline curve of the previous section, changing just one point has a "global" effect in that the entire curve from the first to the last point is affected. Finally, for the cubic splines just studied, the points were given data points. For the two curves we study in this section the points in question are more likely "control" points that we select to determine the shape of the curve we are working on.

not.

1.

For simplicity, we consider mainly the cubic version of these two curves. In what follows, we will express $y = f(x)$ in parametric form. The parametric form represents a relation between x and y by two other equations, $x = F_1(u)$, $y = F_2(u)$. The independent variable u is called the *parameter*. For example, the equation for a circle can be written, with θ as the parameter, as

$$x = r\cos(\theta),$$
$$y = r\sin(\theta).$$

If we express x and y in terms of a parameter, u, the point (x, y) becomes $(x(u), y(u))$. We will use this with values of the parameter u between 0 and 1.

We discuss Bezier curves first. Bezier curves are named after the French engineer, P. Bezier of the Renault Automobile Company. He developed them in the early 1960s to fill a need for curves whose shape can be readily controlled by changing a few parameters. Bezier's application was to construct pleasing surfaces for car bodies.

Suppose we are given a set of control points, $p_i = (x_i, y_i)$, $i = 0, 1, \ldots, n$. (These points are also referred to as *Bezier points.*) Figure 3.6 is an example.

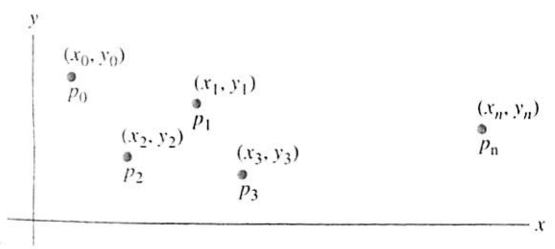


Figure 3.6

These points could be chosen on a computer screen, using a pointing device. The points do not necessarily progress from left to right. We treat the coordinates of each point as a two-component vector,

$$p_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}.$$

The set of points, in parametric form, is

$$P(u) = \begin{bmatrix} x(u) \\ y(u) \end{bmatrix}, \qquad 0 \le u \le 1.$$

The nth-degree Bezier polynomial determined by $n + 1$ points is given by

$$P(u) = \sum_{i=0}^{n} \binom{n}{i} (1 - u)^{n-i} u^i p_i,$$

where

$$\binom{n}{i} = \frac{n!}{i!(n - i)!}.$$

(The preceding formula really represents two other scalar equations, one for x_i and the other for y_i.) For $n = 2$, this would give the quadratic equation defined by three points, p_0, p_1, p_2:

$$P(u) = (1)(1 - u)^2 p_0 + 2(1 - u)(u)p_1 + (1)u^2 p_2,$$

because, for $n = 2$ and $i = 0, 1, 2$, we have $\binom{2}{0} = 1$, $\binom{2}{1} = 2$, $\binom{2}{2} = 1$. The preceding equation represents the pair of equations

$$x(u) = (1 - u)^2 x_0 + 2(1 - u)(u)x_1 + u^2 x_2,$$
$$y(u) = (1 - u)^2 y_0 + 2(1 - u)(u)y_1 + u^2 y_2.$$

Observe that, if $u = 0$, $x(0)$ is identical to x_0 and similarly for $y(0)$. If $u = 1$, the point referred to is (x_2, y_2). As u takes on values between 0 and 1, a curve is traced that goes from the first point to the third of the set. Ordinarily the curve will not pass through the central point of the three. (If the points are collinear, the curve is the straight line through them all.) In effect, the points of the second-degree Bezier curve have coordinates that are weighted sums of the coordinates of the three points that are used to define it. From another point of view, one can think of the Bezier equations as weighted sums of three polynomials in u, where the weighting factors are the coordinates of the three points.

Applying the general defining equation for $n = 3$, we get the cubic Bezier polynomial that we now consider in some detail. The properties of other Bezier polynomials are the same as for the cubic. Here is the Bezier cubic:

$$x(u) = (1-u)^3 x_0 + 3(1-u)^2 u x_1 + 3(1-u)u^2 x_2 + u^3 x_3,$$
$$y(u) = (1-u)^3 y_0 + 3(1-u)^2 u y_1 + 3(1-u)u^2 y_2 + u^3 y_3.$$

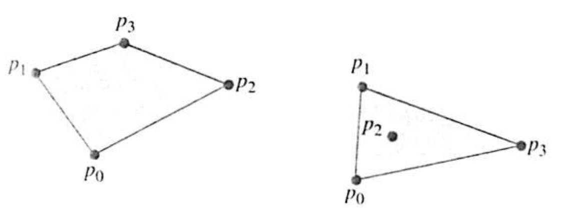
Observe again that $(x(0), y(0)) = p_0$ and $(x(1), y(1)) = p_3$, and that the curve will not ordinarily go through the intermediate points. As illustrated in the example curves in Fig. 3.7, changing the intermediate "control" points changes the shape of the curve. The examples are in Figs. 3.7(a) through 3.7(e). The first three of these show Bezier curves defined by one group of four points.

Figures 3.7(d) and 3.7(e) demonstrate how cubic Bezier curves can be continued beyond the first set of four points; one just subdivides seven points (p_0 to p_6) into two groups of four, with the central one (p_3) belonging to both sets. Figure 3.7(e) shows that p_2, p_3, and p_4 must be collinear to avoid a discontinuity in the slope at p_3.

It is of interest to list the properties of Bezier cubics:

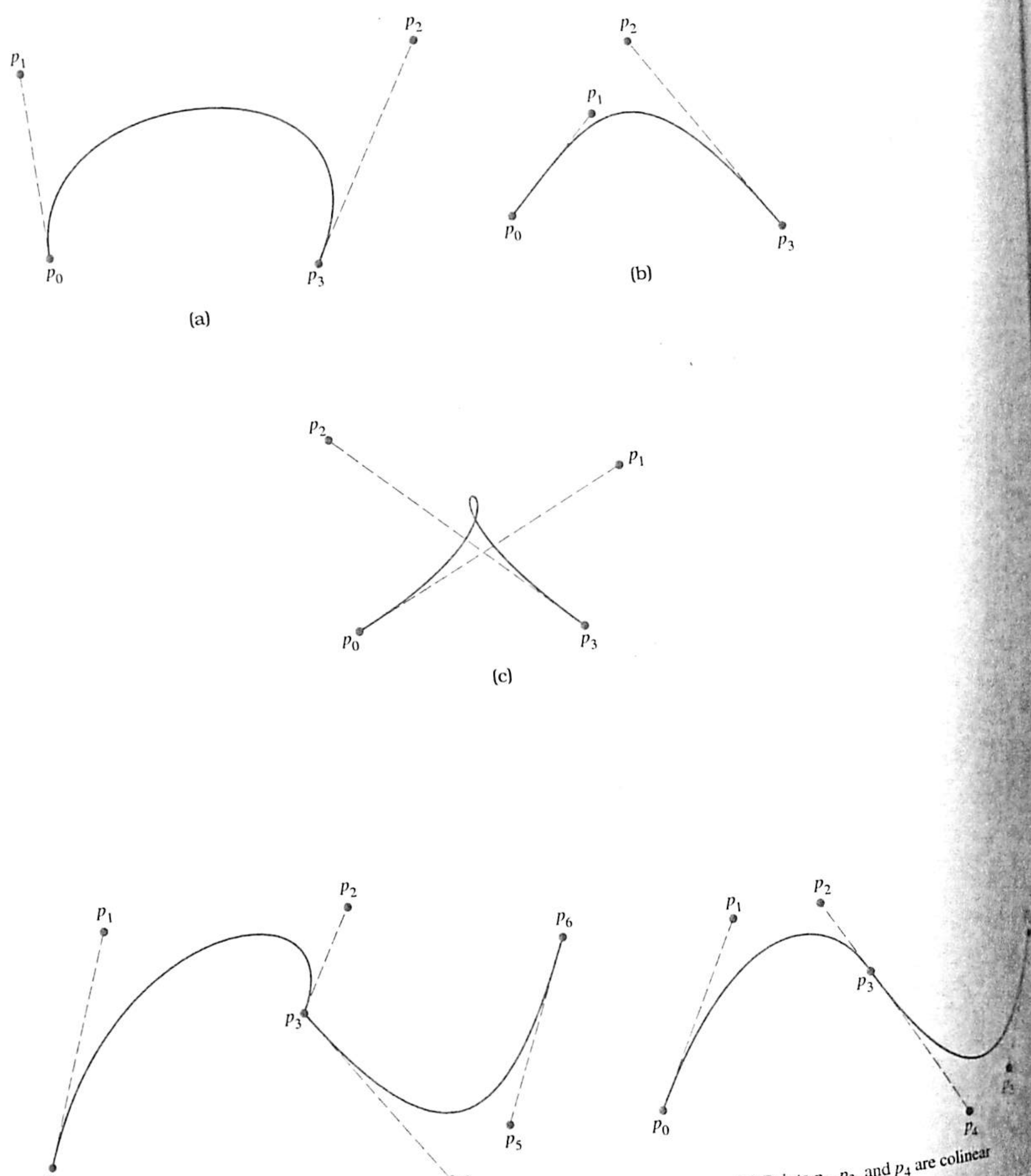
1. $P(0) = p_0$, $\quad P(1) = p_3$.
2. Because $dx/du = 3(x_1 - x_0)$ and $dy/du = 3(y_1 - y_0)$ at $u = 0$, the slope of the curve at $u = 0$ is $dy/dx = (y_1 - y_0)/(x_1 - x_0)$, which is the slope of the secant line between p_0 and p_1. Similarly, the slope at $u = 1$ is the same as the secant line between the last two points. This is indicated in the figures by dashed lines.
3. The Bezier curve is contained in the convex hull determined by the four points.

The *convex hull* of a set of points is the smallest convex set that contains the points. A set, C, is *convex* if and only if the line segment between any two points in the set lies entirely in set C. The following sketches show examples of the convex hull of four points.



It is often convenient to represent the Bezier curve in matrix form. For Bezier cubics, this is

$$P(u) = [u^3, u^2, u, 1]\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix}$$
$$= u^T M_z p.$$



(a)

(b)

(c)

(d) Points p_2, p_3, p_4 are not colinear

(e) Points p_2, p_3, and p_4 are colinear

Figure 3.7

An algorithm for drawing a piece of a Bezier curve is

Given four points, (x_i, y_i), $i = 0, \ldots, 3$:

DO FOR $u = 0$ TO 1 STEP 0.01:
Compute
$x = (1-u)^3 x_0 + 3(1-u)^2 u x_1 + 3(1-u)u^2 x_2 + u^3 x_3$,
$y = (1-u)^3 y_0 + 3(1-u)^2 u y_1 + 3(1-u)u^2 y_2 + u^3 y_3$.
Plot (x, y).
ENDDO.

To continue the curve, repeat this process for the next set of four points, beginning with point p_3.

B-Spline Curves

We now discuss B-splines. These curves are like Bezier curves in that they do not ordinarily pass through the given data points. (The least-squares curves that are described in Section 3.7 are similar in this respect.) They can be of any degree, but we will concentrate on the cubic form. Cubic B-splines resemble the ordinary cubic splines of the previous section in that a separate cubic is derived for each pair of points in the set. However, the B-spline need not pass through any points of the set that are used in its definition.

We begin the description by stating the formula for a cubic B-spline in terms of parametric equations whose parameter is u.

Given the points $p_i = (x_i, y_i)$, $i = 0, 1, \ldots, n$, the cubic B-spline for the interval (p_i, p_{i+1}), $i = 1, 2, \ldots, n-1$, is

$$
\begin{aligned}
B_i(u) &= \sum_{k=-1}^{2} b_k p_{i+k}, \quad \text{where} \\
b_{-1} &= \frac{(1-u)^3}{6}, \\
b_0 &= \frac{u^3}{2} - u^2 + \frac{2}{3}, \\
b_1 &= -\frac{u^3}{2} + \frac{u^2}{2} + \frac{u}{2} + \frac{1}{6}, \\
b_2 &= \frac{u^3}{6}, \qquad 0 \le u \le 1.
\end{aligned}
\tag{3.17}
$$

As before, p_i refers to the point (x_i, y_i); it is a two-component vector. The coefficients, the b_k's, serve as a basis and do not change as we move from one set of points to the next. Observe that they can be considered weighting factors applied to the coordinates of a set of four points. The weighted sum, as u varies from 0 to 1, generates the B-spline curve.

If we write out the equations for x and y from Eq. (3.17), we get

$$x_i(u) = \frac{1}{6}(1-u)^3 x_{i-1} + \frac{1}{6}(3u^3 - 6u^2 + 4)x_i$$
$$+ \frac{1}{6}(-3u^3 + 3u^2 + 3u + 1)x_{i+1} + \frac{1}{6}u^3 x_{i+2};$$
$$y_i(u) = \frac{1}{6}(1-u)^3 y_{i-1} + \frac{1}{6}(3u^3 - 6u^2 + 4)y_i$$
$$+ \frac{1}{6}(-3u^3 + 3u^2 + 3u + 1)y_{i+1} + \frac{1}{6}u^3 y_{i+2}.$$

Note the notation here: $x_i(u)$ and $y_i(u)$ are functions (of u) and x_i, y_i are components of the point p. (The end portions are a special situation that we discuss later.)

As we have said, the u-cubics act as weighting factors on the coordinates of the four successive points to generate the curve. For example, at $u = 0$, the weights applied are 1/6, 2/3, 1/6, and 0. At $u = 1$, they are 0, 1/6, 2/3, and 1/6. These values vary throughout the interval from $u = 0$ to $u = 1$. As an exercise, you are asked to graph these factors. This will give you a visual impression of how the weights change with u.

Let us now examine two B-splines determined from a set of exactly four points. Figures 3.8(a) and 3.8(b) show the effect of varying just one of the points. As you would expect, when p_2 is moved upward and to the left, the curve tends to follow; in fact, it is pulled to the opposite side of p_1. You may be surprised to see that the curve is never very close to the two intermediate points, though it begins and ends at positions somewhat adjacent. It will be helpful to think of the curve generated from the defining equation for B_i as associated with a curve that goes from near p_1 to p_2. It is also helpful to remember that points p_0, p_1, p_2, and p_3 are used to get B_1.

Because a set of four points is required to generate only a portion of the B-spline, that associated with the two inner points, we must consider how to get the B-spline for more than four points as well as how to extend the curve into the region outside of the middle pair. We use a method analogous to the cubic splines of Section 3.5 marching along one point at a time, forming new sets of four. We abandon the first of the old set when we add the new one.

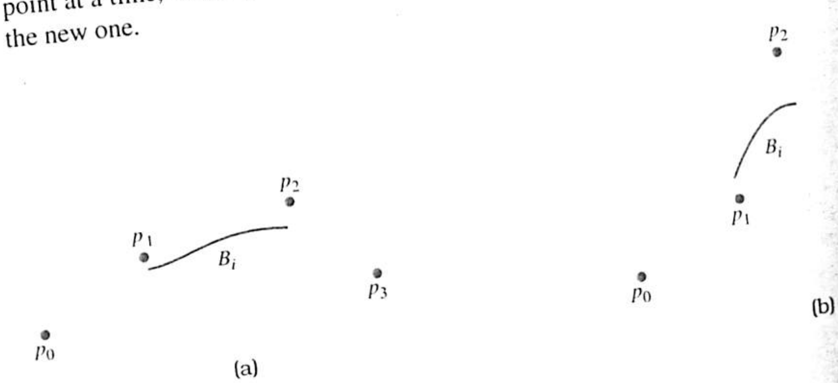


Figure 3.8

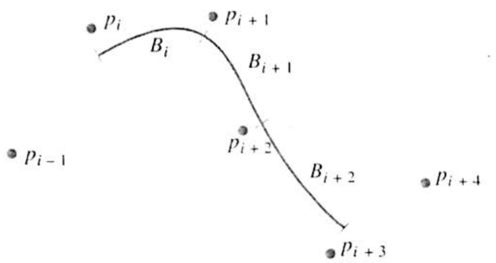


Figure 3.9 Successive B-splines joined together

The conditions that we want to impose on the B-spline are exactly the same as for ordinary splines: continuity of the curve and its first and second derivatives. It turns out that the equations for the weighting factors (the u-polynomials, the b_k) are such that these requirements are met. Figure 3.9 shows how three successive parts of a B-spline might look.

We can summarize the properties of B-splines as follows:

1. Like the cubic splines of Section 3.4, B-splines are pieced together so they agree at their joints in three ways:
   a. $B_i(1) = B_{i+1}(0) = \dfrac{p_i + 4p_{i+1} + p_{i+2}}{6}$,
   b. $B_i'(1) = B_{i+1}'(0) = \dfrac{-p_i + p_{i+2}}{2}$,
   c. $B_i''(1) = B_{i+1}''(0) = p_i - 2p_{i+1} + p_{i+2}$.

   The subscripts here refer to the portions of the curve and the points in Fig. 3.9.
2. The portion of the curve determined by each group of four points is within the convex hull of these points.

Now we consider how to generate the ends of the joined B-spline. If we have points from p_0 to p_n, we already can construct B-splines B_1 through B_{n-2}. We need B_0 and B_{n-1}. Our problem is that, using the procedure already defined, we would need additional points outside the domain of the given points. We probably also want to tie down the curve in some way—having it start and end at the extreme points of the given set seems like a good idea. How can we do this?

First, we can add more points without creating artificiality by making the added points coincide with the given extreme points. If we add not just a single fictitious point at each end of the set, but two at each end, we will find that the new curves not only join properly with the portions already made, but start and end at the extreme points as we wanted. (It looks like we have added two extra portions, but reflection shows these are degenerate, giving only a single point.)

In summary: We add fictitious points p_{-2}, p_{-1}, p_{n+1}, and p_{n+2}, with the first two identical with p_0 and the last two identical with p_n. (There are other methods to handle the starting and ending segments of B-splines that we do not cover.)

The matrix formulation for cubic B-splines is helpful. Here it is:

$$B_i(u) = \frac{1}{6}[u^3, u^2, u, 1]\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}\begin{bmatrix} p_{i-1} \\ p_i \\ p_{i+1} \\ p_{i+2} \end{bmatrix} \qquad (3.18)$$

$$= \frac{u^T M_b p}{6}.$$

This applies on the interval [0, 1] and for the points (p_i, p_{i+1}).

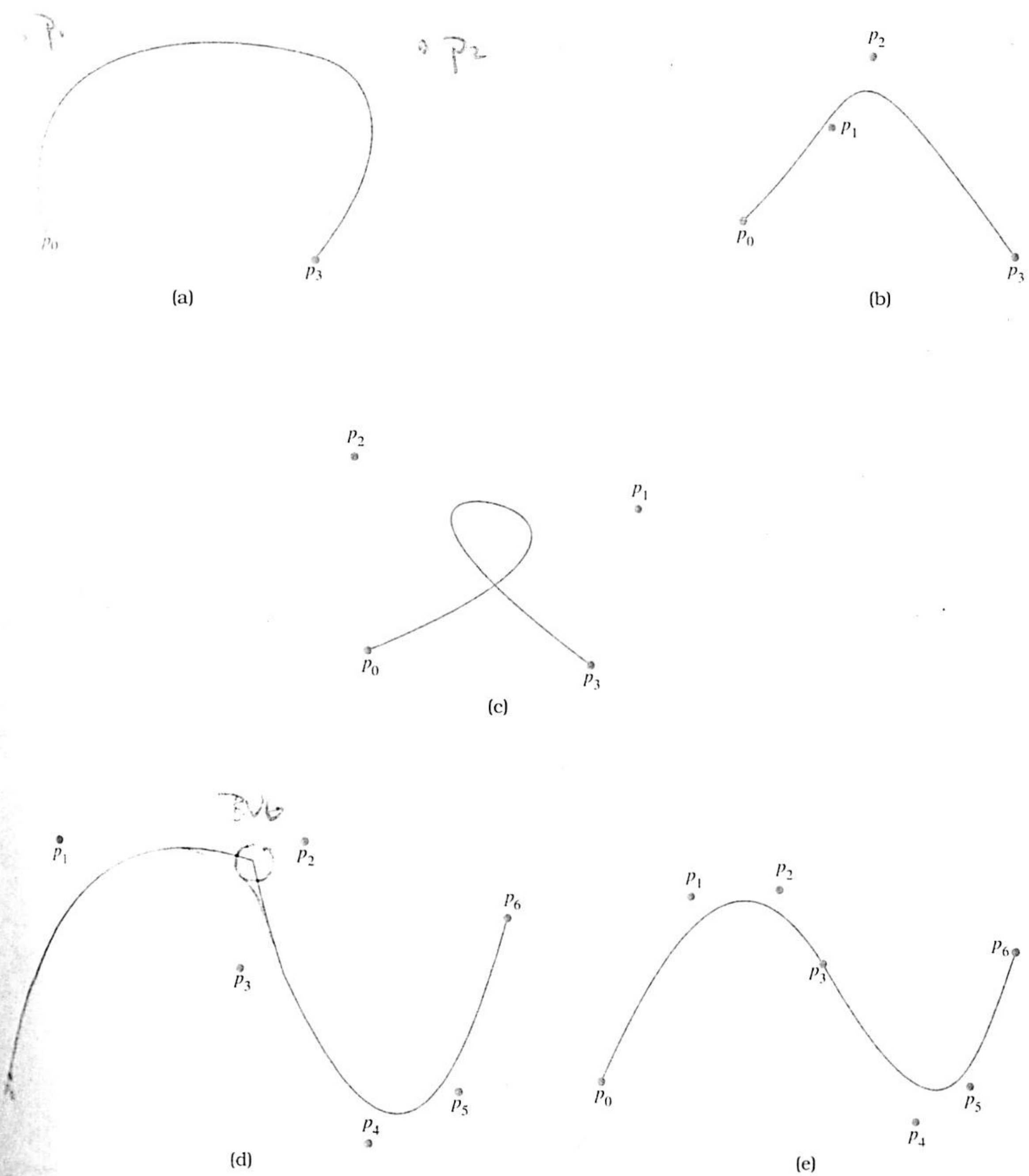
B-splines differ from Bezier curves in three ways:

1. For a B-spline, the curve does not begin and end at the extreme points.
2. The slopes of the B-splines do not have any simple relationship to lines drawn between the points.
3. The endpoints of the B-splines are in the vicinity of the two intermediate given points, but neither the x- nor the y-coordinates of these endpoints normally equal the coordinates of the intermediate points.

An algorithm for drawing a B-spline curve is as follows:

```
Given n + 1 points, p_i = (x_i, y_i), i = 0, . . . , n:
   Set p_-1 = p_-2 = p_0.
   Set p_n+1 = p_n+2 = p_n.
 DO FOR i = 0 to n − 1:
   DO FOR u = 0 TO 1 STEP 0.01:
   Compute
      x = (1 − u)^3/6x_i-1 + (3u^3 − 6u^2 + 4)/6x_i
          + (−3u^3 + 3u^2 + 3u + 1)/6x_i+1 + u^3/6x_i+2,
      y = (1 − u)^3/6y_i-1 + (3u^3 − 6u^2 + 4)/6y_i
          + (−3u^3 + 3u^2 + 3u + 1)/6y_i+1 + u^3/6y_i+2.
   Plot (x, y).
   ENDDO (u).
 ENDDO (i).
```

We conclude this section by looking at several examples of B-splines. The five Fig. 3.10 show B-splines that are defined by the same sets of points as the Bezie in Fig. 3.7. (Fictitious points have been added to complete the end portions B-splines.) There are significant differences.



P_1
P_2
P_0
P_3
(a)
P_2
P_1
P_0
P_3
(b)
P_2
P_1
P_0
P_3
(c)
BUG
P_1
P_2
P_6
P_3
P_5
P_4
(d)
P_1
P_2
P_6
P_3
P_0
P_5
P_4
(e)

Figure 3.10