Using Weighted Power and Exponential Curve Fitting

Palmer Hanson

Introduction

The June 1980 PPC Calculator Journal (V7N5P9-11) presented an HP-41 program by Ron Knapp which would calculate 1,000 digits in 11.5 hours. That result was the basis for a challenge in the so-called "friendly competition" between users of HP and TI machines. Ron provided execution times for a total of five numbers of digits: 30 digits in 2 minutes, 90 digits is 9 minutes, 200 digits in 34 minutes, 1,000 digits in 11.5 hours and 1,160 digits in 15.25 hours. In this article I will use Ron's data in an analysis of curve fitting techniques which will show that the use of weighted data can yield a significant improvement in the quality of a power function fit.

When I took a curve fitting course at the University of Minnesota back in 1949 Professor Eggers opened the course with a discussion of how to select models to test. Professor Eggers said we should always consider the physics of the problem. In a similar vein page 7 of William Kolb's book on curve fitting with programmable calculators [Reference 1] states "We should always select a model for our data on the basis of either theoretical or empirical knowledge." When I considered the nature of pi-finding programs I concluded that the time of execution versus the number of digits calculated should be square function. My reasoning was that the number of iterations and the execution time per iteration would both be proportional to the number of digits. In the olden days I would have plotted the Knapp data on log-log paper, and when the plot approximated a straight line I would have tried a power function fit as my first choice. That was before there were computer or calculator solutions such as that in the HP-48S which will solve using several different functions and select the best function based on some criteria, typically the correlation coefficient.

A Power Function Fit to the Knapp Data

If I use any of the typical multiple curve fitting methodologies and with the elapsed time in minutes l find that the "best fit" is of the form y = Ax^B where with the HP-48S:

A = 4.87796502271E-3

B = 1.70772919979

r = 0.997876797805

The results seem to be roughly consistent with my expectations. But when I use the SCATR and FCN functions of my HP-48S to superimpose a plot of the function on a scatter diagram of the input data I get a plot which shows the fourth and fifth data points well above the plot of the function

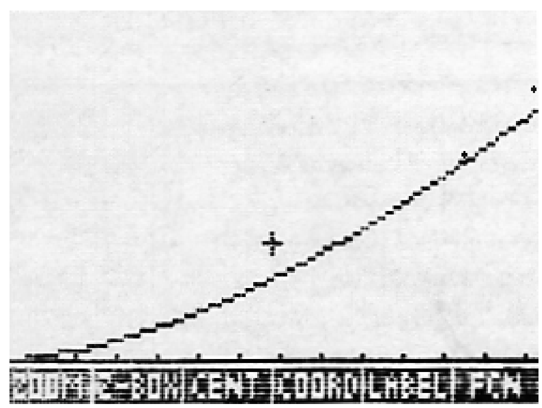

Fig. 1 - HP-48S plot on a scatter diagram.

Calculating the residuals using the PREDY function of my

Table 1 - Knapp's Pi Run Time Residuals

Digits	Time	Residuals
30	2 minutes	0.375342...
90	9 minutes	-1.606099...
200	34 minutes	-7.473892...
1000	690 minutes	42.221870...
1160	915 minutes	80.352548...

HP-48S, where residual = observed value – calculated value, yields the following results which confirm the observations from the plot. The sum of the residuals is 113.8697702 and the sum of the squares of the residuals is 8297.797888 . Those residuals suggest that there might be another power function which would yield smaller residuals. There is. Why don't the solutions from typical curve fitting programs yield that better solution? Because the typical solutions are not done with the measured values but with transformed values. The logarithm is taken of each side of the power equation yielding the transformed equation

$$\ln(y) = \ln(A) + B\ln(x)$$

which is linear in ln(y) and ln(x). When I do the problem at hand with the logarithms of the number of digits as the independent variable and the logarithms of the elapsed time as the dependent variable and solve using the linear function I get

Intercept = -5.32302714963 where the intercept from the linearized solution is the logarithm of the coefficient A of the power function. Then, A = e^intercept = 0.487796502271E-3;

Slope = B = 1.70772919479 = the exponent, and

r = 0.997876797805 which is the same answer as that received from the power function solution. The residuals are

Table 2 - Residuals for a Linear Fit with a Transformed Equation

ln(Digits)	ln(Time)	Residuals
3.401...	0.693...	+0.207850...
4.499...	2.197...	-0.164204...
5.298...	3.526...	-0.198703...
6.907...	6.536...	+0.063143...
7.056...	6.818...	+0.091914...

The sum of the residuals using the complete values in the machine is 1E-10, very near zero as it should be. The sum of the squares of the residuals is. 0.12208... Thus, the solution of the linearized function does appear to be a least squares solution in the linearized coordinate system. The transformation back to the original coordinate system does NOT yield a least squares solution in the original coordinate system. I suspect that we all knew that it might not be. We hoped that it would be close. In some cases such as the Knapp data it is not close at all.

Over many years I had encountered difficulties such as this in my work with power function fitting, but I didn't know what to do about it. The subject hadn't been discussed by Professor Eggers during his course on curve fitting. Then back in 1983, as editor of TI PPC Notes, I received a submission of an article from my friend and mentor George Wm. Thomson. His solution to the problem is derived from a 1943 book by W. Edwards Deming [Reference 2]. That is the same Deming who became famous after World War II by his introduction of statistical process control into Japanese industry. Instead of simply solving with the transformed data, the improved solution is obtained by weighting each transformed data point by the square of the y values. Thomson's solution appears in the volume 8 number 1 issue of TI PPC Notes [Reference 3]. I could not find a similar capability in the HP48S or for any other HP calculators in my collection. I am not saying the capability does not exist but only that I could not find it. I wrote a program which includes that capability on the HP-35s -- see the appendix. Why didn't I write a program for my HP-48S? The answer is lack of sufficient familiarity with RPL. A weighted solution for the Knapp data from the HP-35s program yields the following results:

A = 0.001512745...

B = 1.886571961...

r = 0.999895601...

The residuals are:

Table 3 - Improved Pi Run Time Residuals

Digits	Time	Residuals
30	2 minutes	1.074319...
90	9 minutes	1.644963...
200	34 minutes	0.823973...
1000	690 minutes	-1.005615...
1160	915 minutes	0.705279...

and the sum of the residuals is 3.242919... and the sum of the squares of the residuals is 6.0476825.... That is a much better solution than the one obtained with weighted data.

A Second Power Function Solution Using Weighted Data

My second example of the advantage of using weighted data in the solution for a power function to fit data is not from experimental data. Rather, the problem is a contrived one which provides a better example of just how much the use of a weighted solution can improve the result. To develop the problem I begin with the function y = Ax^B where A = 1.25 and B = 2.5 . Then I generate a table of eight data pairs as in the first two columns of the table by first calculating the exact values for y as a function of x. To introduce some noise in the data pairs to be used in the curve fitting exercise I define each y value to be the nearest integer to the exact y value. See the second and third columns in the table. The fourth column in the table labeled "Error" is the difference between the integerized y values and the exact values. One would expect that the residual errors from a least squares power function curve fit would look something like the induced errors. The weighted solution using the HP-48S or the HP-35s program is

A = 1.226080166

B = 2.507529704

r `= 0.999994365

When I use the SCATR and FCN functions of my HP-48S to superimpose a plot of the function on a scatter diagram of the input data I do not see any points which are clearly not on the curve. This is not because there aren't any such points but because the scale of the plot doesn't allow the errors to appear to the eye.

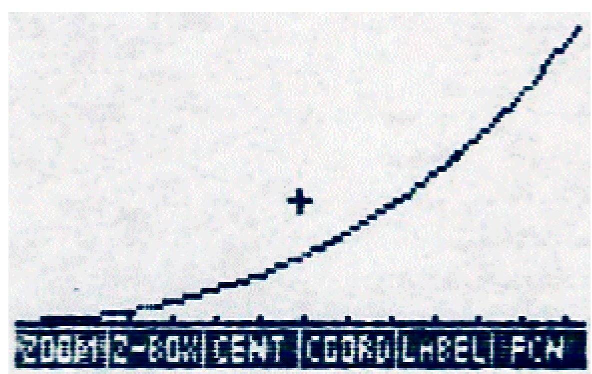

Fig. 2 - SCATR and FCN plot for the second power function problem

The residuals for the weighted solution in the fifth column are similar to the residuals of the weighted power fit to the Knapp data in the sense that the residuals for the larger y values are clearly not close the curve defined by the solution function. The residuals show no obvious similarity to the errors in the contrived function. A linear fit with the errors in the contrived function as the independent variable and the residuals from the weighted power fit as the dependent variable yields an intercept of 0.899..., a slope of 0.386... and a correlation coefficient of 0.0346... Those numbers confirm the idea that there is no linear relationship between the residual and the input errors.

The weighted solution using the HP-35s program is

A = 1.250262080 B = 2.499926401 r = 0.999999844

The residuals for the weighted solution in the sixth column of the table show an obvious similarity to the errors in the contrived function. A linear fit with the errors in the contrived function as the independent variable and the residuals from the weighted power fit as the dependent variable yields an intercept of -0.004969..., a slope of 0.991604... and a correlation coefficient of 0.99864.. Those numbers confirm the idea that there is a close relationship between the residual and the input errors. The weighted solution yields a fit that is what would be expected.

Table 4 - Residuals for a Second Power Fit Problem

x	y	y exact	Error	Weighted Residuals	Weighted
			Residuals		
2	7	7.071...	-0.071..	0.027949	-0.072189...
3	19	19.485...	-0.485...	-0.271458...	-0.488081...
5	70	69.877...	0.123...	0.624373...	0.116503...
7	162	162.052...	-0.052...	0.702622...	-0.063032...
10	395	395.284...	-0.284...	0.498579...	-0.300588...
13	762	761.672...	0.328...	0.333399..	0.311400...
16	1280	1280	0	-1.992519...	-0.007145...
20	2236	2236.067....	-0.067...	-7.314829...	-0.043737
	Sum of Residuals			-7.391882....	-0.546870...
	Sum of Squares of Residuals			0.58.794596..	0.450268....

A Weighted Solution for Exponential Functions

Deming's book indicates that weighting with the square of the dependent variable is also appropriate when using linearized procedures to fit the exponential function Y = Ae^Bx. The program in the appendix includes options to perform curve fitting of an exponential function without and with y^2 weighting. I do not have a set of experimental data to submit to the program. Instead, I generated a contrived set of exponential data in the same manner that I used earlier to obtain a contrived set of data for the investigation of weighting with a power function.

To develop a function to demonstrate the use of a weighted solution for an exponential fit I begin with the function y = Ae^Bx with A = 2 and B = 0.4 . Then I generate a table of eight data pairs by first calculating the exact values for y as a function of x. To introduce some noise in the data pairs to be used in the curve fitting exercise I define each y value to be the nearest integer to the exact y value. See the second and third columns in the table. The fourth column in the table labeled "Error" is the difference between the integerized y values and the exact values. One would expect that the residual errors from a least squares

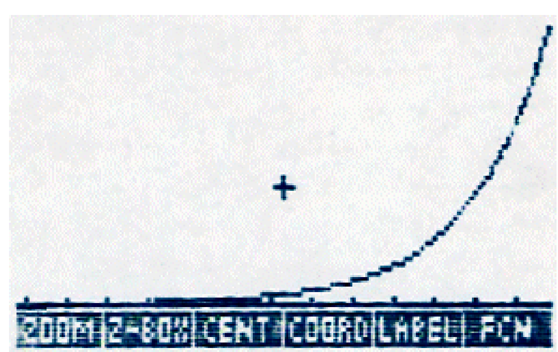

Fig. 3 - SCATR and FCN plot for the exponential function problem.

exponential function curve fit would look something like the induced errors. The weighted solution using the HP-48S or the HP-35s program is:

A = 1.941103046 B = 0.402224248 r = 0.999892737

When I use the SCATR and FCN functions of my HP-48S to superimpose a plot of the function on a scatter diagram of the input data I do not see any points which are clearly not on the curve. As with the situation with the second power function problem this is not because there aren't any such points but because the scale of the plot doesn't allow the errors to appear to the eye.

The residuals for the weighted solution in the fifth column of the table are similar to the residuals of the weighted power fit problems in the sense that the residuals for the larger y values are clearly not close the curve defined by the solution function. The residuals show no obvious similarity to the errors in the contrived function. A linear fit with the errors in the contrived function as the independent variable and the residuals from the weighted power fit as the dependent variable yields an intercept of -3.638..., a slope of -15.587... and a correlation coefficient of -0.316... Those numbers confirm the idea that there is no close relationship between the residual and the input errors.

The weighted solution using the HP-35s program is

A = 2.001474365 B = 0.399961991 r = 0.999999907

The residuals for the weighted solution in the sixth column of the table show a similarity to the errors in the contrived function but not to the same extent as with the power function solution earlier in this paper. A linear fit with the errors in the contrived function as the independent variable and the residuals from the weighted power fit as the dependent variable yields an intercept of -0.0479.., a slope of 0.833... and a correlation coefficient of 0.95772. The weighted solution does yield a substantial reduction in the residuals relative to those for the weighted solution.

Table 5 - Residuals for an Exponential Fit Problem

x	y	y exact	Error	Weighted Residuals	Weighted Residuals
1	3	2.983.. .	0.016...	0.097766...	0.014264...
2	4	4.451...	-0.451...	-0.339264...	-0.454024...
4	10	9.906...	0.093..	0.299733...	0.088139...
6	22	22.046...	-0.046...	0.315409...	-0.057574
8	49	49.065...	-0.065...	0.524894...	-0.086302..
11	163	162.901	0.098...	0.979460...	0.046319...
15	807	806.857...	0.142...	-2.664615...	0.007838...
18	2679	2678.861...	0.138...	-27.178713...	-0.002834...
Sum of residuals				-27.965328...	-0.444173...
Sum of residuals squared				747.331463..	0.227088.

Conclusions

The use of the weighted solutions yields a substantial improvement in the least squares solutions obtained with power and exponential functions.

There are cases in which using a weighted solution may not be appropriate. For example, Deming[(2)]

(page 201) suggests that weighting is not needed if the dependent variable does not change very much. A weighted solution would also not be expected to be appropriate if the error in the dependent values vary with the magnitude of the dependent values. I do not have actual data for such a case. As a possible example I suggest that if the dependent values were electrical values taken with one of those meters which change ranges as the measurements increase, then larger residuals would be expected at larger dependent variable values.

REFERENCES:

1. William Kolb, Curve Fitting for Programmable Calculators Second Edtion, IMTEC, 1983.
2. W. Edward Deming, Statistical Adjustment of Data, Dover Books, 1964. This was a republication of the book published by John Wiley & Sons in 1943.
3. George W. Thomson, "Thoughts about Curve Fitting", TI PPC Notes, V8N1P28-31.

APPENDIX A: Unweighted and Weighted Curve Fitting for the Power and Exponential Functions on the HP-35s

The weighting with this program is limited to be the square of the dependent variable.

The weighted solution will yield values for N which are not integers if the dependent variables include numbers which are not integers. The built-in statistical solutions of the HP-35s will return the error message STAT ERROR if N is not an integer. The HP-33s does the same. Because of that idiosyncrasy it is necessary to solve the weighted case by directly implementing the necessary equations in the program rather than using the built-in linear solution. See program steps Z093 through Z134. Machines that will solve when N is not an integer include the HP-10B, HP-11C, and HP-12C.

Nomenclature:

Equations: y = Ax^B ; y = Ae^BX

Flag 1 is set for a power function solution. Flag 1 is reset for an exponential solution

A = Coefficient	Y = Dependent variable	D = Residual = $Y_{observed}$ - $Y_{calculated}$
B = Exponent	R = Correlation coefficient	S = Sum of the residuals
X = Independent variable	N = Number of data pairs	T = Sum of the Residuals squared

HP-35s Program Listing

Z001 LBL Z
Z002 SF 10
Z003 EQN WEIGHTED CURVE FIT
Z004 CF 10
Z005 0
Z006 STO I
Z007 INPUT N
Z008 STO A
Z009 1
Z010 STO+ I
Z011 INPUT X
Z012 STO (I)
Z013 1
Z014 STO+ I
Z015 INPUT Y
Z016 STO(I)
2017 DSE A
2018 GTO Z009
Z019 SF 1
Z020 SF 10
Z021 EQN 1=PWR, 2=EXP
Z022 CF 10
Z023 1
Z024 x=y?
Z025 GTO Z031
Z026 x<>y
Z027 2
Z028 x ≠ y ?
Z029 GTO Z019
Z051 STO X
Z052 1
Z053 STO+ I
Z054 RCL (I)
Z055 STO Z
Z056 STOx Z
Z057 LN
Z058 STO Y
Z059 -27
Z060 STO J
Z061 RCL Z
Z062 STO+ (J)
Z063 1
Z064 STO- J
Z065 RCL X
Z066 RCLx Z
Z067 STO+ (J)
Z068 1
Z069 STO- J
Z070 RCL Y
Z071 RCLx Z
Z072 STO+ (J)
Z073 1
Z074 STO- J
Z075 RCL X
Z076 x^2
Z077 RCLx Z
Z078 STO+ (J)
Z079 1
Z101 n
Z102 Σx^2
Z103 x
Z104 Σx
Z105 x^2
Z106 -
Z107 /
Z108 STO B
Z109 Σx
Z110 x
Z111 Σy
Z112 -
Z113 +/-
Z114 n
Z115 /
Z116 e^x
Z117 STO A
Z118 RCL R
Z119 n
Z120 Σx^2
Z121 x
Z122 Σx
Z123 x^2
Z124 -
Z125 n
Z126 Σy^2
Z127 x
Z128 Σy
Z129 x^2
Z151 e^x
Z152 GTO Z154
Z153 y^x
Z154 RCLx A
Z155 1
Z156 STO+ I
Z157 R↓
Z158 RCL- (I)
Z159 +/-
Z160 STO+ S
Z161 ENTER
Z162 x^2
Z163 STO+ T
Z164 R↓
Z165 STO D
Z166 VIEW D
Z167 DSE C
Z168 GTO Z144
Z169 VIEW S
Z170 VIEW T
Z171 GTO Z031
Z172 CL Σ
Z173 RCL N
Z174 STO A
Z175 0
Z176 STO I
Z177 1
Z178 STO+ I
Z179 RCL (I)

Z030 CF 1
Z031 SF 10
Z032 EQN 1=WITH WTS, 2=WITHOUT
Z033 CF 10
Z034 2
Z035 x=y?
Z036 GTO Z172
Z037 x<>y
Z038 1
Z039 x ≠ y ?
Z040 GTO Z031
Z041 CL Σ
Z042 RCL N
Z043 STO A
Z044 0
Z045 STO I
Z046 1
Z047 STO+ I
Z048 RCL (I)
Z049 FS? 1
Z050 LN

Z080 STO- J
Z081 RCL Y
Z082 x²
Z083 RCLx Z
Z084 STO+ (J)
Z085 1
Z086 STO- J
Z087 RCL X
Z088 RCLx Y
Z089 RCLx Z
Z090 STO+ (J)
Z091 DSE A
Z092 GTO Z046
Z093 n
Z094 Σxy
Z095 x
Z096 Σx
Z097 Σy
Z098 x
Z099 -
Z100 STO R

Z130 -
Z131 x
Z132 SQRT x
Z133 /
Z134 STO R
Z135 VIEW A
Z136 VIEW B
Z137 VIEW R
Z138 RCL N
Z139 STO C
Z140 0
Z141 STO S
Z142 STO T
Z143 STO I
Z144 1
Z145 STO+ I
Z146 RCL (I)
Z147 RCL B
Z148 FS? 1
Z149 GTO Z153
Z150 x

Z180 FS? 1
Z181 LN
Z182 1
Z183 STO+ I
Z184 R↓
Z185 RCL (I)
Z186 LN
Z187 x<>y
Z188 Σ+
Z189 DSE A
Z190 GTO Z177
Z191 b
Z192 e^x
Z193 STO A
Z194 m
Z195 STO B
Z196 r
Z197 STO R
Z198 GTO Z135
Z199 STOP