Mechanics Corner

A Journal of Applied Mechanics & Mathematics by DrD, #33

Advanced Polynomial Curve Fitting

The use of polynomials to fit engineering data is a common engineering practice. In school, we learn that "A data set consisting of n data points $((x_i, y_i), i = 1, 2, 3, \ldots n)$ can be exactly fitted with a polynomial of degree $n - 1$. Thus three data points can be fitted exactly with a quadratic expression, four data points can be fitted exactly with a cubic expression, and so on. If this approach is pursued much further, something ugly appears: while a polynomial of degree $n-1$ will pass exactly through n data points, for large values of n, it will oscillate wildly in between the data points. Since one of the most common reason for using a polynomial fit in the first place is for interpolation – to be able to estimate a function value at locations between the known data points – this wild oscillation is devastating. It is at this point that least squares fitting is usually introduced to give an approximate fit using a much lower order polynomial. A different approach is employed here.

In a recent engineering problem, the need arose to develop an interpolation for locations between three specific data points. That would seem to indicate that a quadratic polynomial $F_2(x) = a + bx + cx^2$ would be the only choice for an exact fit. Then the fitting equations become

$$\left\{ \begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} \right\} = \left[\begin{array}{ccc} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{array} \right] \left\{ \begin{array}{c} a \\ b \\ c \end{array} \right\}$$

to be solved for the coefficients a, b, and c. But there was more to the problem.

The middle data point was supposed to be the maximum value on the curve, and a quadratic, $F_2(x) = a + bx + cx^2$ that passed exactly through the three data points had a maximum elsewhere, higher and in the wrong place! What to do?

The obvious answer is to add another "degree of freedom" to the polynomial (another adjustable coefficient) and impose the condition that the maximum occur at the required location, thus $F_3(x) = a + bx + cx^2 + dx^3$ for which the coefficients are determined from

$$\left\{ \begin{array}{c} y_1 \\ y_2 \\ 0 \\ y_3 \end{array} \right\} = \left[\begin{array}{cccc} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 0 & 1 & 2x_2 & 3x_2^2 \\ 1 & x_3 & x_3^2 & x_3^3 \end{array} \right] \left\{ \begin{array}{c} a \\ b \\ c \\ d \end{array} \right\}$$

Notice that the third line of the matrix relation is $0 = b + 2cx_2 + 3dx_2^2$ which is the same thing as $F_3'(x = x_2) = 0$, the requirement for an extremum at $x = x_2$. This is a whole lot better, but there was actually more to the problem beyond this!

The interpolating curve was supposed to flow smoothing into another curve at $x = x_3$. Since $F_2(x)$ and $F_3(x)$ each pass through the point (x_4, y_4) continuity is assured with either curve, but what about "smoothly"? To flow smoothly into the second curve extending from x_3 to the right, the slopes should match at that point as well. What to do? Add another degree of freedom to the interpolating polynomial, thus $F_4(x) = a + bx + cx^2 + dx^3 + ex^4$ for which $F_4'(x) = b + 2cx + 3dx^2 + 4ex^3$. It is necessary also to evaluate the slope curve to the right of x_3 so that y_3' becomes a known value. Then the coefficients are determined by

$$\left\{\begin{array}{c} y_1 \\ y_2 \\ 0 \\ y_3 \\ y_3' \end{array}\right\} = \left[\begin{array}{ccccc} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ 0 & 1 & 2x_2 & 3x_2^2 & 4x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 \\ 0 & 1 & 2x_3 & 3x_3^2 & 4x_3^3 \end{array}\right] \left\{\begin{array}{c} a \\ b \\ c \\ d \\ e \end{array}\right\}$$

When the coefficients a, b, c, d, and e are determined and the fitted polynomial is plotted at this level, it will found to be a very close approximation to what intuition expects the actual function to be. There is, however, one more place where the fit can be improved to the eye.

Recall the requirement for a smooth transition to the right at $x = x_3$. The practiced eye can detect changes in curvature at a point, so the fit can be improved by assuring that the curvature is continuous at $x = x_3$. Curvature is controlled by both the first and second derivatives of the function, and the requirement for a first derivative match has already been imposed. All that remains to require continuity in the second derivative at this point. For this, it is assumed that the second derivative of the function to the right of x_3 is known, y_3''.

Adding one more degree of freedom to the polynomial gives $F_5(x) = a + bx + cx^2 + dx^3 + ex^4 + fx^5$. The determination of the coefficients then requires the solution of

$$\left\{\begin{array}{c} y_1 \\ y_2 \\ 0 \\ y_3 \\ y_3' \\ y_3'' \end{array}\right\} = \left[\begin{array}{cccccc} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ 0 & 1 & 2x_2 & 3x_2^2 & 4x_2^3 & 5x_2^4 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 & x_3^5 \\ 0 & 1 & 2x_3 & 3x_3^2 & 4x_3^3 & 5x_3^4 \\ 0 & 0 & 2 & 6x_3 & 12x_3^2 & 20x_3^3 \end{array}\right] \left\{\begin{array}{c} a \\ b \\ c \\ d \\ e \\ f \end{array}\right\}$$

With this last adjustment, the curve transition at x_3 should appear very smooth, so it would seem that all should be wonderful at this point. Is it?

Well, no, not quite. Recall the problem about introducing oscillations as the degree of the polynomial is increased? This is real, and efforts for a smooth transition at x_3 may quite likely have undone the requirement that x_2 be a maximum. To be sure, the derivative will still remain zero at x_2 (that was built into the coefficient determination), but that may no longer represent the true maximum point on the curve. It may be a minimum, a local maximum with a greater maximum elsewhere, or a point of inflection on the fitted polynomial. Thus this process of progressively adding degrees of freedom to the fitted polynomial in order to apply more and more constraint equations must be applied very judiciously. How well it works in any particular case depend on the actual numbers involved and on the exact nature of the constraints to be imposed. For example, is the requirement that x_2 be the maximum point strictly true, or is it only that the curve be fairly flat around x_2 with small local oscillations acceptable? A lot of computer generated plots are called for in the application of this technique. Despite its limitations, this is still a powerful tool in the engineer's mathematical toolbox, and should be ready for use when needed.

DrD is a retired Professor of Mechanical Engineering in the USA. He can be reached for comments, questions, and requests through the message system at ME Forum. Be sure to check back soon at http://mechanical-engg.com/forum/blogs/blog/206-mechanics-corner/ for more articles.