INTERPOLATING POLYNOMIALS

We will construct a polynomial to fit, *exactly*, a discrete set of data, such as

x:	0	1	3	4	7
y:	2	–3	0	1	–2

There are two standard techniques for achieving this (they both result in the *same* polynomial):

Newton's Interpolation (divided differences)

To run a polynomial through all points of the above table, it needs to have 5 coefficients (i.e. degree 4), such that

$$c_0 + c_1 x_i + c_2 x_i^2 + c_3 x_i^3 + c_4 x_i^4 = y_i$$

where $i = 0,\ 1, ...\ 4$. The resulting 5 equations for 5 unknowns are $linear$, having a unique solution (unless there are two or more identical values in the x row). Later on we learn how to solve such systems of linear equations, but here, we can proceed more directly:

- We fit the first point only by a zero-degree polynomial (a constant),
- we extend this to a linear fit through the first two points,
- quadratic fit through the first three points,
- ..., until we reach the end of the data.

Each step of the procedure can be arranged to involve only one unknown parameter, which we can easily solve for.

Expl: To fit the data of the original table, we start with the constant 2 (the first of the y values). Then, we add to it $c_1(x-0)$, where c_1 is a constant to yield, for the complete expression, -3 (the second y value), i.e. $2+c_1 = -3 \Rightarrow c_1 = -5$. Our solution so far is thus $2-5x$. Now, we add to it $c_2(x-0)(x-1)$ so that the total expression has, at $x = 3$, the value of 0, i.e. $2-5\times3+c\times3\times2 = 0 \Rightarrow c_2 = \frac{13}{6}$. This results in $2-5x+\frac{13}{6}x(x-1)$. Add $c_3(x-0)(x-1)(x-3)$ and make the result equal to 1 at $x = 4 \Rightarrow c_3 = -\frac{7}{12}$. Finally, add $c_4(x-0)(x-1)(x-3)(x-4)$ and make the total equal to -2 at $x = 7 \Rightarrow c_4 = \frac{19}{252}$. Thus, the final answer is

$$2 - 5x + \frac{13}{6}x(x-1) - \frac{7}{12}x(x-1)(x-3) + \frac{19}{252}x(x-1)(x-3)(x-4)$$

Expanding (quite time consuming if done 'by hand') simplifies this to

$$2 - \frac{275}{28}x + \frac{1495}{252}x^2 - \frac{299}{252}x^3 + \frac{19}{252}x^4$$

One can easily verify that this polynomial passes, exactly, through all five points.

Computing the c values is made easier by utilizing the following scheme:

$$
\begin{array}{cccccc}
0 & \boxed{2} & & & & \\
 & & \frac{-3-2}{1-0} = \boxed{-5} & & & \\
1 & -3 & & \frac{\frac{3}{3}-(-5)}{3-0} = \boxed{\frac{13}{6}} & & \\
 & & \frac{0-(-3)}{3-1} = \frac{3}{2} & & \frac{-\frac{1}{6}-\frac{13}{6}}{4-0} = \boxed{-\frac{7}{12}} & \\
3 & 0 & & \frac{1-\frac{3}{2}}{4-1} = -\frac{1}{6} & & \frac{-\frac{1}{18}-(-\frac{7}{12})}{7-0} = \boxed{\frac{19}{252}} \\
 & & \frac{1-0}{4-3} = 1 & & \frac{-\frac{1}{2}-(-\frac{1}{6})}{7-1} = -\frac{1}{18} & \\
4 & 1 & & \frac{-1-1}{7-3} = -\frac{1}{2} & & \\
 & & \frac{-2-1}{7-4} = -1 & & & \\
7 & -2 & & & &
\end{array}
$$

Proof: c_0 must clearly equal y_0. Solving $y_0 + c_1(x_1 - x_0) = y_1$ yields $c_1 = \frac{y_1 - y_0}{x_1 - x_0}$. Solving $y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x_2 - x_0) + c_2(x_2 - x_1)(x_2 - x_0) = y_2$ leads to

$$
\begin{aligned}
c_2(x_2 - x_1) &= \frac{y_2 - y_0}{x_2 - x_0} - \frac{y_1 - y_0}{x_1 - x_0} \\
&= \frac{y_2(x_1 - x_0) - y_1(x_2 - x_1 + x_1 - x_0) + y_0(x_2 - x_1)}{(x_2 - x_0)(x_1 - x_0)} \\
&= \frac{(y_2 - y_1)(x_1 - x_0) - (y_1 - y_0)(x_2 - x_1)}{(x_2 - x_0)(x_1 - x_0)}
\end{aligned}
$$

implying

$$
c_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_1}
$$

etc.

Lagrange interpolation

This is a somehow different approach to the same problem (yielding the same solution). The main idea is this: Having $n + 1$ pairs of x-y values, it is relatively easy to construct an n degree polynomial whose value is 1 at $x = x_0$ and 0 at $x = x_1$, $x = x_2$, ... $x = x_n$, as follows

$$
P_0(x) = \frac{(x - x_1)(x - x_2)....(x - x_n)}{(x_0 - x_1)(x_0 - x_2)....(x_0 - x_n)}
$$

where the denominator is simply the value of the numerator at $x = x_0$. Similarly, one can construct

$$P_1(x) = \frac{(x - x_0)(x - x_2)....(x - x_n)}{(x_1 - x_0)(x_1 - x_2)....(x_1 - x_n)}$$

whose values are 0, 1, 0, ... 0 at $x = x_0, x_1, x_2, ... x_n$ respectively, etc.

Having these, we can then combine them into a single polynomial (still of degree n) by

$$y_0 P_0(x) + y_1 P_1(x) + ... + y_n P_n(x)$$

which, at $x = x_0, x_1, ...$ clearly has the value of $y_0, y_1, ...$ and thus passes through all points of our data set.

Expl: Using the same data as before, we can now write the answer almost immediately

$$2\frac{(x-1)(x-3)(x-4)(x-7)}{(-1)(-3)(-4)(-7)} - 3\frac{x(x-3)(x-4)(x-7)}{1(-2)(-3)(-6)} +$$
$$\frac{x(x-1)(x-3)(x-7)}{4\times 3\times 1\times(-3)} - 2\frac{x(x-1)(x-3)(x-4)}{7\times 6\times 4\times 3}$$

which expands to the same polynomial.

Based on this fit, we can now interpolate (i.e. evaluate the polynomial) using any value of x within the bounds of the tabulated x's (taking an x outside the table is called extrapolating). For example, at $x = 6$ the polynomial yields $y = \frac{1}{63} = 0.015873$. Since interpolation was the original reason for constructing these polynomials, they are called interpolating polynomials. We will need them mainly for developing formulas for numerical differentiation and integration.

To conclude the section, we present another **example**, in which the y values are computed based on the $\sin x$ function, using $x = 60$, 70, 80 and 90 (in degrees), i.e.

x	$\sin x^{\mathrm{o}}$
60	0.8660
70	0.9397
80	0.9848
90	1.0000

The corresponding Lagrange interpolating polynomial is

$$\begin{aligned}
&0.866\frac{(x-70)(x-80)(x-90)}{-6000} \\
&+0.9397\frac{(x-60)(x-80)(x-90)}{2000} \\
&+0.9848\frac{(x-60)(x-70)(x-90)}{-2000} \\
&+\frac{(x-60)(x-70)(x-80)}{6000} \\
&= -0.10400 + 2.2797\times 10^{-2}x - 9.7500\times 10^{-5}x^2 \\
&-2.1667\times 10^{-7}x^3
\end{aligned}$$

Plotting the difference between this polynomial and the $\sin x^{\rm o}$ function reveals that the largest error of such and approximation throughout the 60 to 90 degree range is be about 0.00005. This would be quite sufficient when only a four digit accuracy is desired. Note that trying to extrapolate (go outside the 60-90 range) would yield increasingly inaccurate results.