5 Plane curves

5.1 Control points

We shall describe a curve with reference to a set of \(N + 1\) points, \(\{P_0, P_1, \ldots, P_N\}\), called the control points or geometric knots. Not only are the positions, i.e., the coordinates, of the control points important in the description of the curve but so too is the order in which the points are prescribed, e.g. in Fig. 5.1 the control point set \(\{P_0, P_1, P_2, P_4\}\) is different from the set \(\{Q_0, Q_1, Q_2, Q_4\}\). The smallest convex set containing the set is called the convex hull of the set.

![Diagram of control points](image)

Fig. 5.1 The set of control points \(\{P_0, P_1, P_2, P_4\}\) is different from the set \(\{Q_0, Q_1, Q_2, Q_4\}\) even though they have the same positions and the same convex hull.

There are two ways in which a set of points can be used to describe a curve. The first is called interpolation in which the curve passes through each control point. The second is called approximation in which the curve does not necessarily pass through any of the control points but, usually, passes close to all of them, see Fig. 5.2. In general, approximation is easier to compute than is interpolation.

We shall assume that the points \(P_0, P_1, \ldots, P_N\) have position vectors \(r_0, r_1, \ldots, r_N\) where \(r_i = [x_i, y_i]^T\). Then in both cases the problem is to find a function \(f\) such that the point \(P\), with position vector \(r\), where

\[
r(u) = f(u, r_0, r_1, \ldots, r_N) \quad a \leq u \leq b
\]

lies on the required curve, i.e., as the parameter \(u\) varies the locus of the point \(P\) is the required curve. This locus is often called a blend of the control
points. The general form given by equation (5.1.1) is more complicated than necessary. It is convenient to choose the function \( f \) to be linear in \( t \), so that it may be considered as a linear combination of the terms in a set of basis functions, \( \{f_0(u), f_1(u), \ldots, f_N(u)\} \), often called blending functions. Hence we write

\[
r(t) = \sum_{i=0}^{N} r_i f_i(u).
\]  

(5.1.2)

It is the choice of functions, \( f_i \), which gives rise to the different types of interpolation and approximation that we shall meet in this chapter.

For convenience, we usually choose the parameter \( u \) to lie in the interval \([0, 1]\) or in the interval \([-1, 1]\). The parametric form just described is, in general, the most convenient. However, it is possible that the curve may be generated directly in its Cartesian form

\[
y = f(x; r_0, r_1, \ldots, r_N).
\]  

(5.1.3)

### 5.2 Interpolation

#### Polynomial interpolation

\( N + 1 \) points are sufficient for a unique definition of a polynomial of degree \( N \), the point \((x, y)\) on the curve being given by the Cartesian equation

\[
y = a_0 + a_1 x + a_2 x^2 + \ldots + a_N x^N
\]  

(5.2.1)

where the coefficients \( a_0, a_1, \ldots, a_N \) are to be determined. Since we know that the curve is to pass through the points \((x_0, y_0), \ldots, (x_N, y_N)\), we can write down \( N + 1 \) equations for the \( N + 1 \) unknowns \( a_0, a_1, \ldots, a_N \). However, the solution of the equations is computationally expensive and the equations themselves are very often ill-conditioned. We can avoid these computation problems by introducing interpolation polynomials.

#### Lagrange interpolation polynomials

Lagrange interpolation polynomials have the property that, at a control point, the value of the polynomial is either zero or one. For \( N + 1 \) control points, we have \( N + 1 \) polynomials, \( L_i(x) \), each of degree \( N \), with the property

\[
L_i(x_j) = \delta_{ij} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}
\]  

The explicit form of the function \( L_i(x) \) is

\[
L_i(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_N)}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_N)}
\]  

(5.2.2)

The Cartesian equation of the curve may be written in terms of the Lagrange interpolation polynomials as \( y = L(x) \), where

\[
L(x) = \sum_{i=0}^{N} y_i L_i(x).
\]  

(5.2.3)

In general, of course, since each \( L_i(x) \) is an \( N \)th degree polynomial in \( x \), so too is \( y \). However, it is possible for some of the higher degree terms to cancel. For example, if we used three control points lying on a straight line, then the highest degree terms, i.e., the quadratic terms, would cancel leaving a linear function as we would expect. The linear and quadratic Lagrange interpolation polynomials are shown in Fig. 5.3.

#### Example 5.1

Consider the three point \((0, 0), (1, 1), \) and \((2, 2)\). The Lagrange quadratic polynomials are
Plane curves

\[ L_0(x) = \frac{(x - 1)(x - 2)}{(0 - 1)(0 - 2)} = \frac{1}{2}(x - 1)(x - 2) \]
\[ L_1(x) = \frac{(x - 0)(x - 2)}{(1 - 0)(1 - 2)} = -x(x - 2) \]
\[ L_2(x) = \frac{(x - 0)(x - 1)}{(2 - 0)(2 - 1)} = \frac{1}{2}x(x + 1). \]

Hence the curve interpolating the three points is

\[ y = L(x) \]
\[ = \sum_{i=0}^{2} y_i L_i(x) \]
\[ = 0(\frac{1}{2}(x - 1)(x - 2)) + 1[-x(x - 2)] + 2[\frac{1}{2}x(x + 1)] \]

i.e., \( y = x \), which is, of course, the straight line joining the three points.

**Hermite interpolation polynomials**

If at the control points, \( P_i \), we have not only the values of the coordinate \( y_i \) but also the values of the gradient \( y_i' = (dy/ dx)_i \), then we have \( 2(N + 1) \) known values. Consequently we can construct a polynomial of degree \( 2(N + 1) \) which passes through the control points and which takes the known values of the gradient as well as the known function values. We can do this using Hermite interpolation polynomials, \( H_i(x) \) and \( \tilde{H}_i(x) \), which have the following properties:

\[ H_i(x_i) = \delta_{ij} \]
\[ H_i(x) = 0 \]
\[ \tilde{H}_i(x_i) = 0 \]
\[ \tilde{H}_i(x) = \delta_{ij}. \]

Hermite interpolation polynomials have the following forms:

\[ H_i(x) = (1 - 2L_i(x_i)(x - x_i)[L_i(x)]^2 \]
\[ \tilde{H}_i(x) = (x - x_i)[L_i(x)]^2. \] (5.2.4)

The Cartesian equation of the curve may be written in terms of the Hermite interpolation polynomials as \( y = H(x) \), where

\[ H(x) = \sum_{i=0}^{2} y_i H_i(x) + \sum_{i=0}^{2} y_i' \tilde{H}_i(x). \] (5.2.5)

The Hermite cubic interpolation polynomials are shown in Fig. 5.4.

Using polynomial interpolation usually works in a satisfactory manner for low numbers of control points. However, it is not sensible to try to use one polynomial to represent a curve over a large range of values. As the degree of the polynomial increases, so a tendency to introduce unrepresentative oscillations occurs, as shown in the following simple example.

**Example 5.2**

Consider the curve \( y = x^{1/3} \) and the control points (0, 0), (1, 1), (8, 2), and (27, 3) which lie on the curve.

Using Lagrange cubic interpolation polynomials we obtain the interpolation function:

\[ L(x) = \frac{x(x - 5)(x - 27)}{182} - \frac{x(x - 1)(x - 27)}{532} + \frac{x(x - 1)(x - 8)}{4446}. \]

A comparison between \( y = x^{1/3} \) and \( y = L(x) \) is shown in Fig. 5.5 from which we see that the interpolation polynomial is totally unrepresentative of the original function.
Piecewise interpolation

We can overcome the problem of unwanted oscillations by producing a composite interpolation curve which is constructed using low-degree polynomials in a piecewise manner. The concept is illustrated in Fig. 5.6.

Fig. 5.6 Piecewise polynomial interpolation (a) linear; (b) a combination of quadratic, cubic, and quadratic

For piecewise interpolation it is convenient to write the Lagrange and Hermite interpolation polynomials as functions of a parameter $u$ which takes values in the range $[0, 1]$.

In this range, the interpolation polynomials are written as follows (see Exercises 5.2 and 5.3).

**Lagrange linear polynomials**

$L_0(u) = (1 - u) \quad L_1(u) = u$. (5.2.6)

**Lagrange quadratic polynomials**

$L_0(u) = 2u^2 - 3u + 1 \quad L_1(u) = -4u^2 + 4u \quad L_2(u) = 2u^2 - u$. (5.2.7)

These functions are shown in Fig. 5.7.

**Hermite cubic polynomials**

$H_0(u) = 2u^3 - 3u^2 + 1 \quad H_1(u) = -2u^2 + 3u^2$

$H_2(u) = u^3 - 2u^2 + u \quad H_3(u) = u^3 - u^2$. (5.2.8)

These functions are shown in Fig. 5.8.

**Linear interpolation between the control points $P_i$ and $P_{i+1}$, using equation (5.2.6)** is given by

$x = x_i L_0(u) + x_{i+1} L_1(u) \quad y = y_i L_0(u) + y_{i+1} L_1(u)$.

(5.2.9)

Fig. 5.7 Lagrange interpolation polynomials defined on the interval $[0, 1]$ (a) linear; (b) quadratic

Fig. 5.8 Hermite cubic interpolation polynomials defined on the interval $[0, 1]$

Now $\frac{dy}{dx} = \frac{dy/du}{dx/du}$ and $L_0'(u) = -1, \quad L_1'(u) = 1$. Hence $\frac{dy}{dx} = \frac{(y_{i+1} - y_i)/(x_{i+1} - x_i)}{x_{i+1} - x_i}$.

Similarly for interpolation between $P_{i-1}$ and $P_{i+2}$

$\frac{dy}{dx} = \frac{y_{i+2} - y_{i+1}}{x_{i+2} - x_{i+1}}$

and it follows that in general the composite curve has a discontinuity in slope at the control points.

**Quadratic interpolation between the control points $P_{i-1}, P_i$, and $P_{i+1}$, using equation (5.2.7)** is given by

$x = x_{i-1} L_0(u) + x_i L_1(u) + x_{i+1} L_2(u) \quad y = y_{i-1} L_0(u) + y_i L_1(u) + y_{i+1} L_2(u)$.

(5.2.10)
Cubic interpolation between the points \( P_i \) and \( P_{i+1} \), using equation (5.2.8) is given by

\[
x = x_i + u(x_{i+1} - x_i)
\]
\[
y = y_i + u(y_{i+1} - y_i) + \frac{u^2}{2} \frac{d^2y}{dx^2}(i)
\]
\[
y = \frac{u^3}{6} \frac{d^3y}{dx^3}(i)
\]

where \( x_i \) and \( y_i \) are the values of \( x \) and \( y \) at control point \( i \), given by \( x_i = x_{i+1} - x_i \) and \( y_i = y_{i+1} - y_i \).}

We note here that, in general, it is difficult to measure slopes accurately. In consequence, if the data set is supplied using a drawing, it is likely that the resulting curve will not be a good representation of the actual curve.

For piecewise Lagrange interpolation, the resulting curve has a discontinuity in the slope at the control points, which may be avoided using Hermite interpolation. However, it is only possible to have Hermite interpolation if the values of the derivatives are known at the control points.

In both Lagrange and Hermite interpolation the second derivative is discontinuous at the control points and it follows that the curvature is discontinuous at these points.

We now introduce a method in which we can interpolate in a piecewise manner in which first and second derivatives are continuous but for which we need very little knowledge concerning derivative values.

**Cubic splines**

Cubic splines provide one of the most commonly used interpolation techniques. The control points are joined in pairs and piecewise cubic interpolation is used between them.

The point \( P \) on the spline curve between the control points \( P_i \) and \( P_{i+1} \) has position vector given by

\[
r(u) = a_0 + a_1 u + a_2 u^2 + a_3 u^3 \quad 0 \leq u \leq 1
\]

where the vectors \( a_0, \ldots, a_3 \) are to be determined so that \( r(0) \) is the point \( P_i \) and \( r(1) \) is the point \( P_{i+1} \).

We shall consider the calculation for the \( x \)-coordinate only since the \( y \)-coordinate can be determined in a similar manner.

Suppose that

\[
x = \phi(u) = a_0 + a_1 u + a_2 u^2 + a_3 u^3
\]

and that the second derivative at \( P_i \) is given by \( M_i \). Then, since \( \phi(u) \) is cubic, it follows that \( \phi''(u) \) is linear in \( u \) and hence

\[
\phi''(u) = M_i + (M_{i+1} - M_i) u.
\]

Integrate twice to obtain

\[
\phi(u) = \frac{1}{3} M_i u^3 + \frac{1}{2} (M_{i+1} - M_i) u^2 + A + B u.
\]

We find the constants \( A \) and \( B \) using the fact that \( \phi(0) = x_i \) and \( \phi(1) = x_{i+1} \).

Hence \( A = x_i \) and \( B = (x_{i+1} - x_i) - \frac{1}{2} (M_{i+1} - M_i) \).

Consequently, the cubic expression for \( \phi(u) \) is

\[
\phi(u) = x_i + \left[ (x_{i+1} - x_i) - \frac{1}{2} (M_{i+1} - M_i) \right] u + \frac{1}{3} M_i u^3 + \frac{1}{2} (M_{i+1} - M_i) u^2.
\]

Continuity of \( \phi(u) \) at the control points \( P_i, P_{i+1}, \ldots, P_{N-1} \) determines a set of equations for the values of \( M_i \).

For the section of the spline between \( P_i \) and \( P_{i+1} \)

\[
\phi'(u) = (x_{i+1} - x_i) - \frac{1}{2} M_{i+1} - \frac{1}{2} M_i + M_i u + \frac{1}{2} (M_{i+1} - M_i) u^2.
\]

Similarly for the section of the spline between \( P_{i-1} \) and \( P_i \)

\[
\phi'(u) = (x_i - x_{i-1}) - \frac{1}{2} M_i - \frac{1}{2} M_{i-1} + M_{i-1} u + \frac{1}{2} (M_i - M_{i-1}) u^2.
\]

The value of the derivative at \( P_i \) may be obtained by setting \( u = 0 \) in equation (5.2.15) or \( u = 1 \) in equation (5.2.16). Continuity of this derivative value leads to the equation

\[
M_{i-1} + 4M_i + M_{i+1} = 6(x_{i-1} - 2x_i + x_{i+1}) \quad i = 1, 2, \ldots, N - 1.
\]

Hence we have \((N-1)\) equations for the \((N+1)\) unknowns \( M_0, M_1, \ldots, M_N \).

Consequently we require two further relations and these will depend upon the situation in any specific problem. The two relations are obtained by specifying conditions at the ends of the curve. There is a variety of possibilities but the following three cases are the most usual:

(i) **natural spline** for which the second derivative is zero at the end, i.e., \( M_0 = 0 \) or \( M_N = 0 \);

(ii) **specified first derivatives**, \( \phi'(u) = g_0 \) at \( P_0 \), or \( \phi'(u) = g_N \) at \( P_N \);

(iii) **quadratic end spans** for which the second derivative is constant throughout the span, i.e.,

\[
M_0 = M_1 \quad \text{or} \quad M_N = M_{N-1}.
\]
The overall system of equations (5.2.17) may thus be written

\[
\begin{bmatrix}
    a_{00} & a_{01} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
    1 & 4 & 1 & 0 & 0 & \cdots & & & \\
    0 & 1 & 4 & 1 & 0 & \cdots & & & \\
    0 & 0 & 1 & 4 & 1 & \cdots & & & \\
    \vdots & & & & & & & & \\
    1 & 4 & 1 & & & & & & \\
    0 & 0 & 0 & 0 & \cdots & 0 & a_{NN-1} & a_{NN} & M_N \\
\end{bmatrix}
\begin{bmatrix}
    M_0 \\
    M_1 \\
    M_2 \\
    M_3 \\
    \vdots \\
    M_{N-1} \\
    M_N \\
\end{bmatrix}
\]

where (see Exercise 5.4) for the three cases above:

(i) \( a_{00} = 1 \) and \( a_{01} = b_0 = 0 \) or \( a_{NN} = 1 \) and \( a_{NN-1} = b_N = 0 \);

(ii) \( a_{00} = 1 \) and \( a_{01} = 1 \), \( b_0 = x_1 - x_0 - a_0 \) or \( a_{NN} = 2 \) and \( a_{NN-1} = 1 \), \( b_N = y_N - x_N + x_{N-1} \);

(iii) \( a_{00} = 1 \) and \( a_{01} = -1 \), \( b_0 = 0 \) or \( a_{NN} = 1 \) and \( a_{NN-1} = -1 \), \( b_N = 0 \).

Notice that the matrix of coefficients is tridiagonal so that the solution may be calculated accurately and efficiently using standard algorithms. Once the values \( M_0, M_1, \ldots, M_N \) have been found, then the interpolation between the control points is given by equation (5.2.14). For situations in which we have a closed curve, i.e., \( P_0 = P_N \), then it is usual to invoke one of the following two conditions at this point:

(iv) **Cyclic spline** \( y_0 = y_N \) and \( M_0 = M_N \);

(v) **Anticyclic spline** \( y_0 = -y_N \) and \( M_0 = -M_N \).

Since in both these cases \( y_0, y_N, M_0, \) and \( M_N \) are unknown, we eliminate \( y_0 \) and \( y_N \) between the first and last equations. The resulting set of equations for the unknowns \( M_0, M_1, \ldots, M_N \) is now, however, no longer tridiagonal.

The cubic spline representation for \( y \) is given by an expression similar to equation (5.2.13):

\[ y = \psi(u) = b_0 + b_1 u + b_2 u^2 + b_3 u^3. \]

While it is true that the cubic spline is a definite improvement over Lagrange or Hermite interpolation it does still have two disadvantages:

(i) oscillation problems may arise at points at which the second derivative is not continuous;

(ii) local modifications require that the whole spline curve be re-computed, since every equation affects every value \( M_i \).

**B-splines**

In the previous section we used cubic interpolation over each span of the spline. In this section we shall see how to form a cubic blending function over four spans. In this case we shall consider the control points, \( P_{-2}, P_{-1}, P_0, P_{+1}, P_{+2} \), with the parameter \( u \in [-2, 2] \).

The cubic B-spline blending function is given by

\[
B(u) = \begin{cases}
(2 + u)^3/6 & -2 \leq u \leq -1 \\
(4 - 6u + 3u^2)/6 & -1 \leq u \leq 0 \\
(4 - 6u^2 + 3u^3)/6 & 0 \leq u \leq 1 \\
(2 - u)^3/6 & 1 \leq u \leq 2 \\
0 & \text{otherwise}
\end{cases}
\]

and the function \( B(u) \) is shown in Fig. 5.9.
Plane curves

The derivative is given by

\[
B'(u) = \begin{cases} 
(2 + u)^2/2 & -2 \leq u \leq -1 \\
(-4u - 3u^2)/2 & -1 \leq u \leq 0 \\
(-4u + 3u^2)/2 & 0 \leq u \leq 1 \\
-(2 - u)^2/2 & 1 \leq u \leq 2 \\
0 & \text{otherwise.}
\end{cases}
\] (5.2.21)

Sometimes the four constituent cubics which occur in the definition of \(B(u)\) in equation (5.2.20) are written individually and defined for \(v \in [0, 1]\) as follows:

\[
b_{-2}(v) = v^3/6 \\
b_{-1}(v) = (1 + 3v + 3v^2 - 3v^3)/6 \\
b_{0}(v) = (4 - 6v^2 + 3v^3)/6 \\
b_{1}(v) = (1 - 3v + 3v^2 - v^3)/6.
\]

Hence

\[
B(u) = \begin{cases} 
b_{-2}(u + 2) & -2 \leq u \leq -1 \\
b_{-1}(u + 1) & -1 \leq u \leq 0 \\
b_{0}(u) & 0 \leq u \leq 1 \\
b_{1}(u - 1) & 1 \leq u \leq 2 \\
0 & \text{otherwise.}
\end{cases}
\]

This cubic curve has the following properties:

(i) it is an even function, i.e., \(B(-u) = B(u)\);
(ii) it has continuous first and second derivatives;
(iii) the cubic B-spline has support 4, this is the number of spans over which it is non-zero;
(iv) \(b_{-2}(v) + b_{-1}(v) + b_{0}(v) + b_{1}(v) = 1\), which may be written as

\[
B(u - 2) + B(u - 1) + B(u) + B(u + 1) = 1 
\]

0 \leq u \leq 1

or

\[
B(Nu - 2) + B(Nu - 1) + B(Nu) + B(Nu + 1) = 1 
\]

0 \leq \mu \leq \frac{1}{N}.

In general, interpolate the parametric knot set but it is introduced to allow flexibility when setting the derivative values at the ends of the spline.

Suppose that the point \(A_i\) has position vector \(R_i\), and that the parameter, \(\mu\), varies from 0 at \(P_0\) to 1 at \(P_N\) and takes the value \(i/N\) at \(P_i\). Then the equation of the spline interpolating the points \(P_i\) is

\[
r(\mu) = \sum_{j=-1}^{N+1} B(N \mu - j) R_j.
\] (5.2.23)

We find the position vectors \(R_j\) by noting that the geometric knots, \(P_0, P_1, \ldots, P_N\) occur when \(\mu = i/N\) (i = 0, 1, ..., N). Hence

\[
r_j = r(j/N) = (R_{j-1} + 4R_j + R_{j+1})/6 \quad j = 0, 1, \ldots, N.
\] (5.2.24)

This is a set of \((N + 1)\) equations for the \((N + 3)\) unknowns \(R_j\). The remaining two relations may be obtained by setting the gradient at the ends to \(g_0\) and \(g_N\), i.e., \(r'(0) = g_0\) and \(r'(1) = g_N\). Now

\[
r(\mu) = B(N\mu + 1)R_{\mu-1} + B(N\mu)R_{\mu} + B(N\mu - 1)R_\mu + \ldots \\
+ B(N\mu - N - 1)R_{N+1}
\]

so that

\[
r'(\mu) = NB(N\mu + 1)R_{\mu-1} + NB(N\mu)R_{\mu} + NB(N\mu - 1)R_\mu + \ldots \\
+ NB(N\mu - N - 1)R_{N+1}
\]

Hence

\[
r'(0) = N[B(1)R_{\mu-1} + B(0)R_{\mu} + B(-1)R_\mu] \quad \text{and, using equation (5.2.21), it follows that } g_0 = (N/2)(R_1 - R_{-1}).
\]

Similarly, \(g_N = (N/2)(R_{N+1} - R_{N-1})\). Finally we obtain three sets of equations of the form

\[
M x = h,
\] (5.2.25)

where \(M\) is the matrix given by

\[
M = \begin{bmatrix}
-N & 0 & N & 0 & 0 & \ldots & 0 \\
1 & 4 & 1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 4 & 1 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 1 & 4 & 1 \\
0 & 0 & \ldots & -N & 0 & N
\end{bmatrix}
\]

where \(x\) is the vector of unknowns and \(h\) is the vector of known values.
Plane curves given by

\[
\begin{bmatrix}
R_1 \\
R_0 \\
\vdots \\
R_N \\
R_{N+1}
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
r_0 \\
r_1 \\
\vdots \\
r_N \\
r_{N+1}
\end{bmatrix}
\]

In practice this matrix system would be solved as a multiple set

\[
\begin{bmatrix}
X_1 & Y_1 & Z_1 \\
X_0 & Y_0 & Z_0 \\
\vdots & \vdots & \vdots \\
X_N & Y_N & Z_N \\
X_{N+1} & Y_{N+1} & Z_{N+1}
\end{bmatrix}
\begin{bmatrix}
\frac{3}{2}g_{x0} & \frac{3}{2}g_{y0} & \frac{3}{2}g_{z0} \\
x_0 & y_0 & z_0 \\
\vdots & \vdots & \vdots \\
x_N & y_N & z_N \\
\frac{3}{2}g_{xN} & \frac{3}{2}g_{yN} & \frac{3}{2}g_{zN}
\end{bmatrix} = 6
\]

5.3 Approximation

Cubic B-splines

The cubic B-spline introduced in the previous section may be used to approximate the set of control points \(P_0, P_1, \ldots, P_N\) with position vectors \(r_0, r_1, \ldots, r_N\) so that the approximating curve is similar to equation (5.2.23)

\[
r(\mu) = \sum_{i=0}^{N} B(N\mu - i) r_i, \quad 0 \leq \mu \leq 1.
\]

(5.3.1)

The control points at the ends of the spline correspond to \(\mu = 0\) and \(\mu = 1\). In contrast with interpolating curves, this approximating curve will not necessarily pass through the control points.

Now \(r(0) = B(0) r_0 + B(-1) r_1 = \frac{3}{2} r_0 - \frac{1}{2} r_1 \neq r_0\) and \(r(1) = B(1) r_{N-1} + B(0) r_N = \frac{3}{2} r_{N-1} + \frac{1}{2} r_N \neq r_N\). Hence, in general, the spline does not pass through these end knots. However, it can be forced to pass through them by extending the set so that we have two phantom control points, \(P_{-1}\) and \(P_{N+1}\), which are chosen so that the spline passes through \(P_0\) and \(P_N\), in this case

\[
r(\mu) = \sum_{i=-1}^{N+1} B(N\mu - i) r_i, \quad 0 \leq \mu \leq 1.
\]

(5.3.2)

For the curve to pass through \(P_0, P_N\) we require

\[
r_0 = r(0) = B(1) r_{-1} + B(0) r_0 + B(-1) r_1 = \frac{3}{2} r_{-1} + \frac{3}{2} r_0 + \frac{1}{2} r_1.
\]

Hence we choose

\[
r_{-1} = 2r_0 - r_1.
\]

(5.3.2)

Similarly, at the other end we choose \(r_{N+1} = 2r_N - r_{N-1}\).

Theorem 5.1

The slope at an end point is the same as that of the line joining the end knots.

Proof

\[
r'(\mu) = N \sum_{i=-1}^{N+1} B'(N\mu - i) r_i,
\]

so that

\[
r'(0) = N \sum_{i=-1}^{N+1} B'(-i) r_i
\]

\[
= N[B'(1) r_{-1} + B'(0) r_0 + B'(-1) r_1].
\]

Using equation (5.2.21) it follows that \(B'(-1) = \frac{1}{2}, B'(0) = 0, B'(1) = -\frac{1}{2}\). Hence \(r'(0) = -(N/2)(r_{-1} - r_1)\) and, using equation (5.3.2)

\[
r'(0) = N[r_1 - r_0]
\]

(5.3.3)

i.e., the slope at the end \(\mu = 0\) is the same as that of the line \(P_0 P_1\). A similar result holds at the end \(\mu = 1\) at which the slope is the same as that of the line \(P_{N-1} P_N\).

It is also possible to use multiple control points. The effect of a double point is to pull the curve towards that point. If the point is a triple point, then the spline will pass through that point. However, it is possible to introduce discontinuities this way. Closed curves may be generated by choosing the double point \(P_k = P_0\). The gradient will be continuous at this point provided that \(P_0, P_0, P_k\), and \(P_{k+1}\), are collinear, since the slope is the same as the line joining the end control points, see Fig. 5.10. The cubic B-spline has some interesting geometric properties which are useful for the purposes of sketching:

1. For any non-integer value of \(N\mu\), only four of the terms in equation (5.3.1) are non-zero, consequently each span of the curve is determined by at most four consecutive vertices of the defining polygon.

Consider the span defined by the points \(P_{-2}, P_{-1}, P_1, P_2\) given by

\[
r(\mu) = B(N\mu - (k - 2)) r_{k-2} + B(N\mu - (k - 1)) r_{k-1}
\]

\[
+ B(N\mu - k) r_k + B(N\mu - (k + 1)) r_{k+1},
\]

where \((k - 1)/N \leq \mu \leq k/N\). The coefficients of the vectors on the right-hand side are all positive, and, using equation (5.2.22), it has unit sum. Hence \(r(\mu)\) is a convex combination of four control points. It follows that the vector...
4. Consider the case \( N\mu = (k - 1) + \frac{1}{2} (k \neq 0 \text{ and } k \neq N) \), then
\[
r(\mu) = B(\frac{3}{2})r_{k-2} + B(\frac{1}{2})r_{k-1} + B(-\frac{1}{2})r_k + B(-\frac{3}{2})r_{k+1}.
\]
Since \( B(\frac{3}{2}) = B(-\frac{1}{2}) = 0.0208333 \) and \( B(-\frac{1}{2}) = B(\frac{1}{2}) = 0.4791667 \), it follows that \( r \approx (r_{k-1} + r_k)/2 \).

This means that the spline curve passes 'close' to the midpoint of the side \( P_{k-1}P_k \). Hence the curve passes 'close' to the midpoints of all the sides of the characteristic polygon, except for the first and the last.

5. Consider the case \( N\mu = k \) (\( k \neq 0 \) and \( k \neq N \)). In this case three coefficients only in equation (5.3.1) are non-zero and
\[
r(\mu) = B(2)r_{k-2} + B(1)r_{k-1} + B(0)r_k + B(-1)r_{k+1}
\]
\[
= (r_{k-1} + 4r_k + r_{k+1})/6
\]
\[
= 2r_k/3 + [(r_{k-1} + r_{k+1})/2]/3
\]

i.e., the spline curve passes through the point which is one-third of the way along the line joining \( P_k \) to the midpoint of the line joining \( P_{k-1} \) to \( P_{k+1} \), see Fig. 5.12. Properties 4 and 5 allow us to sketch cubic splines as shown in Fig. 5.13.

2. Since each span is defined in terms of four control points only, we see that changing, say, the point \( P_k \) affects the four spans \( k - 1, k, k + 1, \) and \( k + 2 \) only. Consequently we have the opportunity for local control of the B-spline.

3. If four consecutive control points are collinear, then the corresponding span of the B-spline is a straight line segment.
Recursive B-splines

So far we have considered third-order B-splines which have used cubic blending functions. It is of course possible to produce B-splines using other polynomial blending functions. The cubic blending function given by equation (5.2.20) is one of a family of functions which are defined recursively as follows:

$$B_j(u) = \begin{cases} 
1 & 0 \leq u \leq 1 \\
0 & \text{otherwise}
\end{cases}$$

$$B_j(u) = \frac{1}{j+1} \left[ uB_{j-1}(u) + (j+1-u)B_{j-1}(u-1) \right].$$

The range over which $B_j(u)$ is non-zero is given by $0 < u < j + 1$, $(j \neq 0)$.

Example 5.3

$$B_j(u) = \frac{[uB_0(u) + (2-u)B_0(u-1)]}{2}$$

$$= \begin{cases} 
u & 0 \leq u \leq 1 \\
\frac{1}{2} & 1 \leq u \leq 2 \\
0 & \text{otherwise}
\end{cases}$$

$$B_j(u) = \begin{cases} 
-\frac{u^2}{6} + \frac{2u}{3} - 3 & 0 \leq u \leq 1 \\
\frac{1}{6} & 1 \leq u \leq 2 \\
\frac{u^2}{6} + \frac{2u}{3} + 3 & 2 \leq u \leq 3 \\
0 & \text{otherwise}
\end{cases}$$

We have seen $B_3(u)$ in the previous section, in fact $B_3(u) = \frac{1}{4}B(u-2)$, see Exercise 5.10. We can centre the spline on the origin by considering

$$B_j\left(\tau + \frac{j+1}{2}\right), \quad -\frac{j+1}{2} \leq \tau \leq \frac{j+1}{2}$$

and we can centre the spline on the point $\mu/N$ with

$$B_j\left(\mu - i + \frac{j+1}{2}\right), \quad i - \frac{j+1}{2} \leq \mu \leq i + \frac{j+1}{2}.$$ 

The spline representation with the $j$th order blending function is

$$r(\mu) = \sum_{i=0}^{\infty} B_j(\mu - i + \frac{j+1}{2})r_i$$

and this gives a uniform B-spline because the $N + 1$ knots are uniformly spaced at $\mu = 0, 1/N, 2/N, \ldots, (N-1)/N, 1$.

Fig. 5.14 The first six recursive B-spline blending functions

The first six blending functions are shown in Fig. 5.14.

β-splines

B-splines are popular because of their reliability and the high speed with which they can be computed. Sometimes, however, they do exhibit undesirable properties such as oscillations which can arise due to the presence of extraneous points of inflection. These situations can be dealt with by using different shaped blending curves. One popular approach is to use the β-spline which has properties similar to those of the B-spline. We take $\mu = 0, 1/N, 2/N, \ldots, (N-1)/N$.

Two extra parameters are used: $\sigma$, the skew and $\tau$, the tension. When $\sigma = 1$ (no skew) and $\tau = 0$ (no tension) the β-spline reduces to the B-spline.

We shall consider the third-order β-spline

$$\beta(\mu) = \begin{cases} 
\frac{2}{\delta} & (2 + u)^3, \quad -2 \leq u \leq -1 \\
\frac{1}{\delta} & [(\tau + 4\sigma + 4\sigma^2 - 6(1 - \sigma^2)u - 3(2 + \tau + 2\sigma)u^2 - 2(1 + \tau + \sigma + 3\sigma^2)u^3] \quad -1 \leq u \leq 0 \\
\frac{1}{\delta} & [(\tau + 4\sigma + 4\sigma^2 - 6(\sigma - \sigma^3)u - 3(\tau + 2\sigma^2 + 2\sigma^3)u^2 + 2(\tau + \sigma + \sigma^2 + 3\sigma^3)u^3] \quad 0 \leq u \leq 1 \\
\frac{2}{\delta} & \sigma^2(2 - u)^3, \quad 1 \leq u \leq 2 \\
0 & \text{otherwise}
\end{cases}$$

where $\delta = \tau + 2 + 4\sigma + 4\sigma^2 + 2\sigma^3$. 
6.3 Rational parametric curves

We use homogeneous coordinates in which \( \mathbb{R}^3 \) is imbedded in \( \mathbb{R}^4 \) by \( r \rightarrow [r^1 r^2] \). The homogeneous coordinates are defined by the vector \( \mathbf{R} = \omega [r^1 r^2] \).

**Rational quadratic curves—conic sections**

We mention here the rational quadratic curve

\[
\mathbf{R}(u) = (1 - u)^2 \mathbf{R}_0 + 2u(1 - u)\mathbf{R}_1 + u^2 \mathbf{R}_2
\]

which is considered in Exercises 6.7, 6.8, 6.9, and 6.10 where we show that, by suitable choice of the weights, the rational curve reduces to:

(i) the quadratic Bézier curve \( r(u) = (1 - u)^2 r_0 + 2u(1 - u)r_1 + u^2 r_2 \);

(ii) the straight line \( r(u) = (1 - u)r_0 + ur_2 \);

(iii) the unit circle, centre the origin.

The quadratic curve is a plane curve lying in the plane defined by the points \( P_0, P_1, \) and \( P_2 \) with position vectors \( r_0, r_1, \) and \( r_2 \). It may be shown (Faux and Pratt, 1987) that the Bézier cubic curve is in fact a parabolic section.

Further it can also be shown (Rogers and Adams, 1990) that with the rational quadratic written in the form

\[
r(u) = \frac{(1 - u)^2 r_0 + 2u(1 - u)\omega_1 r_1 + u^2 r_2}{(1 - u)^2 + 2u(1 - u)\omega_1 + u^2}
\]

the curve reduces to:

(i) a straight line if \( \omega_1 = 0 \);

(ii) an elliptic curve segment if \( 0 < \omega_1 < 1 \);

(iii) a parabolic curve segment if \( \omega_1 = 1 \);

(iv) a hyperbolic curve segment if \( \omega_1 > 1 \).

**Rational cubic curves**

Consider a rational cubic curve defined by the control points \( P_0, P_1, P_2, \) and \( P_3 \) with homogeneous position vectors \( \mathbf{R}_i \). The cubic curve is given, using a direct analogy with the Bézier curve, by

\[
\mathbf{R}(u) = (1 - u)^3 \mathbf{R}_0 + 3u(1 - u)^2 \mathbf{R}_1 + 3u^2(1 - u)\mathbf{R}_2 + u^3 \mathbf{R}_3.
\]

In a similar manner to the development of equations (6.2.1) and (6.2.3) we can show (see Exercise 6.6) that

\[
\begin{align*}
\mathbf{R}(0) &= \mathbf{R}_0, \\
\mathbf{R}(1) &= \mathbf{R}_3, \\
\dot{\mathbf{R}}(0) &= 3(\mathbf{R}_1 - \mathbf{R}_0), \\
\dot{\mathbf{R}}(1) &= 3(\mathbf{R}_2 - \mathbf{R}_3),
\end{align*}
\]

and it follows that

\[
\begin{align*}
\mathbf{R}(0) &= 3\omega_1 (r_1 - r_0), \quad (6.3.1) \\
\mathbf{R}(1) &= 3\omega_2 (r_2 - r_1).
\end{align*}
\]

Similarly

\[
\begin{align*}
\dot{\mathbf{R}}(0) &= 3\omega_2 (\dot{r}_1 - \dot{r}_0), \quad (6.3.2) \\
\dot{\mathbf{R}}(1) &= 3\omega_3 (\dot{r}_2 - \dot{r}_1).
\end{align*}
\]

Here then we see that the tetrahedron defined by the vertices with position vectors \( r_0, r_1, r_2, \) and \( r_3 \) has the same significance as Bézier's characteristic tetrahedron, i.e., it has the variation diminishing property and the curve is tangential to the defining polygon at the end points. However, in this case, we also have the freedom to adjust the parameterization of the curve as well as its shape via the weights \( \omega_0, \omega_1, \omega_2, \) and \( \omega_3 \).
To obtain the curvatures we use the result of Theorem 2.3, \( \kappa(0) = |\dot{r}(0)| \times |\ddot{r}(0)| / |\dot{r}(0)|^3 \). Since \( R(u) = \begin{bmatrix} \omega(u) \dot{r}(u) \\ \omega(u) \ddot{r}(u) \end{bmatrix} \) it follows that
\[
\dot{R} = \begin{bmatrix} 0 & \omega F \\ \omega \dot{F} \end{bmatrix}, \quad \text{hence} \quad \frac{\omega F}{\omega} = \dot{R} - \frac{\dot{\omega} \dot{r} + 2 \ddot{\omega}}{\omega} \dot{r}.
\]
Now \( \dot{R} = 6(R_0 - 2R_1 + R_2) \). Hence
\[
\begin{bmatrix} \omega_0 \dot{r}(0) \\ \omega_0 \ddot{r}(0) \end{bmatrix} = 6 \begin{bmatrix} \omega_0 \dot{r}_0 - 2 \omega_1 \dot{r}_1 + \omega_2 \dot{r}_2 \\ \omega_0 - 2 \omega_1 + \omega_2 \end{bmatrix} - \frac{\dot{\omega}(0) \dot{r}_0 + 2 \ddot{\omega}(0) \ddot{r}_0}{\omega_0} \begin{bmatrix} \omega_0 \ddot{r}_0 \\ \omega_0 \ddot{r}_0 \end{bmatrix} = 6 \begin{bmatrix} \omega_0 \dot{r}_0 - 2 \omega_1 \dot{r}_1 + \omega_2 \dot{r}_2 \\ \omega_0 - 2 \omega_1 + \omega_2 \end{bmatrix} - \frac{(6 \omega_0 - 12 \omega_1 + 6 \omega_2) \dot{r}_0}{\omega_0}
\]
so that \( \dot{\omega}(0) = 6(\omega_0 - 2 \omega_1 + \omega_2) \) and \( \ddot{\omega}(0) = 3(\omega_1 - \omega_0) \). Thus
\[
\begin{align*}
\omega_0 \dot{r}(0) &= 6 \omega_0 \dot{r}_0 - 12 \omega_1 \dot{r}_1 + 6 \omega_2 \dot{r}_2 - (6 \omega_0 - 12 \omega_1 + 6 \omega_2) \dot{r}_0 \\
&= -6(\omega_1 - \omega_0) \frac{3 \omega_0}{\omega_1} \left( \dot{r}_1 - \dot{r}_0 \right)
\end{align*}
\]
\[
= \left( 12 \omega_1 - 6 \omega_2 + \frac{18 \omega_0^2}{\omega_0} - 18 \omega_0 \right) \dot{r}_0 \\
+ \left( -12 \omega_1 - \frac{18 \omega_1^2}{\omega_0} + 18 \omega_1 \right) \dot{r}_1 + 6 \omega_2 \dot{r}_2
\]
\[
= \left( -6 \omega_1 - 6 \omega_2 + \frac{18 \omega_1^2}{\omega_0} \right) \dot{r}_0 + \left( 6 \omega_1 - 18 \omega_1 \right) \dot{r}_1 + 6 \omega_2 \dot{r}_2.
\]

Hence
\[
\dot{r}(0) = \frac{1}{\omega_0} \begin{bmatrix} 6 \omega_0 - 18 \omega_0^2 \\ 6 \omega_0 \end{bmatrix} \dot{r}(0) = \frac{6 \omega_0}{\omega_0} \left( \dot{r}_1 - \dot{r}_0 \right) + \frac{6 \omega_2}{\omega_0} \left( \dot{r}_2 - \dot{r}_0 \right)
\]
from which we obtain
\[
\kappa(0) = 2 \frac{\omega_0 \omega_2}{3 \omega_0^2} \left( \dot{r}_1 - \dot{r}_0 \right) \times (\dot{r}_2 - \dot{r}_0). \tag{6.3.5}
\]
Similarly we can show that
\[
\kappa(1) = 2 \frac{\omega_0 \omega_2}{3 \omega_0^2} \left( \dot{r}_2 - \dot{r}_1 \right) \times (\dot{r}_3 - \dot{r}_2).
\]
The advantage here is that, by choice of the weights \( \omega_0, \omega_1, \omega_2, \) and \( \omega_3 \), we can control the curvatures at the end points which contrasts with equation (6.2.4) for the Bézier curve in which the end point curvatures are fixed.

**Rational B-splines**

Using equation (5.3.7) and by analogy with equation (6.3.2), we have the rational B-spline given by

\[
r(\mu) = \sum_{i=0}^{n} \omega_i B_i(\mu) \frac{N_{\mu-i+j+1}}{N_{\mu-i+j+1}^2}, \tag{6.3.6}
\]

Sometimes we write the rational B-spline in the form
\[
r(\mu) = \sum_{i=0}^{n} \omega_i B_i(\mu) \frac{N_{\mu-i+j}}{N_{\mu-i+j}^2},
\]
where
\[
R_i(\mu) = \sum_{k=0}^{n} \omega_k B_k(\mu) \frac{N_{\mu-k+j}}{N_{\mu-k+j}^2}.
\]
The rational basis functions have the properties \( R_i(\mu) \geq 0, \sum_{i=0}^{n} R_i(\mu) = 1 \) so that the curve is contained within the convex hull of the defining polygon.

Rational B-splines have similar geometric properties to those of the non-rational B-splines described in Section 5.2 but with an increases controllability by virtue of the weights \( \omega_0, \omega_1, \ldots, \omega_n \), e.g., rational B-splines have the variation diminishing property. Details can be found in Rogers and Adams (1990).

### 6.4 Non-uniform rational B-splines

The B-splines described in Section 5.2 and the rational B-splines described in Section 6.3 are defined on a uniform knot set \( [\mu = 0, 1/N, 2/N, \ldots, (N-1)/N] \). In **non-uniform** representations, the parameter interval is not necessarily uniform. In this case the blending functions are not the same for each interval—they are different for each curve segment. The recursive definition of the non-uniform B-spline is a generalization of equation (5.3.4).

Suppose that the \( N \)th degree B-spline is defined on the parametric knot set \( [u_0, u_1, u_2, \ldots, u_N] \). Then the recursive definition of the non-uniform B-spline, \( B_i(\mu) \), is given by

\[
B_i(\mu) = \begin{cases} 1 & u_i \leq \mu \leq u_{i+1} \\ 0 & \text{otherwise} \end{cases}
\]

\[
B_i(\mu) = \frac{u - u_i}{u_{i+1} - u_i} B_{i+1}(\mu) + \frac{u_{i+1} - u}{u_{i+1} - u_{i+2}} B_{i+1}(\mu). \tag{6.4.1}
\]

The only restriction on the knot set is that it is nondecreasing. Multiple knots are allowed, e.g., the knot set \( [0, 0, 0, 1, 1, 2, 3, 4, 4, 4, 5] \) has knot values 2, 3, and 5 with multiplicity 1, knot value 1 with multiplicity 2, and knot values 0 and 4 with multiplicity 3.
6.5 Composite curves in three dimensions

Consider the problem of joining the curve segment \( r^1(u) \) (\( 0 \leq u \leq 1 \)) with the curve segment \( r^2(v) \) (\( 0 \leq v \leq 1 \)). We would usually wish to ensure continuity of the curve and its slope at the join, the point which corresponds to \( u = 1 \) and \( v = 0 \). Hence we have the join conditions

\[
r^1(1) = r^2(0) \quad \text{and} \quad \dot{r}^1(1) = e_1 \dot{e}_1 \quad \dot{r}^2(0) = e_2 \dot{e}_2. \quad (6.5.1)
\]

Here \( \dot{\mathbf{e}} \) is the common unit tangent and \( e_1 \) and \( e_2 \) are the parameters which determine the fullness of each segment.

For continuity of curvature we proceed as follows:

\[
\kappa \mathbf{b} = \frac{\kappa \times \dot{\kappa}}{|\dot{\kappa}|^3} \quad (6.5.2)
\]

curvature continuity requires that

\[
\frac{\ddot{r}^2(0) \times \ddot{r}^1(0)}{|\ddot{r}^2(0)|^3} = \frac{\ddot{r}^1(1) \times \ddot{r}^1(1)}{|\ddot{r}^1(1)|^3}.
\]

Hence, using equation (6.5.1), \( \dot{\mathbf{e}} \times \ddot{r}^2(0) = \theta \dot{\mathbf{e}} \times \ddot{r}^1(1) \) where \( \theta = e_2/e_1 \). This equation is satisfied by

\[
\ddot{r}^2(0) = \theta \ddot{r}^1(1) + \lambda \ddot{r}^1(1) \quad (6.5.3)
\]

where \( \lambda \) is an arbitrary scalar which is sometimes set to zero for ease of computation but, of course, allows more flexibility if it is non-zero.

6.6 Composite Ferguson curves

Equation (6.1.2) gives the Ferguson cubic curve in the form

\[
\mathbf{r}(u) = \left(1 - 3u^2 + 2u^3\right)\mathbf{r}(0) + \left(3u^2 - 2u^3\right)\mathbf{r}(1) + (u - 2u^2 + u^3)\mathbf{r}'(0) + (-u^2 + u^3)\mathbf{r}'(1).
\]

(6.6.1)

An obvious approach to curvature continuity across segment joins is to match \( r, \dot{r}, \) and \( \ddot{r} \) across the join and set \( e_1 = e_2 = 1 \) in equation (6.5.1) to satisfy equation (6.5.2) directly.

Continuity of \( \ddot{r} \) yields, since \( \ddot{r}^1(1) = \ddot{r}^2(0) \),

\[
6\ddot{r}^1(0) - 6\ddot{r}^1(1) + 2\ddot{r}^1(0) + 4\ddot{r}^1(1) = -6\ddot{r}^2(0) + 6\ddot{r}^2(1) - 4\ddot{r}^2(0) - 2\ddot{r}^2(1). \quad (6.6.2)
\]
Note that if multiple knots are used, it is possible that the recursion in equation (6.4.1) will lead to division by zero. In this case it is conventional to replace the result by zero.

We sometimes say that \( B_{j,r}(u) \) is the \( j \)th order blending function for the control point \( P_i \).

To illustrate the effect of multiplicity on spline curves we shall discuss the non-uniform cubic B-spline in which the control point set \( \{ P_0, P_1, \ldots, P_k \} \) is approximated by cubic curve segments defined by the knot set \( \{ u_0, u_1, \ldots, u_{k+4} \} \).

(i) The spline evaluated at a single knot lies inside the convex hull and has a continuous second derivative.

(ii) The spline evaluated at a double knot, say \( u_i = u_{i+1} \), lies on the line segment \( P_{i-1}P_{i+1} \) and the second derivative is no longer continuous at the knot.

(iii) The spline evaluated at a triple knot, say \( u_i = u_{i+1} = u_{i+2} \), yields the point \( P_{i-1} \) and the first derivative is discontinuous here.

(iv) The spline evaluated at a quadruple knot, say \( u_i = u_{i+1} = u_{i+2} = u_{i+3} \), yields both \( P_{i-1} \) and \( P_i \), i.e., the curve is discontinuous.

It is property (i) that yields an increased flexibility of the non-uniform B-spline over its uniform counterpart.

The non-uniform cubic B-spline curve is given in the region defined by control points \( P_{i-3}, P_{i-2}, P_{i-1}, P_i \) and \( P_{i+1} \) by

\[
r^{(i)}(u) = B_{i-3,3}(u)r_{i-3} + B_{i-2,3}(u)r_{i-2} + B_{i-1,3}(u)r_{i-1} + B_{i,3}(u)r_i, \\
u_i \leq u \leq u_{i+1} \quad 3 \leq i \leq N. \tag{6.4.2}
\]

We have seen already that rational curves give more flexibility than their equivalent non-rational form by virtue of the choice of weights \( \{ w_i \} \). Non-uniform rational B-splines are usually referred to by the acronym NURBS. The cubic NURBS curves are given by an expression similar to equation (6.4.2)

\[
r^{(i)} = \sum_{k=i-3}^i R_{k,3}(u)r_k
\]

where

\[
R_{k,3}(u) = \frac{\alpha_k B_{k,3}(u)}{\sum_{j=k-3}^i \alpha_j B_{j,3}(u)} \quad u_i \leq u \leq u_{i+1} \quad 3 \leq i \leq N. \tag{6.4.3}
\]

An advantage of the rational splines compared with the non-rational ones is that conic sections can be defined exactly (cf. the conic sections generated by rational quadratics in Section 6.3), see Rogers and Adams (1990).

### 6.5 Composite curves in three dimensions

Consider the problem of joining the curve segment \( r^{(1)}(u) \) (0 \( \leq u \leq 1 \)) with the curve segment \( r^{(2)}(v) \) (0 \( \leq v \leq 1 \)). We would usually wish to ensure continuity of the curve and its slope at the join, the point which corresponds to \( u = 1 \) and \( v = 0 \). Hence we have the join conditions

\[
r^{(1)}(1) = r^{(2)}(0) \quad \text{and} \quad \frac{\hat{r}^{(1)}(1)}{\|\hat{r}^{(1)}(1)\|} = \frac{\hat{r}^{(2)}(0)}{\|\hat{r}^{(2)}(0)\|} \hat{r}^{(2)}(0). \tag{6.5.1}
\]

Here \( \hat{r} \) is the common unit tangent and \( \hat{e}_1 \) and \( \hat{e}_2 \) are the parameters which determine the fullness of each segment.

For continuity of curvature we proceed as follows: since

\[
\frac{\hat{r}^{(1)}(0) \times \hat{r}^{(2)}(0)}{\|\hat{r}^{(1)}(0)\|^3} = \frac{\hat{r}^{(1)}(1) \times \hat{r}^{(2)}(1)}{\|\hat{r}^{(1)}(1)\|^3}.
\]

Hence, using equation (6.5.1), \( \hat{r}^{(2)}(0) = \hat{e}_2 \hat{r}^{(2)}(0) \) where \( \hat{e}_2 = \hat{e}_1 \).

This equation is satisfied by

\[
\hat{r}^{(2)}(0) = \hat{e}_2 \hat{r}^{(1)}(1) + \alpha \hat{r}^{(1)}(1), \tag{6.5.3}
\]

where \( \alpha \) is an arbitrary scalar which is sometimes set to zero for ease of computation but, of course, allows more flexibility if it is non-zero.

### 6.6 Composite Ferguson curves

Equation (6.1.2) gives the Ferguson cubic curve in the form

\[
r(u) = (1 - 3u^2 + 2u^3)r(0) + (3u^2 - 2u^3)\hat{r}(1) + (u - 2u^2 + u^3)\hat{r}(0) + (-u^2 + u^3)\hat{r}(1). \tag{6.6.1}
\]

An obvious approach to curvature continuity across segment joins is to match \( r, \hat{r}, \) and \( \hat{r} \) across the join and set \( \hat{e}_1 = \hat{e}_2 = 1 \) in equation (6.5.1) to satisfy equation (6.5.2) directly.

Continuity of \( \hat{r} \) yields, since \( \hat{r}^{(1)}(1) = \hat{r}^{(2)}(0) \),

\[
6\hat{r}^{(1)}(0) - 6\hat{r}^{(1)}(1) + 2\hat{r}^{(1)}(0) + 4\hat{r}^{(1)}(1) = -6\hat{r}^{(2)}(0) + 6\hat{r}^{(2)}(1) - 4\hat{r}^{(2)}(0) - 2\hat{r}^{(2)}(1). \tag{6.6.2}
\]
hence $t_1 = t_2 = -\frac{4}{5}$, so that
\[ \mathbf{i}_0 = [0 \ 1 \ 0]^T, \quad \mathbf{i}_1 = \frac{1}{\sqrt{161}} [-10 \ 5 \ -6]^T, \]
\[ \mathbf{i}_2 = \frac{1}{\sqrt{161}} [10 \ 5 \ -6]^T, \quad \mathbf{i}_3 = [0 \ 1 \ 0]^T. \]

6.15 For the Ferguson cubic of Exercise 6.1
\[ \mathbf{r}(u) = \begin{bmatrix} -1 + u^2 - u^3 \\ 2 - 3u^2 + 2u^3 \\ 3 + u - 16u^2 + 11u^3 \end{bmatrix}, \quad 0 \leq u \leq 1. \]

We now split the curve at $u = \frac{1}{4}$.

(i) $r_1(\xi) = r(u)$ where $u = \frac{1}{4} + \xi$, $0 \leq u \leq \frac{1}{4}$, $0 \leq \xi \leq 1$
\[ r_1(\xi) = \begin{bmatrix} -1 + \xi^2/16 - \xi^3/64 \\ 2 - 3\xi^2/16 + \xi^3/32 \\ 3 + \xi/4 - \xi^2/16 + 11\xi^3/64 \end{bmatrix}. \]

(ii) $r_2(\eta) = r(u)$ where $u = \frac{1}{4} + \frac{1}{4} \eta$, $\frac{1}{4} \leq u \leq 1$, $0 \leq \eta \leq 1$
\[ r_2(\eta) = \begin{bmatrix} -1 + (1 + 3\eta)^2/16 - (1 + 3\eta)^3/64 \\ 2 - 3(1 + 3\eta)^2/16 + (1 + 3\eta)^3/32 \\ 3 + (1 + 3\eta)/4 - (1 + 3\eta)^2 + 11(1 + 3\eta)^3/64 \end{bmatrix}. \]

Now
\[ r_1(1) = \begin{bmatrix} 17/64 \\ -9/32 \\ -91/64 \end{bmatrix}, \quad \text{and} \quad r_2(0) = \begin{bmatrix} 51/64 \\ -27/32 \\ -273/64 \end{bmatrix}. \]
Hence $r_2(0) = 3r_1(1)$, i.e., the tangent direction is continuous but we have a discontinuity in tangent magnitude with factor 3.

### 7 Surface patches

In this chapter, we extend the ideas concerning curve segments, presented in Chapters 5 and 6, to develop similar results for surface patches. In many cases, the procedure is almost identical, the only difference being that the algebra is considerably more tedious. In such cases we shall refer to the relevant sections and then quote the results. In some cases the elements of the matrices are themselves matrices, e.g., the matrix $A$ in Section 7.1 is a $16 \times 16$ matrix written as a $4 \times 4$ block of $4 \times 4$ matrices $a_{ij}$.

We shall consider the topologically rectangular patch bounded by the four curves $u = 0$, $u = 1$, $v = 0$, and $v = 1$, as shown in Fig. 7.1. The equation of the surface patch is $r = \mathbf{a}(u, v)$, the corners have position vectors $r_{00}$, $r_{10}$, $r_{01}$, and $r_{11}$, and the boundary curves have equations $r = \mathbf{a}(0, v)$, $r = \mathbf{a}(1, v)$, $r = \mathbf{a}(u, 0)$, and $r = \mathbf{a}(u, 1)$.

![Fig. 7.1 Rectangular surface patch](image)

#### 7.1 Coons patches

Consider the first two boundary curves only. We can use linear interpolation between these curves to generate the ruled surface
\[ r_1(u, v) = (1 - u)\mathbf{a}(0, v) + u\mathbf{a}(1, v) \quad (7.1.1) \]
7.2 Ferguson bicubic patches

Consider equation (6.1.1) for the Ferguson cubic curve with \( 0 \leq u \leq 1 \). If we allow the coefficients, \( a_i \), to be functions of a second parameter, \( v \), with \( 0 \leq v \leq 1 \), then the resulting position vector can be written as

\[
\mathbf{r}(u, v) = a_0(v) + a_1(v)u + a_2(v)u^2 + a_3(v)u^3.
\]

(7.2.1)

This position vector will, as \( u \) and \( v \) vary, describe a surface patch. If we use a cubic parameterization for the coefficients \( a_i(v) \), then

\[
a_i(v) = a_{i0} + a_{i1}v + a_{i2}v^2 + a_{i3}v^3.
\]

(7.2.2)

We can now write

\[
\mathbf{r}(u, v) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij}u^jv^i
\]

i.e.,

\[
\mathbf{r}(u, v) = \mathbf{u}^T \mathbf{A} \mathbf{v}
\]

(7.2.4)

where \( \mathbf{u} = [1 \ u \ u^2 \ u^3]^T \), \( \mathbf{v} = [1 \ v \ v^2 \ v^3]^T \), and \( \mathbf{A} = [a_{ij}] \).

There are sixteen vectors, \( a_{ij} \), to be determined by specifying \( \partial \mathbf{r}/\partial u \) and \( \partial \mathbf{r}/\partial v \) at the four corners of the patch, \( r_{00}, r_{10}, r_{01}, r_{11} \). Equation (7.2.4) can be recast so that the matrix \( \mathbf{A} \) is written explicitly in terms of the known quantities at the corner of the patch. In Exercise 7.3, we show that

\[
\mathbf{r}(u, v) = \mathbf{u}^T \mathbf{C} \mathbf{T} \mathbf{v}
\]

(7.2.5)

where

\[
\mathbf{T} = \begin{bmatrix}
    r(0,0) & r(0,1) & r(0,1) & r(0,1) \\
    r(1,0) & r(1,1) & r(1,1) & r(1,1) \\
    r(0,0) & r(0,1) & r(0,1) & r(0,1) \\
    r(1,0) & r(1,1) & r(1,1) & r(1,1)
\end{bmatrix}
\]

(7.2.6)

and \( \mathbf{C} \) is given by equation (5.2.6).

7.3 Bézier patches

We can generalize the Bézier–Bernstein curve, equation (5.3.11), in an analogous manner to that which leads to equation (7.2.3).
\[ r(u, v) = \sum_{i=0}^{M} \sum_{j=0}^{N} W(i, M; v)W(j, N; u)r_{ij} \]  
(7.3.1)

with \(0 \leq u \leq 1\), \(0 \leq v \leq 1\).

For the case \(N = M = 3\), we obtain the Bézier bicubic surface patch defined by the control points \(P_{00}, P_{01}, \ldots, P_{33}\) with position vectors \(r_{00}, r_{01}, \ldots, r_{33}\), as shown in Fig. 7.3. We can follow a procedure similar to that in Section 6.2 to obtain the matrix form of the Bézier bicubic surface, see Exercise 7.5, as

\[ r = u^TMBM^Tv \]  
(7.3.2)

where \(u\) and \(v\) are the usual parameter column vectors given by equation \(7.2.4\), \(M\) is the matrix given by equation \(6.2.2\), and \(B\) is the matrix of position vectors defining the characteristic polyhedron given by \(B = [r_{ij}]\).

The derivative values \(r_u, r_v, \) and \(r_{uv}\) at the corners may then be shown to be given by, see Exercise 7.5,

\[ T = NBN^T \]  
(7.3.3)

where \(T\) is the matrix of corner derivative values given by equation \(7.2.5\) and

\[ N = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & -3 & 3 \end{bmatrix} \]  
(7.3.4)

We can thus see that the Bézier surface patch has properties that are generalizations of those for the Bézier curve. For example, see Exercise 7.5, at the corner \(P_{00}\), \(r(0, 0) = r_{00}\), \(r_u(0, 0) = 3(r_{10} - r_{00})\), \(r_v(0, 0) = 3(r_{01} - r_{00})\), i.e., the surface patch passes through the corners and the edges of the characteristic polyhedron are tangential to the patch at the corners.

Also, expanding the matrix product on the right-hand side of equation \(7.3.3\) we have

\[ T = \begin{bmatrix} r_{00} & r_{03} & 3(r_{03} - r_{00}) & 3(r_{03} - r_{02}) \\ r_{30} & r_{33} & 3(r_{33} - r_{30}) & 3(r_{33} - r_{32}) \\ 3(r_{10} - r_{00}) & 3(r_{13} - r_{03}) & 9(r_{10} - r_{13} + r_{11}) & 9(r_{10} - r_{12} - r_{03} + r_{13}) \\ 3(r_{30} - r_{20}) & 3(r_{33} - r_{23}) & 9(r_{30} - r_{33} - r_{23} + r_{31}) & 9(r_{32} - r_{33} - r_{23} + r_{31}) \end{bmatrix} \]

and we see that the four 'internal' control points, \(r_{11}, r_{12}, r_{21},\) and \(r_{22}\), affect only the twist at the patch corners and have no influence on the patch corner derivatives.

### 7.4 Rational surface patches

Consider the same characteristic polyhedron used for the Bézier bicubic surface patch in Section 7.3.

As in Section 6.3, we use homogeneous coordinates \(R = a[r \ 1]^T\) and, following the procedure described in Section 7.3, we find

\[ R(u, v) = u^TMPM^Tv \]  
(7.4.1)

where \(P = [r_{ij}].\)

Hence

\[ \begin{bmatrix} a(u, v) & a_u(u, v) \\ a_v(u, v) \end{bmatrix} = u^TM \begin{bmatrix} \omega_{ij} r_{ij} \\ \omega_{ij} \end{bmatrix} M^Tv. \]  
(7.4.2)

Thus

\[ a(u, v) = u^TM[\omega_{ij}]M^Tv \]  
(7.4.3)

and it follows that

\[ r(u, v) = u^TM[\omega_{ij}]M^Tv. \]  
(7.4.4)

In Section 6.8 we saw that four weights are required for the rational cubic curve but that it is the ratios \(a_{10}/a_{00}, a_{20}/a_{00}, a_{01}/a_{00}, \) and \(a_{21}/a_{00}\) which are important in the determination of the shape. Similarly for the bicubic patch, we have six weights but it is the ratios \(a_{10}/a_{00}, a_{11}/a_{00}, a_{20}/a_{00}, a_{21}/a_{00}, a_{01}/a_{00}, \) and \(a_{12}/a_{00}\) at \(P_{00}\), see Exercise 7.10, together with three other similar ratios at \(P_{01}, P_{02},\) and \(P_{30}, P_{23}\), which determine the shape for this patch.
7.5 Possible difficulties

Continuity of derivatives

The surface patches described so far allow continuity of position across interpatch boundaries. Continuity of derivative could be achieved by using Hermite blending functions. However, this is not just a matter of taking products of Hermite interpolation polynomials and we shall describe the necessary procedure in Section 8.1.

Degenerate patches

So far in this chapter, we have considered only surface patches which are topologically rectangular. However, it is quite common for triangular patches to be needed. Such patches can be considered to be degenerate rectangular patches by choosing two corner points, say \( P_{00} \) and \( P_{10} \), to be coincident, see Fig. 7.4. For the degenerate patch in Fig. 7.4, the boundary edge \( e = 0 \) has length zero. This patch will be well-defined if and only if there is a unique surface normal at the double corner, \( P \). Since \( P_{00} \) and \( P_{10} \) are coincident, it follows that \( \mathbf{r}(u, 0) \) is a constant vector for \( 0 \leq u \leq 1 \) and so \( \mathbf{r}(u, 0) = \mathbf{0} \). Consequently we are unable to find the direction of the normal by using the vector product \( \mathbf{r}(0, 0) \times \mathbf{r}(0, 0) \).

Instead, we use a Taylor series in \( v \) about the point \( P \), \( \mathbf{r}(u, v) = \mathbf{r}(u, 0) + \mathbf{O}(v^2) \) and

\[
\mathbf{r}(u, v) = \mathbf{r}(u, 0) + v \mathbf{r}_u(u, 0) + \mathbf{O}(v^2)
\]

and

\[
= v \mathbf{r}_u(u, 0) + \mathbf{O}(v^2).
\]

\[
\hat{\mathbf{n}}(u, v) = \begin{vmatrix} v \mathbf{r}_u(u, 0) \times \mathbf{r}_u(u, 0) + \mathbf{O}(v^2) \end{vmatrix} \]  

and the unit normal at \( P \) is given by

\[
\lim_{v \to 0} \hat{\mathbf{n}}(u, v) = \frac{\mathbf{r}_u(u, 0) \times \mathbf{r}_u(u, 0) \times \mathbf{r}_u(u, 0)}{|\mathbf{r}_u(u, 0) \times \mathbf{r}_u(u, 0)|}.
\]

If the right-hand side is independent of \( u \), then this limit is unique. Now the required unique normal is the vector \( \hat{\mathbf{n}} \) as shown in Fig. 7.5 and this is normal to the tangent plane at \( P \) which contains all the vectors \( \mathbf{r}_u(u, 0) \) with \( 0 \leq u \leq 1 \). Hence if \( \mathbf{r}_u(u, 0) \) is chosen to lie in this tangent plane for all \( u \) with \( 0 \leq u \leq 1 \), then \( \hat{\mathbf{n}} \) will be the unique normal.

7.6 Exercises

7.1 Show that the Coons bilinear surface patch, developed in Section 7.1 may be written in the form \( \mathbf{r}(u, v) = [1 - u \quad u \quad 0] \mathbf{H}[1 - v \quad v \quad 1]^T \), where \( \mathbf{H} \) is a \( 3 \times 3 \) matrix to be determined.

7.2 Consider the surface patch bounded by the four curves \( y = z = 0 \); \( y^2 + z^2 = 1 \), \( y \geq 0, z \geq 0 \); \( x = y = 0 \); \( x^2 + y^2 = 1 \), \( x \geq 0, y \geq 0 \). Obtain an expression for the Coons surface patch and find the point on the surface corresponding to the parameter values \( u = \frac{1}{2}, v = \frac{1}{2} \).
8 Composite surfaces

Most composite surfaces are constructed on a framework of two intersecting families of curves, the curves usually being of the type described in Chapter 5. Consequently, these composite surfaces will comprise topologically rectangular surface patches of the type described in Chapter 7.

8.1 Coons surfaces

The Coons patch described in Section 7.1 is easily constructed in composite form. However, it suffers from the serious drawback that only positional continuity is assured across the patch boundaries. In general, derivatives will be discontinuous across the patch boundaries.

Continuity of gradient is usually essential in most design applications and this may be obtained if we extend the ideas which lead to equation (7.14) and use the Hermite interpolation polynomials given by equation (5.2.8).

Suppose that, as in Chapter 7, the equation of the surface patch is \( r = \mathbf{a}(u,v) \) so that the boundary curves are given by \( r = \mathbf{a}(0,v) \), \( r = \mathbf{a}(1,v) \), \( r = \mathbf{a}(u,0) \), and \( r = (u,1) \).

Also, the derivatives along the boundary will be given by the functions:

\[
\begin{align*}
\mathbf{a}_u(0,v) & = \mathbf{a}_u(0,v) \\
\mathbf{a}_v(0,v) & = \mathbf{a}_v(0,v) \\
\mathbf{a}_u(1,v) & = \mathbf{a}_u(1,v) \\
\mathbf{a}_v(1,v) & = \mathbf{a}_v(1,v) \\
\mathbf{a}_u(u,0) & = \mathbf{a}_u(u,0) \\
\mathbf{a}_v(u,0) & = \mathbf{a}_v(u,0) \\
\mathbf{a}_u(u,1) & = \mathbf{a}_u(u,1) \\
\mathbf{a}_v(u,1) & = \mathbf{a}_v(u,1)
\end{align*}
\]

Hence, by analogy with equation (7.14), we can obtain the equation of the surface patch as

\[
\mathbf{r}(u,v) = H_{00}(u) H_{0v}(v) \mathbf{H}_0(v) + H_{0v}(u) H_{0v}(v) \mathbf{H}_0(v) + H_{00}(u) H_{00}(v) \mathbf{H}_0(v) + H_{0v}(u) H_{00}(v) \mathbf{H}_0(v) + H_{00}(u) H_{0v}(v) \mathbf{H}_0(v) + H_{0v}(u) H_{0v}(v) \mathbf{H}_0(v)
\]

8.2 Tensor product surfaces

Suppose that on the boundary \( u = 0 \)

\[
\mathbf{a}(0,v) = H_{00}(v) \mathbf{a}(0,0) + H_{0v}(v) \mathbf{a}(0,1) + H_{00}(v) \mathbf{a}(0,0) + H_{0v}(v) \mathbf{a}(0,1)
\]

\[
= \begin{bmatrix} \mathbf{H}_0(v) \\ H_{0v}(v) \\ H_{00}(v) \\ H_{0v}(v) \end{bmatrix} \begin{bmatrix} \mathbf{a}(0,0) \\ \mathbf{a}(0,1) \\ \mathbf{a}(0,0) \\ \mathbf{a}(0,1) \end{bmatrix}
\]

If we write \( s = [\mathbf{a}(0,0) \mathbf{a}(0,1) \mathbf{a}(0,0) \mathbf{a}(0,1)] \) then \( \mathbf{a}(0,v) = s^T \mathbf{H}(v) \), where \( \mathbf{H}(v) = [H_{00}(v) H_{0v}(v) H_{0v}(v) H_{00}(v)]^T \) with similar expressions for the boundaries \( u = 1, v = 0, \) and \( v = 1, u = 1 \). Then we see that equation (8.1.1) becomes

\[
\mathbf{r}(u,v) = \mathbf{H}^T(u,v) \mathbf{T} \mathbf{H}(v)
\]

where \( \mathbf{T} \) is the matrix of corner values and \( \mathbf{r}(u,v) \) is given by equation (7.2.4).

A patch such as this is called a tensor product surface patch and to ensure continuity of gradient across interpatch boundaries we need only ensure that the derivative values \( r_u, r_v, r_{uv} \) are matched at the corners.

We could have set up equation (8.2.1) using any suitable set of blending functions, see Exercise 8.1, but we choose to use only the Hermite cubic polynomials. Using equation (5.2.8) we see that \( \mathbf{H}(v) = C^T v \), where \( C \) is the matrix given by equation (6.1.2) and \( v = \begin{bmatrix} 1 & v & v^2 & v^3 \end{bmatrix}^T \).

Consequently we can write the boundary curve \( u = 0 \) in the form

\[
r = \mathbf{a}(0,v) = s^T \mathbf{H} = \mathbf{H}^T s = v^T C s.
\]

This is precisely the equation of the Ferguson curve, equation (6.1.3). Hence Hermite polynomials lead to a composite surface defined on a mesh consisting of Ferguson bicubic curves, the surface patch being given by

\[
\mathbf{r}(u,v) = u^T C T C^T v.
\]
The equivalence of the Hermite tensor product surface patch and the Ferguson bicubic patch in a specific case is considered in Exercise 8.2.

The major problem associated with this patch is the determination of the mixed derivative terms at the corners.

### 8.3 Spline surfaces

Following the ideas presented in Section 6.10, we can use a suitable three-dimensional set of points to generate a mesh of cubic splines. We can do this if each point in the set is considered to be a knot for a curve with parameter $u$ and a curve with parameter $v$, see Fig. 8.1. We shall generate the surface bounded by the curves $u = 0$, $u = 1$, $v = 0$, and $v = 1$. The spline surface thus generated is the same as a composite surface of Ferguson bicubic surface patches. The composite surface is then interpolated over each patch using equation (8.2.2), i.e.,

$$r(u, v) = u^T C T C^T v. \quad (8.3.1)$$

If we know the values of $r_{ij}$ at the knots, together with $r_u$, along the boundaries $u = 0$, $u = 1$, and, $r_v$, along the boundaries $v = 0$, $v = 1$, then we can construct the spline surface as follows.

Splines are generated in the $u$ direction using the $N + 1$ knots on each $u$-curve and then splines are generated in the $v$ direction using knots on each $v$-curve. However, we need on each curve, two extra items of data other than the position vectors $r_{ij}$, see Section 5.2; we use the known values of the derivatives $r_u$ and $r_v$. At this stage we have generated the mesh and we now know the values of $r_{ij}$ and $r_u$ at the knots. We now calculate the two splines which interpolate the boundary gradients $r_{ij}$ in the $u$ direction from which we obtain the twist vector $r_u$ along the boundary curves $u = 0$ and $u = 1$. To construct these two splines we need the corner values of $r_{ij}$ which are usually taken to be zero. For the final step, we calculate the $M + 1$ splines which interpolate the gradient vectors $r_u$. We now have all the values to calculate the matrix $T$ in equation (8.3.1).

Strictly speaking this surface is not a composite surface. We do not develop each patch separately and then match the degree of continuity via suitable blending functions. The surface is constructed on a mesh of splines and it is the chosen mesh that ensures the degree of continuity.

### 8.4 Cubic B-spline surfaces

We develop composite cubic B-spline surfaces as a tensor product surface in a manner very similar to that for B-spline curves in Section 5.2.

We use the same geometric knot set as in Section 8.3

$$\{ P_{ij} \} \quad i = 0, 1, \ldots, N \quad j = 0, 1, \ldots, M.$$ 

The points on the edges cause similar difficulties as for B-spline curves. To control the surface at the edges we introduce the phantom knots so that the knot set becomes

$$\{ P_{ij} \} \quad i = -1, 0, \ldots, N + 1 \quad j = -1, 0, \ldots, M + 1.$$ 

The blending function formulation takes the form

$$r(\mu, \nu) = \sum_{i=-1}^{N+1} \sum_{j=-1}^{M+1} B(N\mu - i)B(M\nu - j)r_{ij}$$

$$= b^T(\mu)Bb(\nu) \quad (8.4.1)$$

where

$$b(\nu) = [B(N\nu) + 1] \cdot [B(N\nu - 1) \cdot B(N\nu - N) \cdot B(N\nu - (N + 1))]T$$

and $B = [r_{ij}]$.

At a knot, $\mu = i/N$ and $\nu = j/M$ so that

$$r(\mu, \nu) = \begin{bmatrix} r_{i-1,j-1} & r_{i-1,j} & \cdots & r_{i-1,j+1} \\ r_{i-1,1} & r_{i-1,2} & \cdots & r_{i-1,j+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{i+1,j-1} & r_{i+1,j} & \cdots & r_{i+1,j+1} \\ \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ \end{bmatrix}.$$ 

$$r(i/N, j/M) = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ \end{bmatrix}. \quad (8.4.2)$$
Hence, to ensure that the surface passes through the corner \( r_{00} \), say, we can choose double knots so that
\[
\begin{align*}
    r_{-1} &= r_{00} & i &= 0, 1, \ldots, N \\
    r_{-1j} &= r_{00} & j &= 0, 1, \ldots, M.
\end{align*}
\]  \hspace{1cm} (8.4.3)

Then using equations (8.4.2) and (8.4.3) we find
\[
r_{00} = r(0, 0) = \left[ \begin{array}{c}
    r_{-1} \\
    r_{00} \\
    r_{10} \\
    \frac{1}{3}
\end{array} \right] = \frac{1}{3}(r_{-1} + r_{10} + 3r_{00}).
\]

Hence
\[
r_{-1} = 12r_{00} - 5r_{10} - 5r_{00} - r_{11}.
\]  \hspace{1cm} (8.4.4)

Clearly we can obtain similar expressions for \( r_{v+1,1} \), \( r_{-1M} \), and \( r_{-1M+1} \). With these conditions satisfied we have a composite B-spline surface as shown in Fig. 8.2.

---

8.5 Bicubic Bézier surfaces

Consider two neighbouring patches as shown in Fig. 8.3. Then using equation (7.3.2) we may write
\[
r^{(1)}(u, v) = u^3B_1(0)M^T v \quad \text{and} \quad r^{(2)}(u, v) = u^3B_2M^Tv.
\]  \hspace{1cm} (8.5.1)

Continuity of position will be assured across the interpatch boundary if
\[
r^{(1)}(u, 1) = r^{(2)}(u, 0), \quad \text{i.e.,} \quad u^3B_1(1)M^T[1 1 1] = u^3B_2M^T[1 0 0].
\]

Since this is true for all \( u \) it follows that, after premultiplying by \( M^{-1} \) and evaluating \( M^T[1 1 1] \) and \( M^T[1 0 0] \),
\[
B_1(0 0 0) = B_2(1 0 0).
\]

It then follows, see Exercise 8.4, that \( r^{(1)}_i = r^{(2)}_i, \ i = 0, 1, 2, 3 \), which means that the control points on the boundary edge of patch 1 are the same as those on the boundary edge of patch 2. This is of course what we would expect and has been implied in Fig. 8.3.

To ensure continuity of gradient across the interpatch boundary, we require that the tangent plane on the boundary, considered to be in patch 1, is the same as that considered to be in patch 2. In this case, both tangent and surface normal will be continuous.

The condition is that the normals have the same direction, i.e.,
\[
ax^{(1)}(u, 1) \times r^{(1)}(u, 1) = r^{(2)}(u, 0) \times r^{(2)}(u, 0),
\]  \hspace{1cm} (8.5.2)

where \( a \) is arbitrary.

Now \( r^{(1)}(u, 1) = r^{(2)}(u, 0) \), hence a possible solution is
\[
r^{(1)}(u, 1) = r^{(2)}(u, 0),
\]  \hspace{1cm} (8.5.3)

i.e.,
\[
au^3B_1M^T[0 1 2 3] = au^3B_2M^T[0 1 0].
\]

Since this holds for all \( u \) it follows that, after premultiplying by \( M^{-1} \) and evaluating \( M^T[0 1 0] \) and \( M^T[0 1 2 3] \),
\[
zB_1(0 0 -3 3) = B_2(-3 3 0 0).
\]

Hence it follows, see Exercise 8.4, that \( az_i = r^{(2)}_i - r^{(2)}_i, i = 0, 1, 2, 3 \),