The Imperfect Best
by Gerald Fitton

Ring the bells that still can ring
Forget your perfect offering.
There is a crack in everything,
That's how the light gets in.

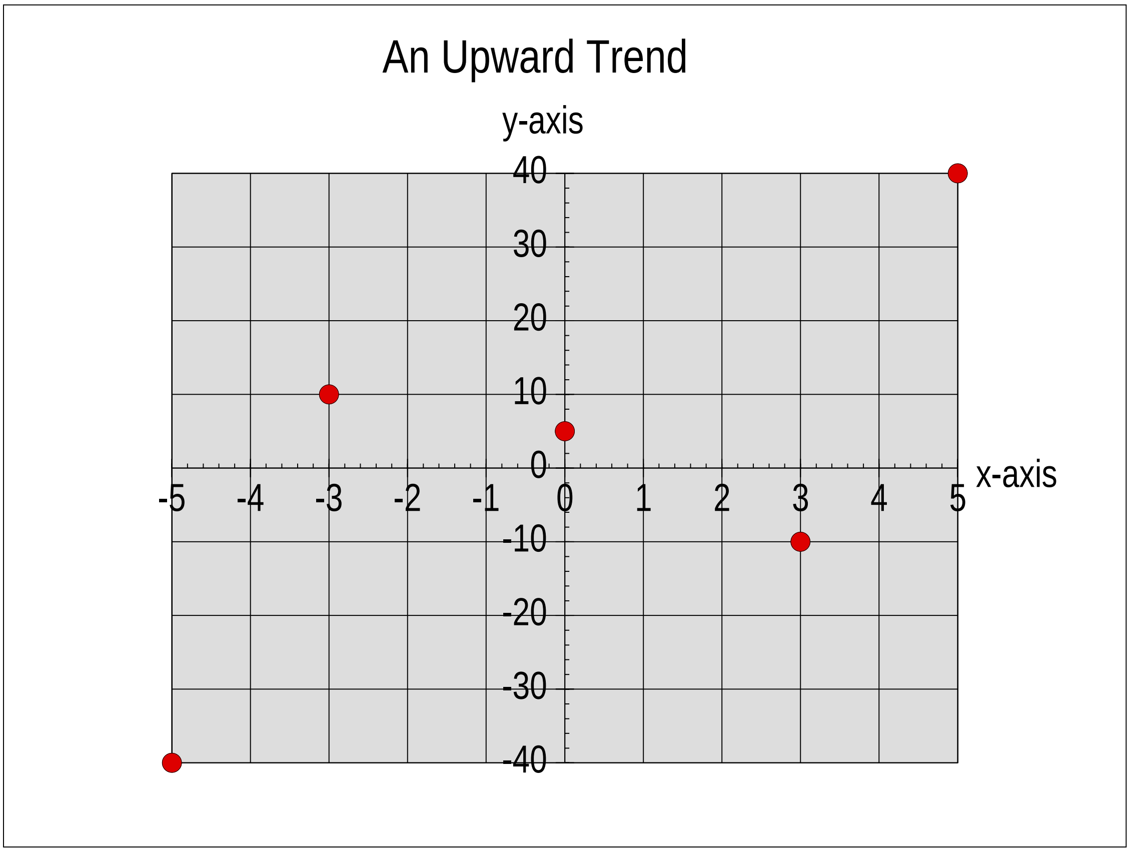
"Anthem" by Leonard Cohen (1934 -)
Released 17th November 1992

The Canadian, Leonard Cohen, published his first book of poetry whilst he was still an undergraduate at McGill University. He wrote poetry and fiction through the 1960s before becoming a folk singer and singer-songwriter. His song, "Anthem" with the oft quoted line, "Ring the bells that can still ring", featured in the film, "Natural Born Killers".

If we pursue 'Perfection' with relentless enthusiasm then we can become so obsessed with detail that we lose track of our destination. We need to accept the "cracks" if we are to make progress along that road which leads to our distant goals.

Too many points

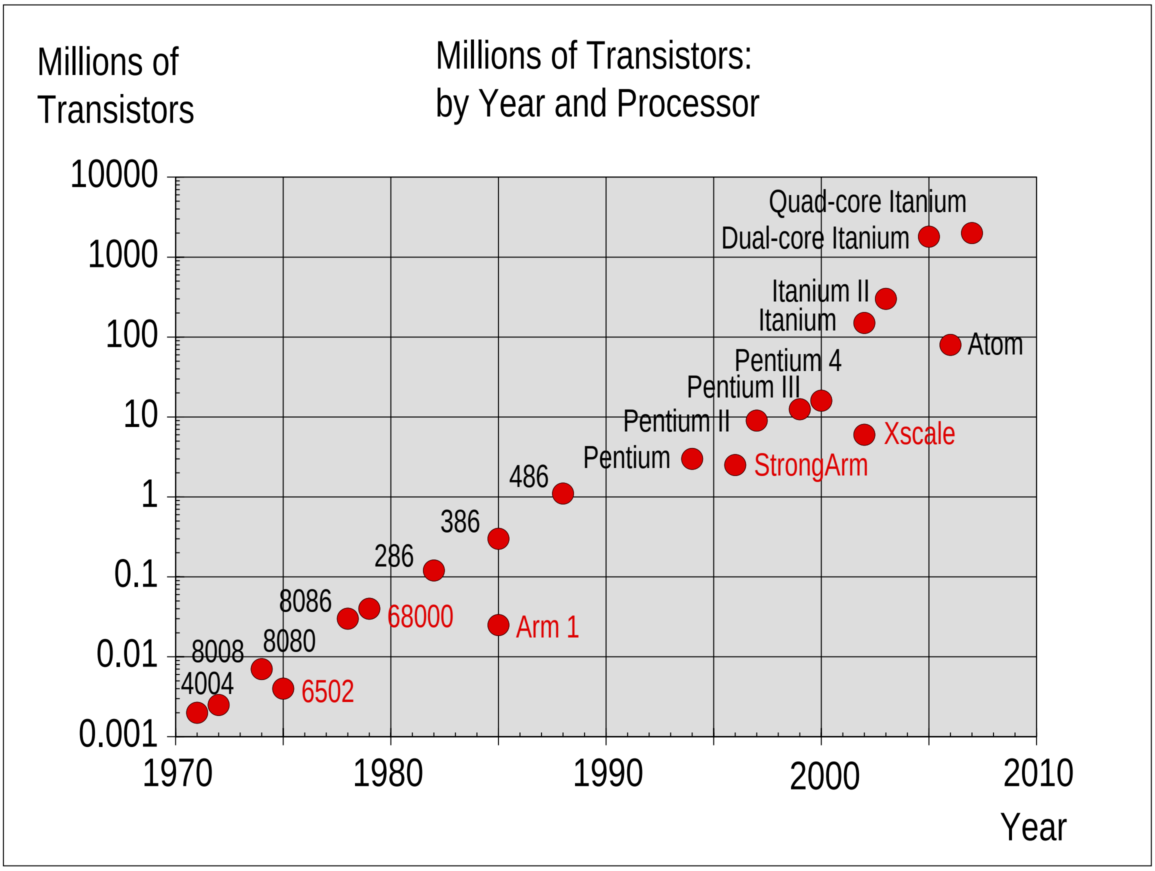
In the last couple of articles I described how, using matrix algebra, we can find the formulae for a straight line through 2 points, a parabola through 3 points and a cubic through 4 points. In the graph below you will see 5 points. “What order formula is required in order to pass, ‘perfectly’ through these 5 points?” Of course you, my most intelligent reader, will know that what is required is a quartic, a formula with an x^4 in it.



If you have followed the matrix algebra of my previous articles then I am sure you will have no trouble constructing a spreadsheet containing the 5x5 matrix which, using Bob Ardler’s custom function, you can invert and thus find the parameters, a, b, c, d and e of the unique quartic, the one and only quartic, which passes ‘perfectly’ through all 5 points.

However, a ‘perfect fit’ is, perhaps, not what is wanted. When we have, say, 21 points scattered in such a way that the general trend is upwards then do we really want to find the 20th order function which passes ‘perfectly’ through all 21 points? Or should we be seeking the formula for a less than ‘perfect’ line, one which predicts the general trend?

The diagram below illustrates Moore’s Law; it contains 21 points. Should we find the 20th power formula for the wiggly line that passes ‘perfectly’ through every one of these 21 points? Of course not. What we want is a smoother, ‘imperfect’ line showing the upward trend. The 21 points will be scattered around this line with few, if any, actually on the line.



The Imperfect Best Straight Line

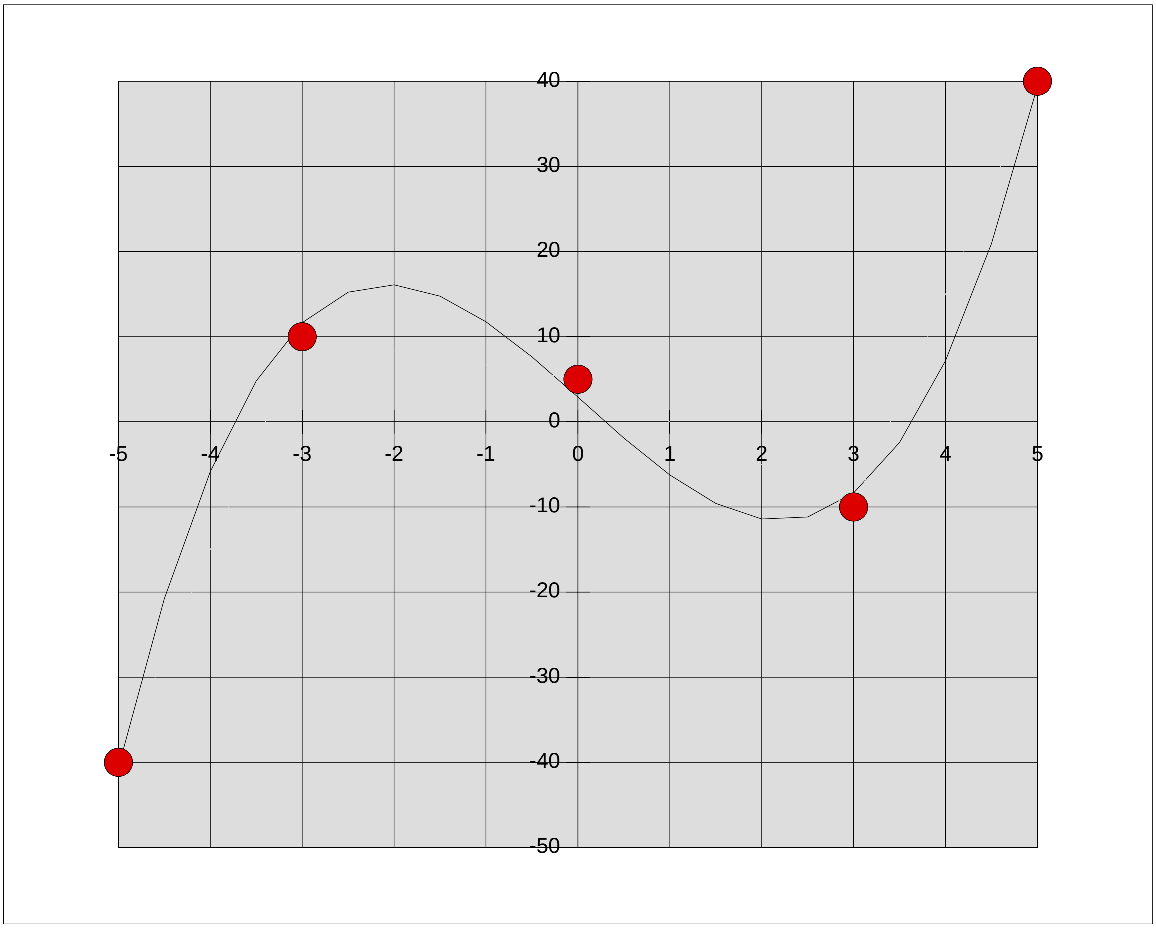
I am sure that (nearly?) all of you who have read this far will have been shown at school, university or, perhaps, much later how to find the formula of the 'best' straight line which doesn't go through every point. I have no doubt that every one of you will have a book somewhere with the 2 magical formulae for the parameters a and b of the 'best' straight line, y = ax + b. Most spreadsheets have a 'line of best fit' facility built into them.

One of the points on this best straight line is the point at the 'centre' of this mass of points. The co-ordinates of this 'centre' point is at the average of the x values and the average of the y values. The slope of the line is such that the sum of the squares of the offsets (of the actual points from the line) is a minimum. As the slope is varied from near horizontal to near vertical, this sum (of the squares of the offsets) passes through a minimum value. That line is the 'best' straight line 'through' (albeit not 'perfectly through'!) these points.

What I shall not do in this article is to limit myself to such a 'best' straight line. Instead I shall describe to you a matrix technique which will find the parameters of the 'best' parabola, cubic or quartic, etc. This line is the 'Imperfect Best Fit' for a number of points, which is far too many for any us to consider the 'Perfect Fit' to be the 'Best Fit'.

The Best Cubic

For my worked example I shall find the parameters of the 'Best Cubic' for 5 points.



It may look as though this cubic passes through the points at x = -5 and x = +5. You can see from the table below that it doesn't. The point at x = 0, y = 5 is the one furthest from this 'Imperfect' but 'Best' cubic; it is (5.00 - 2.90) = 2.10 away from the line. The sum of the squared error (sometimes called the 'squared deviations') is 10.51 (to 2 places).

x	y	y cubic	err^2
- 5.00	- 40.00	- 40.59	0.35
- 3.00	10.00	11.64	2.70
0.00	5.00	2.90	4.42
3.00	- 10.00	- 8.36	2.70
5.00	40.00	39.41	0.35
		sum(err^2)=	10.51

The formula for this cubic (the coefficients shown here are to 2 decimal places) is:

$y = 0.71x^3 - 0.14x^2 - 9.71x + 2.90$

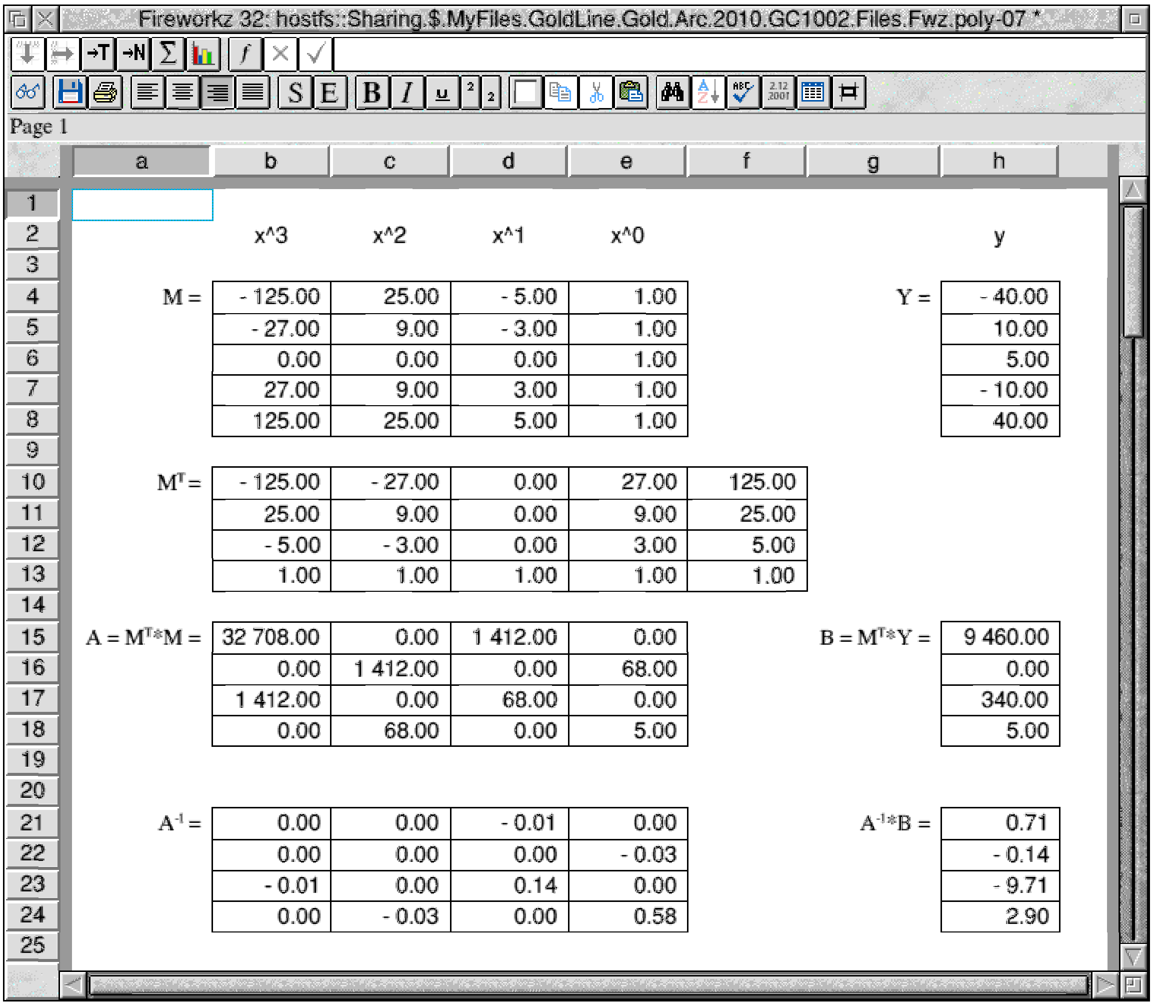
If you can find a formula for a cubic which misses these 5 points in such a way that the squared error for your cubic is less than 10.51 (to 2 places) then let me know! If the matrix method which I shall describe in the next few paragraphs does not return parameters (the coefficients of the third order formula) giving the 'best' cubic then let me know!

The Spreadsheet

The screenshot below shows the working part of the Fireworkz spreadsheet which I have used to find this ‘Best Cubic’. The x values of the 5 points are entered into the block d4d8 and the corresponding y values into the block h4h8. There are no other manual entries required; the spreadsheet does all the necessary calculations. It returns the parameters a = 0.71, b = 0.14, c = -9.71 and d = 2.90 in the block h21h24.

These values of the parameters can be entered into the general formula:

$y = ax^3 + bx^2 + cx = d$ to give the ‘best’ (least squares) cubic for these 5 points.



	a	b	c	d	e	f	g	h
1								
2		x^3	x^2	x^1	x^0			y
3								
4	M =	- 125.00	25.00	- 5.00	1.00		Y =	- 40.00
5		- 27.00	9.00	- 3.00	1.00			10.00
6		0.00	0.00	0.00	1.00			5.00
7		27.00	9.00	3.00	1.00			- 10.00
8		125.00	25.00	5.00	1.00			40.00
9								
10	M^T =	- 125.00	- 27.00	0.00	27.00	125.00		
11		25.00	9.00	0.00	9.00	25.00		
12		- 5.00	- 3.00	0.00	3.00	5.00		
13		1.00	1.00	1.00	1.00	1.00		
14								
15	$A = M^T*M$ =	32 708.00	0.00	1 412.00	0.00		$B = M^T*Y$ =	9 460.00
16		0.00	1 412.00	0.00	68.00			0.00
17		1 412.00	0.00	68.00	0.00			340.00
18		0.00	68.00	0.00	5.00			5.00
19								
20								
21	A^{-1} =	0.00	0.00	- 0.01	0.00		$A^{-1}*B$ =	0.71
22		0.00	0.00	0.00	- 0.03			- 0.14
23		- 0.01	0.00	0.14	0.00			- 9.71
24		0.00	- 0.03	0.00	0.58			2.90
25								

There is a part of my spreadsheet (rows 26 to 55) which is not visible in this screenshot. This hidden part is used to draw the graph of the cubic using these calculated parameters.

The spreadsheet and graph are ‘live’ so that if you change any or all of the five points then this Fireworkz spreadsheet will recalculate the parameters a, b, c and d, and use the new values to redraw the best cubic, the cubic having the least squares error for the new points.

You will find this Fireworkz spreadsheet and other relevant files on the website.

Higher Order Lines

Although this spreadsheet is suitable for finding the ‘Best Cubic’ the same principles can be applied to the creation of a spreadsheet which will find the parameters of higher order ‘Best Lines’ such as the ‘Best Quartic’ or ‘Best Quintic’ when there are far too many points to make an exact fit (the ‘perfect fit’) the most appropriate line for those points.

In the spreadsheet shown in my example (and available on the website) you will see that column b contains x^3; to construct a spreadsheet to find the parameters of the ‘Best Quartic’ an extra column is required for the x^4 term and matrix M will be a column wider.

The Transpose of a Matrix

The matrix M has 5 rows and 4 columns. It is said to have dimension 5 by 4; this is often abbreviated to 5x4. Now have a look at the matrix filling the block b10f13. I have called this matrix M^T. The ‘T’ stands for ‘Transpose’.

In cell b10 you will find the formula set_value (b10f13, transpose(b4e8)). The function set_value(destination,source) expands the ‘source’, the transpose of b4e8, into the ‘destination’ block, b10f13.

If you look at M^T and compare it with M then you’ll see that what has happened is that rows have become columns and columns have become rows.

An important thing to note is that the dimension of M^T is 4x5.

Matrix Multiplication

I have explained the detail (and the mechanism) of matrix multiplication in an earlier issue of Archive. Look at page 25 of Volume 22 Number 3. At the bottom of the first column of that page you’ll find:

“Matrix multiplication requires that the first matrix has the same number of columns as the second matrix has rows.”

You will appreciate that by transposing the original matrix, M, to create M^T, we ensure that M^T has the same number of columns as M has rows. Consequently we can use matrix multiplication to find a result for M^T*M. I have called this result matrix A; you will find it in the block b15e18.

Similarly we can find B = M^T*Y because M^T has the same number of columns as Y has rows. You will find B (a matrix of dimension 5x1) in the block h15h18.

The Inverse of A

The matrix A^{-1} is the Inverse of A. I have expanded it into the block b21e24.

Note that A has dimension 4x4; it is a 'square' matrix. An Inverse matrix exists only for matrices having the same number of rows as it has columns. By using the operation M^T*M to generate A we have ensured that A is square and therefore has an Inverse.

The Matrix Algebra

So that's nearly all the definitions sorted out. There is one more definition and that is the matrix of the coefficients of the cubic equation $y = ax^3 + bx^2 + cx = d$ which I shall write as C (for coefficients). The matrix C is a column matrix having dimension 4x1.

Although what I shall write here is not strictly logical I do find it a good way of arriving at the final result which is:

$C = (M^T*M)^{-1}*M^T*Y$

Our starting point for finding C is a matrix equation which is not quite true; it is not quite true because of the offsets from the cubic curve:

$M*C = Y$

We pre-multiply both side by M^T to get:

$M^T*M*C = M^T*Y$

This can be rewritten as:

$A*C = B*Y$

Now we pre-multiply both sides of this matrix equation by A^{-1} to get:

$A^{-1}*A*C = A^{-1}*B*Y$

The reason why we pre-multiply A by the Inverse of A is because:

$A^{-1}*A = I$ where I is the identity matrix

Substituting the Identity Matrix, I for $A^{-1}*A$, we get our answer:

$I*C = C = A^{-1}*B*Y$

Or, in extended form:

$C = (M^T*M)^{-1}*M^T*Y$

Compare this with the matrix formulation of the solution when we have exactly the right number of points (eg 4 points for a cubic); this is the problem I dealt with in the last two issues of Archive.

$C = M^{-1}*Y$

The only complication caused by the excess of points is the introduction of the Transpose of the matrix M. It is this introduction (twice) of the Transpose which allows us to do the necessary matrix multiplications because it ensures that the dimensions of all the matrices which are multiplied together are such that the number of columns in the pre-multiplying matrix is always equal to the number of rows in the post-multiplied matrix. Also, by using $M^T*M = A$ we have found an A which is ‘square’ and consequently has an Inverse.

Have we found the right answer?

Although I have shown you some ‘clever’ matrix operations, has they done the trick?

What I mean by this is that, although the introduction of the Transpose Matrix has allowed us to do some fancy multiplications, and create an Inverse matrix, does the value of C (the column matrix of the coefficients of the cubic) which results from these operations, really generate a cubic which has the property that the square of the y offsets is a minimum?

You can breath a huge sigh of relief now because (a) I assure you that it does and (b) I am not going to prove it to you; I’m sure that you will believe that I could - once upon a time!

The proof involves finding general formulae for the ‘Residual’ (the sum of the squares of the offsets) and then differentiating (using partial differentiation) this sum with respect to the various coefficients. For those of you interested enough to pursue this proof then I suggest that you look at the properties and uses of what is called a Vandermonde Matrix. This type of matrix is such that the values across each row form a geometrical progression; it is named after Alexandre-Théophile Vandermonde (1735 - 1796). I am sure that somewhere on the internet (by entering the right key words in Google) you’ll find that one of the uses of this type of matrix, the Vandermonde Matrix, is to do exactly what I’ve done here, namely to find the ‘Best Polynomial’ which doesn’t pass ‘perfectly’ through an over abundance of points.

Computing ‘Best Polynomials’

Matrix Multiplication generally requires much less computing power than is required to Invert a matrix. Earlier I showed you a chart, the Moore’s Law chart, containing 21 points. Suppose we decide to fit part of a cubic to these 21 points.

Although we will have to pre-multiply a 21x4 matrix by its 4x21 transpose, when it comes to finding the inverse of M^T*M we are inverting a 4x4 just as we did in this 5 point cubic problem. Consequently the computing power required to solve the 21 point problem (finding the coefficients of the ‘Best Cubic’) is only marginally more than that used with the 5 point problem I have used for my example.

Do you need any help?

If you wish to draw the ‘Best Cubic’ (or quartic or quintic) through a set of points and find yourself in trouble then please get in touch with me and I’ll see if I can help you. If you do need my help then please send me a zipped up copy of your not quite working spreadsheet because that way I shall be able to discover exactly what is going wrong for you.

Box-Out   The Rejection of Perfection

Like so many great poets, novelists and folk singers, Leonard Cohen was able to accept his limitations. He knew that he was flawed, imperfect but, as he says in the lines that I've quoted, "There is a crack in everything, That's how the light gets in".

We, each of us should try to know and then accept our own imperfections. We need to, "Ring the bells that can still ring" rather than waste all too much time and effort seeking unachievable 'Perfection'. Once we are able to acknowledge, accept and build into our lives those imperfections then we can use those working 'Bells' rather than wait in vain for that " ... perfect offering" which, like the end of a rainbow, is always just out of reach.

Of course I'm not suggesting that seeking perfection is a bad thing; what I'm suggesting is that we should not use the search for perfection to put off using those, "Bells that can still ring" just because there are other 'bells' which we acknowledge are not working perfectly (and, perhaps, never will).

We have to do what we can with whatever imperfect facilities we have got. There are times when we need to accept that rejecting 'perfection' in favour of the achievable is the best that we can do.

Bio-Bit

There was a time some 10 or 20 years ago that Gerald, looking way back to the days of his youth, wondered if he rang his bells aright. Now he finds that his enjoyment of those distant memories is heightened by the knowledge that, although many of those youthful bells can no longer ring for him, he knows they rang loudly and with great enthusiasm.