5
Curves other than circles

5.1 General Implicit Quadratic Equations

The general equation of a quadratic (note the slight change from our usual notation to accommodate the number of coefficients, and also the factors of 2, which make certain calculations easier) is:

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0$$

Such general quadratics are called *conic sections* as they can represent all the shapes that it is possible to get by cutting a cone with a plane. The three shapes that can be obtained in this way are the ellipse (of which the circle is a special case), the parabola and the hyperbola. If we calculate three values from the coefficients:

$$\Delta = a(cf - e^2) + b(bf - de) + d(be - dc)$$

$$\delta = ac - b^2$$

$$S = a + c$$

(We have changed our usual notation here to comply with the usual mathematical conventions for these values.) Then, no matter how the quadratic is moved about in the plane using translation (sliding) and rotation, these three values will stay the same as long as the shape of the quadratic stays the same.

Categorising a given quadratic into one of the three possible forms it can take (ellipse, parabola, or hyperbola) can be done using the three invariants whose calculation was described above. The categorisation is sumarised as follows:

If Δ is 0 then the quadratic is degenerate and represents two straight lines (which may not always exist) otherwise:

$\delta < 0$ Quadratic is a hyperbola

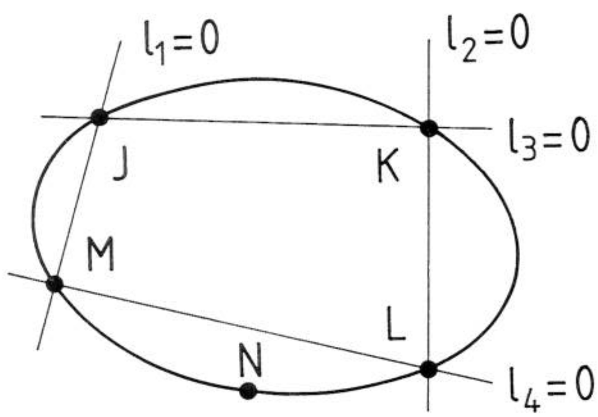
$\delta = 0$ Quadratic is a parabola

$\delta > 0$ Quadratic is an ellipse

In the latter case the ellipse only exists if ΔS is negative. Note that, when coding this, it is unlikely that a calculation of Δ will yield exactly zero for a degenerate quadratic, and similarly δ will be small, but not exactly zero, for a parabola. This imprecision is the result of rounding error, and is unavoidable on a machine that uses floating point arithmetic. General quadratics are dealt with in more detail in Bronshtein and Semendyayev.

The general quadratic can be useful for fitting smooth curves through patterns of given points and tangential to given lines. This is covered in the section on Liming multipliers in Section 5.2.

5.2 Interpolation Using General Implicit Quadratics



Suppose that we have two pairs of straight lines l_1, l_2, and l_3, l_4 where

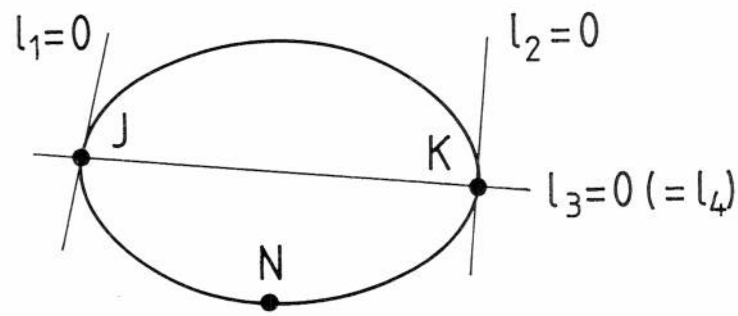
$$l_i = a_i x + b_i y + c_i , \quad i = 1, 2, 3 \text{ and } 4$$

and we multiply these line equations together with a factor λ, which is known as a *Liming multiplier*:

$$(1 - \lambda)l_1 l_2 + \lambda l_3 l_4 = 0$$

We will have generated a family of implicit quadratic equations, each different value that we choose for λ giving a different quadratic. All the quadratics will have the property that they will pass through the intersection points of the pairs of lines, J,K,L, and M. To specify one quadratic we need a value for λ. This can be found by specifying a fifth point, N, through which the quadratic must pass and then substituting the value of (x_N, y_N) into the quadratic to find λ. (To find l_1 - l_4 given J,K,L and M use the method given in Section 1.6.)

If we reduce the number of lines to three by making two of them equal ($l_3 = l_4$, say) then we get quadratics tangential to l_1 and l_2 at the points where l_3 cuts those two lines.



and the quadratics become:

$$(1 - \lambda)l_1l_2 + \lambda l_3^2 = 0$$

Again we can tie down the value of λ by specifying another point, N, through which the quadratic is to pass and substituting its x and y coordinates back into the quadratic.

This technique is very useful for constructing piecewise quadratics that join smoothly (in other words which have common tangents where they join) at given points or with given straight lines. The technique is described more fully in Chapter 1 of Faux and Pratt.

5.3 Parametric Polynomials

Implicit equations which are higher order than quadratic (ie have terms in x^3, x^2y etc.) are not generally useful because of the problems encountered in solving them to obtain a value of y for a given x coordinate, or vice versa. Extending the parametric line to higher orders by adding terms in t^2, t^3 and so on does not give rise to this problem as values of x and y are easily calculated from a value of t. These *parametric polynomials* may therefore be used to generate more flexible curves than is possible with the implicit quadratic.

The simplest non-linear parametric curve is the quadratic:

$$x = a_1 + b_1 t + c_1 t^2$$

$$y = a_2 + b_2 t + c_2 t^2$$

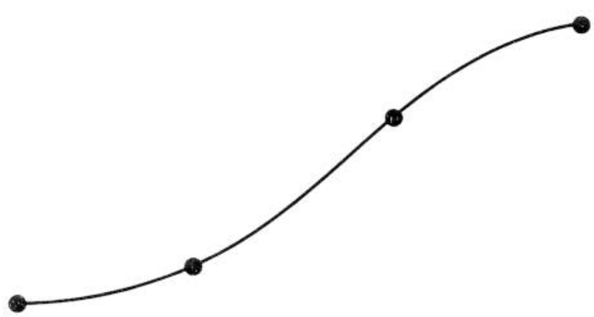
(Note the change in notation from that used in previous parametric equations to accommodate the additional terms.) The next form is the parametric cubic

$$x = a_1 + b_1 t + c_1 t^2 + d_1 t^3$$

$$y = a_2 + b_2 t + c_2 t^2 + d_2 t^3$$

and so on, with higher order equations being formed by adding more terms. Three-dimensional curving lines in space may be formed by adding a third equation in z.

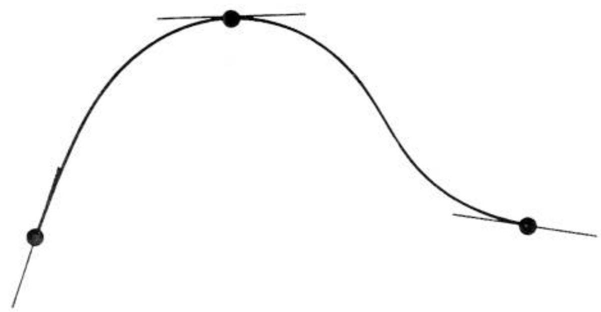
5.4 Interpolation Using Parametric Polynomials



Parametric polynomials are often used to interpolate a curve through a set of data points. To do this it is first necessary to choose the value of t which will correspond to each given point, thus determining the order in which the curve passes through the points. The chosen values of t and the corresponding x and y values for the points are substituted into the parametric equation at each point. This gives two sets of linear simultaneous equations in the coefficients of the parametric polynomials. If the order of the curve (the highest power of t used) is one less than the number of points (3 points for a quadratic, 4 for a cubic etc.), then the simultaneous equations can be solved. The curve is thus defined, and it may then be drawn or used in other calculations.

Programs for the solution of many simultaneous equations are beyond the scope of this book, but the reader employing a large mainframe computer may well find that he has a library package (such as the NAG library) available for the purpose. Users of small computers may not be in such a fortunate position, though a subset of the NAG library is available that runs under the CP/M microcomputer operating system. Alternatively, readers should consult Wilkinson and Reinsch.

Interpolation through points is often called *Lagrangian Interpolation*. *Hermite Interpolation*, on the other hand, is concerned with fulfilling slope constraints as well.



To achieve this, the equations are differentiated. For example a cubic

$$x = a_1 + b_1 t + c_1 t^2 + d_1 t^3$$

$$y = a_2 + b_2 t + c_2 t^2 + d_2 t^3$$

has differentials:

$$\frac{dx}{dt} = b_1 + 2c_1 t + 3d_1 t^2$$

$$\frac{dy}{dt} = b_2 + 2c_2 t + 3d_2 t^2$$

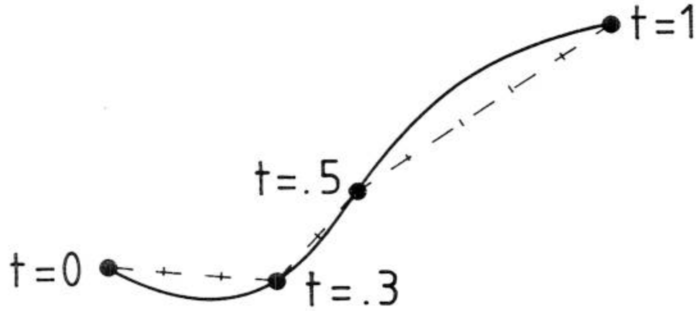
The values of dx/dt and dy/dt are generated from the specified slopes at each given point for the values of t that it has been decided to use at that point. These values are substituted into the differential equations to give yet more simultaneous linear equations in the polynomial coefficients, and the entire set, from original and differentiated polynomials, is solved together to give values for the coefficients.

It is not, of course, necessary to have slope constraints at every point; position and slope constraints can be mixed as required.

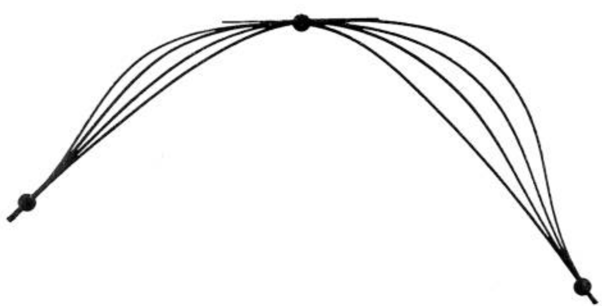
Interpolation does not always yield the curve that the reader might intuitively expect. Here are three points to watch:

1) If the initial points are approximately evenly spread along the course of the desired line,

then the points can be parameterised at even intervals of t. If the points are uneven, however, this should be reflected in the parameterisation. Making the intervals in the parameter proportional to the distance between points is a solution commonly adopted:

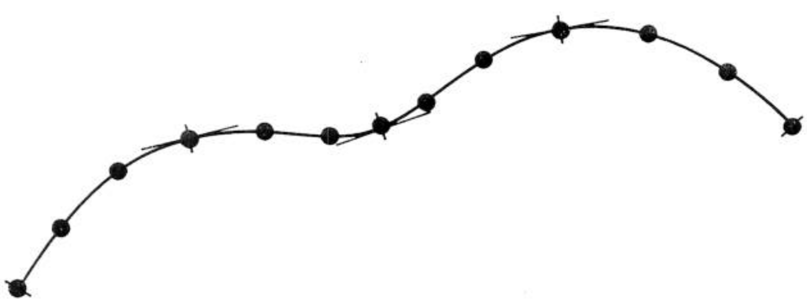


2) In Hermite interpolation, there are no unique values of dx/dt and dy/dt for a required dy/dx (slope), only the ratio of (dy/dt)/(dx/dt) must correspond. Increasing the actual values of dx/dt and dy/dt specified will lead to a flatter curve at the point being considered, but may perhaps produce unwanted effects elsewhere.



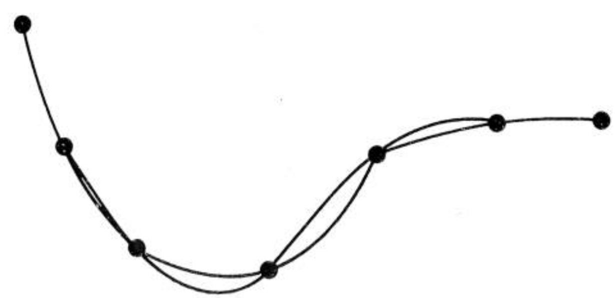
3) As the order of curves becomes higher, undesired oscillations - waviness - will tend to occur. Fifth and sixth order curves may be regarded as a conservative limit.

5.5 Parametric Spline Curves



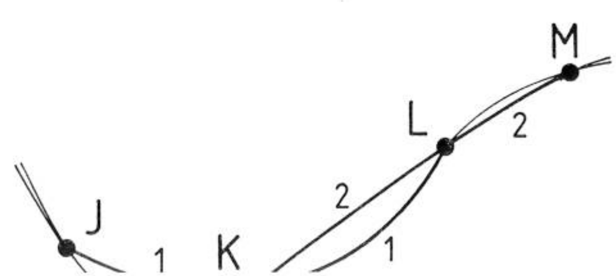
When too many points or slope constraints must be met for a single polynomial to be used several polynomials may be joined end to end. This can be done by dividing the data points into groups or *spans* and interpolating different polynomials over each span with the constraint that the slopes at the joins should match. The joins are called *knots*. All the simultaneous equations needed to find all the coefficients of all the parametric polynomials are consequently inter-dependent, and the whole linear system needs to be solved in one, rather complicated, operation. The completed structure is called a *spline*. As each span will generally contain few points high order polynomials are not often used in this technique. Despite this, waviness may still occur, and there have been a number of sophisticated spline formulations that attempt to overcome such difficulties.

A much simpler technique, which has a number of advantages, is the parametric *Overhauser* curve. This is generated by dividing the set of points through which the curve is to pass (which we will assume to be fairly regularly spaced) into overlapping groups of three. A parametric quadratic is fitted to each set as outlined in Section 5.4.



A linear blending function is then used to combine the sets of curves into a single curve through all the data points. This blending function has the value one at the middle of each quadratic and zero at the ends. The resulting curve is smooth, and will not exhibit waviness, however many points are to be interpolated.

Looking in more detail at part of the diagram above, suppose that two sets of three points, J K L and K L M



are parameterised uniformly as follows:

Span 1, J K L:

At J $s = 0$

At K $s = 0.5$

At L $s = 1$

Span 2, K L M:

At K $t = 0$

At L $t = 0.5$

At M $t = 1$

Suppose also that solving for the constant terms in the parametric quadratics gives the equations:

For J K L

$$x = a_{11} + b_{11}s + c_{11}s^2$$

$$y = a_{12} + b_{12}s + c_{12}s^2$$

and for K L M

$$x = a_{21} + b_{21}t + c_{21}t^2$$

$$y = a_{22} + b_{22}t + c_{22}t^2$$

Then any point on the curve *between K and L* has coordinates

$$x = f(a_{11} + b_{11}s + c_{11}s^2) + g(a_{21} + b_{21}t + c_{21}t^2)$$

$$y = f(a_{12} + b_{12}s + c_{12}s^2) + g(a_{22} + b_{22}t + c_{22}t^2)$$

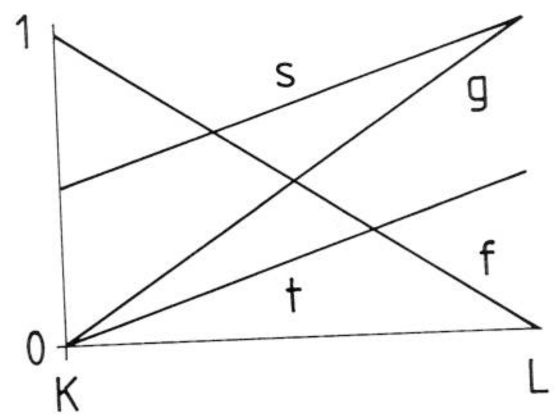
where the parameters are linked

$$s = t + 0.5$$

and f and g are the linear blending multipliers:

$$f = 2 - 2s = 1 - 2t$$

$g = 2s - 1 = 2t$



The curve between the first and second points, and between the last and last but one points is not, of course, blended in the parametric Overhauser scheme, but is the unaltered quadratic curve in each case.

5.6 Radius of Curvature

The radius of curvature, R, of an arbitrary curve, $f(x,y) = 0$, at some point on the curve (x,y) is the radius of a circle that would have the same curvature as that of the curve at that point. As this radius goes to infinity at points of inflection it is not as useful as curvature, κ, its inverse:

$$\kappa = \frac{1}{R}$$

κ is given by

$$\kappa = \frac{f_{xx} f_y - 2f_{xy} f_x f_y + f_{yy} f_x}{[f_x^2 + f_y^2]^{3/2}}$$

where f_{xx} is the second partial derivative of the function with respect to x and so on. For parametric curves, $x = f(t)$, $y = g(t)$, κ is given by:

$$\kappa = \frac{f_t g_{tt} - g_t f_{tt}}{[f_t^2 + g_t^2]^{3/2}}$$