DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANITOBA

SUPPLEMENTARY

NOTES

for

136.382 : INTRODUCTION TO MATHEMATICAL MODELLING

Table of Contents

CHAPTER 1 - INTRODUCTION

1.1 - THE PROCESS OF MATHEMATICAL MODELLING

Mathematical Modelling is concerned with the development of a mathematical system to describe (within some acceptable limits) a real-world phenomenon. This process may be represented schematically as in the following diagram:

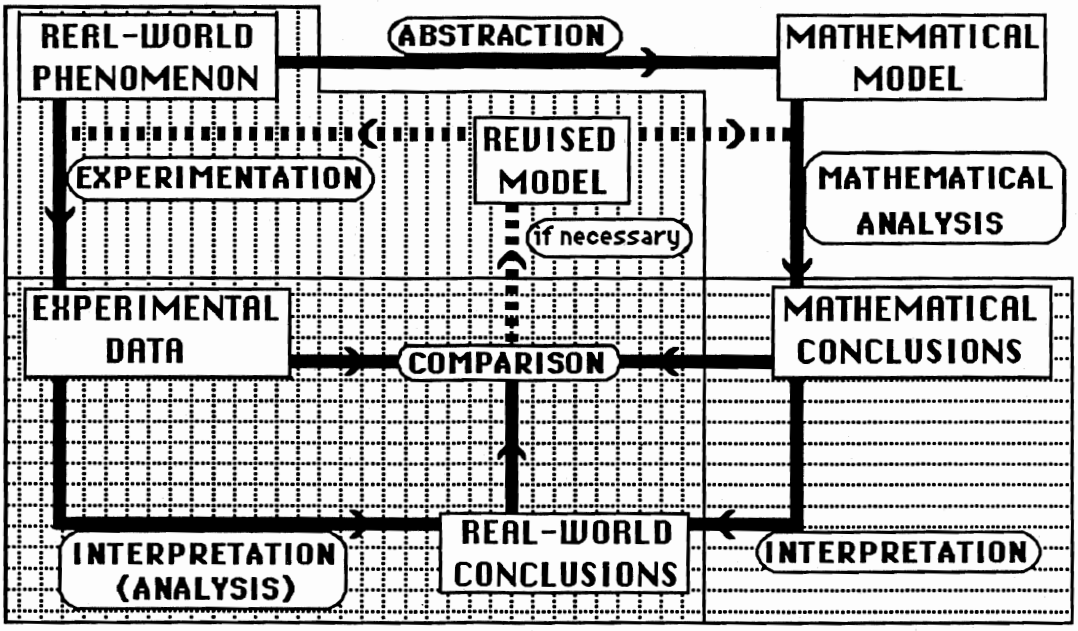


Figure 1 : THE PROCESS OF MATHEMATICAL MODELLING

The section at the left side of this diagram, shaded by vertical lines, may be thought of as representing (in an over-simplified manner) **"empirical science"** or **"empirical modelling"**, in which experimental data is collected by a scientist in the construction of an "empirical model" for the phenomenon. It should be noted that the last step, of comparing "real-world conclusions" and experimental data, may suggest new experiments that should be performed in an effort to verify or disprove "real-world conclusions" for which no existing data may be found. Thus, it may be necessary to pass around this loop several times before the "FINAL CONCLUSIONS" may be drawn. It should be pointed out that empirical models are generally **descriptive** models, and in most instances are **not predictive**.

However, as can be seen from the above (again over-simplified) diagram, the process of mathematical modelling is much more complicated and elaborate than empirical modelling, but includes the latter as one of its building blocks. Below is given a brief description of the various additional procedures used in the development of a mathematical model:

ABSTRACTION:
In order to set up a mathematical model of a real-world problem it is necessary to

(a) identify and define the crucial variables

and

(b) formulate appropriate mathematical relations amongst them.

Usually this process involves making ASSUMPTIONS, either out of necessity or merely for convenience (simplicity). It is crucial to record these assumptions concisely for future reference, so that any limitations of the model are clearly understood.

MATHEMATICAL ANALYSIS:
This analysis involves the use of the unambiguous terms of mathematics and logical deduction, and thus introduces a clarity of thought not necessarily available through use of other techniques. However, it may be necessary to introduce additional ASSUMPTIONS at this stage, again either for convenience or as the result of the introduction of specific mathematical techniques (e.g. approximation techniques). Again it is crucial to record these assumptions in unambiguous terms to avoid confusion at a later time.

INTERPRETATION:
At this step it is necessary to recall the original real-world problem, the process of abstraction used to develop the model, and the assumptions made in earlier stages in order to provide meaningful "real-world conclusions" based on the model being studied.

PERHAPS THE MOST IMPORTANT STEP IN THE PROCESS OF MATHEMATICAL MODELLING OCCURS IN THE REGION AT THE BOTTOM OF Figure 1 (shaded by horizontal lines) IN WHICH THE CONCLUSIONS OF THE MODEL ARE DRAWN AND REVISION OF THE MODEL OCCURS (IF NECESSARY). THIS SECTION IS REFERRED TO AS **"MODEL TESTING"** AND MAY INVOLVE A NUMBER OF ITERATIONS IN ORDER TO OBTAIN AN ACCEPTABLE MATHEMATICAL MODEL.

MODEL TESTING:

(a) If the agreement between the conclusions of the model and the experimental data is "GOOD" (note that this may be a very subjective evaluation), then the modeller may make the decision to end the process, feeling he has adequately captured the essential features of the phenomenon in his model. However, he must not lose sight of the inherent limitations of his model as a result of the assumptions made in arriving at it. Any undesirable features of the model suggest that a

revision of the model is necessary. In addition, if the model exhibits properties for which there is no experimental evidence, it may be advisable to devise additional experiments to test these conclusions.

(b) "POOR" agreement between the model conclusions and the experimental data may suggest that the model should be abandoned, or further analysis may give rise to underlying assumptions, which have not previously been recognized, which suggest major modifications to the model in an effort to obtain a more realistic model.

(c) "FAIR" agreement may allow for significant improvement to the model merely through one or more minor revisions to the model.

CAUTION: Before any revision of the model is attempted, it is wise to check for errors, either logical or mathematical, in the analysis. In addition, when approximation techniques have been used, it may be worthwhile to question their appropriateness and if necessary seek alternative, and hopefully more accurate and appropriate techniques.

If revision of the model is necessary, for any reason whatsoever, the whole process is repeated until an acceptable model is obtained. Subsequent revisions may lead to a better agreement, or may even provide additional conclusions which suggest more experiments for the empirical scientist to perform in an effort to verify or disprove the model.

1.2 - GUIDING PRINCIPLES OF MATHEMATICAL MODELLING

Although it is desirable to construct mathematical models which represent reality as completely and accurately as possible, there are several principles of modelling which are commonly followed:

THE SIMPLICITY PRINCIPLE:
Keep in mind the intended use of the model, and employ the SIMPLEST possible model in relation to these intended uses.

THE FAMILIARITY PRINCILPLE:
Mathematical modelling may give rise to the same basic mathematical structure for completely different physical phenomena, and hence familiarity with one of these immediately imposes at least some familiarity with the other(s). For example, simple spring-mass systems and simple electrical circuits are governed by the same basic mathematical equations, although the notations used to describe these systems are usually not the same. Thus, these systems provide very useful analogues for each other and knowledge of one of them is helpful in understanding the other.

THE APPROXIMATION PRINCIPLE:
Keep in mind the experimental error involved in obtaining experimental data when mathematical approximation techniques are introduced into a model. There is no point in using a mathematical approximation technique with an error

of less than 0.1% in an effort to describe, through the model, experimental data with a known experimental error of 5%. Of course, this principle is only useful when the error in experimental data is known, and moreover when it can be assumed to be a "random error". (Of course, with appropriate analysis, "systematic errors" can be removed from the data.)

1.3 - THE CLASSIFICATION OF MATHEMATICAL MODELS

As pointed out earlier, **empirical models are generally descriptive** in nature.

On the other hand, **mathematical models are generally predictive** in nature, and may be used to study the future state of a given system. Nonetheless, there are basically two types of mathematical models which can be constructed to represent physical phenomena. These may be classified as either **deterministic** or **probabilistic** models. Below is given a brief description of these two types of models:

DETERMINISTIC MODELS:
In a deterministic model full knowledge of the present "state" of the system and of all present and subsequent "forces" acting on the system is sufficient to determine the state of the system at all future times.

Such models are commonly used to describe phenomenon in the physical sciences such as Physics and Engineering. The techniques most commonly used in such models are those of the theory of differential equations.

This type of model, when used to describe some phenomenon, (such as social phenomenon), which are less well understood than those mentioned above, is subject to a number of criticisms such as:

(a) such models do not allow for an expression of "free will",
(b) problems concerning ecosystems usually involve variables which are not continuous in time (such as population size), and hence cannot be "successfully" modelled through the use of differential equations.

Even for purely physical science problems to which the above criticisms do not apply, the Heisenburg Uncertainty Principle (commonly associated with Quantum Mechanics) is often raised as an objection against the use of a deterministic model. This principle, in very loose form, states that "the operation of performing a physical measurement on a physical system affects the state of the system, so that "full" knowledge of the state of a system is unattainable."

PROBABILISTIC MODELS:
In a probabilistic model, it is usually assumed that the system can only occupy a certain number of known distinct "states" at any time each with its own probability. If the probability distribution of the system for these states at a given time is known, then the probability distribution for these states at any later time can be determined, through knowledge of the subsequent "forces" acting on the

system. Thus, the state of the system at subsequent times is never known with certainty, although the probability distribution for the admissible states is known.

Most probabilistic models usually involve discrete variables, as well as continuous variables (such as time), and hence are expressed in terms of differential and difference equations and sometimes in terms of differential-difference equations.

1.4 - A COMMENT CONCERNING THE STRUCTURE AND PURPOSE OF THIS COURSE

Most modellers regard mathematical modelling as a skill, or even an "art" (as opposed to a "science"), which is learned by means of example. Although mathematical analysis plays a fundamental role in the modelling process, it is often the most straightforward part of the process. Thus, throughout this course we will primarily employ mathematical techniques with which you are already familiar, so that we may concentrate our efforts on the more "artistic" features of the process of mathematical modelling. However, whenever necessary, we will review mathematical techniques, or even introduce new ones briefly, in order to aid in the development and analysis of mathematical models.

CHAPTER 2 - "CURVE-FITTING" (THE REPLICATION OF "EXPERIMENTAL " DATA)

2.1 - INTRODUCTION

In many cases, the experimental scientist or mathematical modeller has at his disposal a collection of experimental data, and he would like to obtain a mathematical function which reproduces this data either accurately, or more often only approximately, so that he may more easily make use of this data. Although this procedure is not itself mathematical modelling, it often plays a fundamental role in modelling. In this chapter, we shall discuss several techniques applicable to this problem of "replication of data".

The first such "curve-fitting" technique discussed is known as the "polynomial fit" for the given data, and is used when a precise replication of the data is desired.

Moreover, we shall discuss a family of procedures (applicable to a number of classes of functions), known collectively as "least-squares" techniques, which may be applied to find a function which approximates the given data. These techniques are characterized by the property that, within a prescribed class of functions, they determine the most appropriate choice for the function (within the prescribed class) to replicate the data.

Finally, a brief (and non-rigourous) discussion is presented concerning how finite differences and the corresponding difference tables may be used as a guide in determining the most appropriate degree of a polynomial function to be used to approximate a given set of data using the method of least-squares, under the assumption that a polynomial replication is desired.

2.2 - POLYNOMIAL FIT

To illustrate the procedure for a polynomial fit, suppose we wish to find a function $y = y(x)$ which agrees precisely with the data in Table 1 , it being assumed that these data are themselves precise:

x	y
-1	-8
0	3
1	6
2	31

TABLE 1 : DATA FOR "POLYNOMIAL FIT"

Since there are four given data points, it seems reasonable to assume that this data can be represented by a function $y = y(x)$ which contains four unknown constants (parameters). Thus if we assume that the desired function is a polynomial, under the simplicity principle, the most appropriate choice is merely a cubic polynomial of the form

(1) $$y(x) = ax^3 + bx^2 + cx + d\,,$$

in which a , b , c and d are unknown constants whose values are to be determined.

To determine the constants appearing in (1) , we note that each of the four points (x, y) of Table 1 must lie on the curve represented by (1) , so that we must have

$$\begin{aligned} -a + b - c + d &= -8 \\ d &= 3 \\ a + b + c + d &= 6 \\ 8a + 4b + 2c + d &= 31 \end{aligned}$$

which is a linear system of four equations in the four unknowns a , b , c and d . This system may be solved by a variety of techniques and has the unique solution

$$a = 5\,,\ b = -4\,,\ c = 2\,,\ d = 3\,.$$

Thus, for the above set of data, the desired cubic polynomial is simply

$$y(x) = 5x^3 - 4x^2 + 2x + 3\,.$$

More generally, it can be shown that, for n any positive integer, given n data points (x_1, y_1) , (x_2, y_2) , ... , (x_n, y_n) , with $x_i \neq x_k$ for $i \neq k$, there is precisely one polynomial of degree (n-1) of the form

(2) $$y(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0\,,$$

(in which a_{n-1} , a_{n-2} , ... , a_1 , a_0 are n unknown constants) whose graph passes through the n given data points. Indeed, this polynomial is given by the LAGRANGE INTERPOLATION FORMULA , namely

(3) $$y(x) = \sum_{i=1}^{n} C_i^* \, y_i \quad ,$$

with

$$C_i^* = \frac{(x-x_1)(x-x_2)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_n)}{(x_i-x_1)(x_i-x_2)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_n)}\,.$$

Although we shall not attempt to derive the above result, the following observations should be noted:

(i) this is clearly a polynomial of degree (n-1) , since each of the coefficients C_i^* is a polynomial of degree (n-1) ,

and

(ii) for $x = x_k$ ($k = 1,2,...,n$), $y(x_k) = y_k$, as required.

To demonstrate the use of Lagrange's Formula, in the case of the data of Table 1 , we merely let $n = 4$, $(x_1, y_1) = (-1,-8)$, $(x_2, y_2) = (0,3)$, $(x_3, y_3) = (1,6)$, $(x_4, y_4) = (2,31)$. Thus, in this case, (3) provides

$$y(x) = \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}\, y_1 + \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}\, y_2$$

$$+ \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}\, y_3 + \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}\, y_4$$

$$= \frac{x(x-1)(x-2)(-8)}{(-1)(-2)(-3)} + \frac{(x+1)(x-1)(x-2)(3)}{1(-1)(-2)} + \frac{(x+1)x(x-2)(6)}{2(1)(-1)}$$

$$+ \frac{(x+1)x(x-1)(31)}{3(2)(1)}$$

$$= \frac{4}{3}(x^3-3x^2+2x) + \frac{3}{2}(x^3-2x^2-x+2) - 3(x^3-x^2-2x) + \frac{31}{6}(x^3-x)$$

$$= 5x^3 - 4x^2 + 2x + 3 ,$$

as obtained earlier.

2.3 - THE METHOD OF "LEAST-SQUARES" FOR LINEAR FUNCTIONS

Consider the problem of replicating the "experimental" data presented in Table 2 below:

x	y
1	6.05
2	8.32
3	10.74
4	13.43
5	15.90
6	18.38
7	20.93
8	23.32
9	24.91
10	28.36

TABLE 2 : DATA FOR LEAST-SQUARES LINEAR APPROXIMATION

Clearly, using the results of the previous section, it is possible to fit this data precisely with a polynomial of degree 9. However, two objections to this procedure immediately are evident , namely:

(i) the evaluation of the coefficients of this polynomial is extremely tedious, even if we use the Lagrange Interpolation formula,

and

(ii) since the data in Table 2 is described as "experimental", it is probably subject to experimental error (either random or systematic) and hence an attempt to duplicate the data precisely will not only duplicate the data itself but also any errors built into the experiment.

Thus it seems prudent to merely attempt to approximate this data through some function, chosen to exhibit the "trend" of the data.

To identify the appropriate type of function to use for this approximation, we plot the data of Table 2 as in Figure 2 .

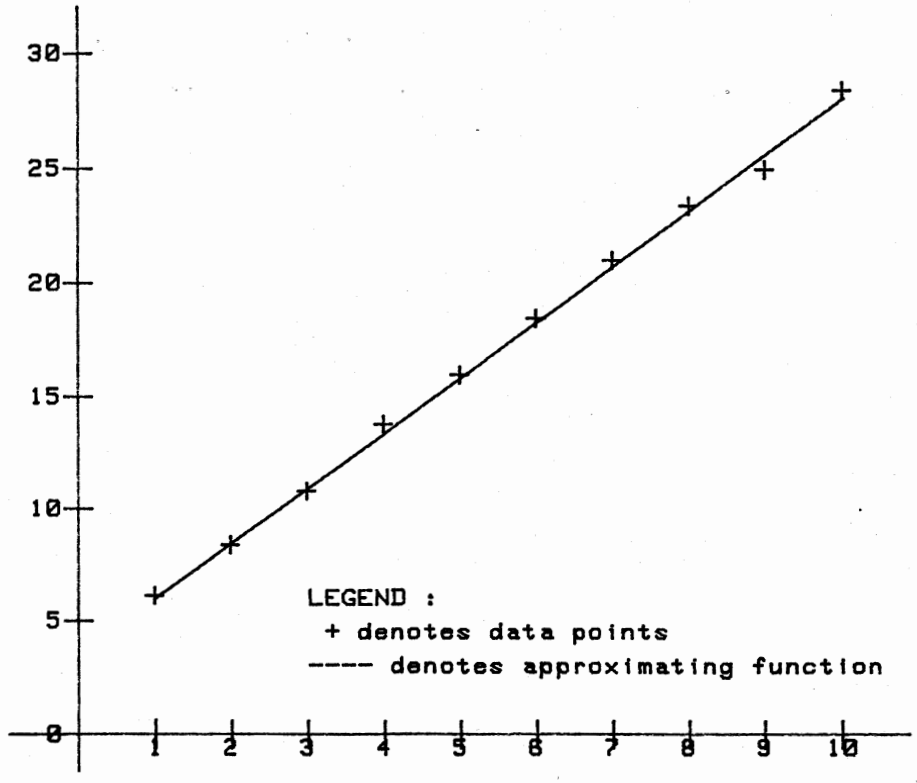


Figure 2 : A LINEAR LEAST-SQUARES FIT TO GIVEN DATA

A casual glance at Figure 2 suggests that this data could be approximated quite reasonably by means of a linear function

(3) $$y(x) = ax + b\,,$$

for some unknown constants a and b .

The problem may then be expressed simply as that of determining a and b in order to obtain a "best fit" from amongst all linear functions of the form (3) . The question that arises then is **"what criterion shall we use in order to determine when the "best fit" has been obtained?"** Although we could choose any criterion we wished, subject only to its suitability, we shall adopt the following:

Suppose that the data points are denoted by (x_i, y_i) for $i = 1,2,...,n$ with n a positive integer (in the above case $n = 10$), and let

$$y(x_i) = ax_i + b$$

denote the calculated values of y corresponding to the values x_i (for $i = 1,2,..,n$), under the ASSUMPTION that the given data may be represented by the linear function (3) . Then, the function

(4) $$S = \sum_{i=1}^{n} (y(x_i) - y_i)^2$$

$$= \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

is a function only of the unknowns a and b , since x_i and y_i are given constants,

i.e., $$S = S(a,b) .$$

Since S(a,b) represents the sum of the squares of the differences between corresponding observed values y_i and calculated values $y(x_i)$ for the range of values x_i (i = 1,2,..., n), and since it is clear that

> if $y(x_i) = y_i$ for i = 1,2,...,n (which would be the case if the fit were precise) then S(a,b) = 0 ,

while

> if $y(x_i) \neq y_i$ then S(a,b) > 0 ,

we adopt as our **optimality criterion** the requirement that

> **S(a,b) should attain a minimum value for appropriate choices of a and b .**

However, we recall that for a function of two variables to attain a minimum value it is NECESSARY that its partial derivatives vanish simultaneously. Thus, for the desired optimality criterion to be satisfied we require

(5) $$\frac{\partial S}{\partial a} = 0 \qquad \text{and} \qquad \frac{\partial S}{\partial b} = 0$$

or equivalently, by differentiation of (4)

$$2 \sum_{i=1}^{n} (ax_i + b - y_i)\, x_i = 0$$

$$2 \sum_{i=1}^{n} (ax_i + b - y_i) = 0 .$$

The latter two conditions may be written in the form of a linear system of two equations in a and b , namely

(6) $$\left(\sum_{i=1}^{n} x_i^2\right) a + \left(\sum_{i=1}^{n} x_i\right) b = \sum_{i=1}^{n} x_i y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a + n\,b = \sum_{i=1}^{n} y_i$$

in which the coefficients may be evaluated entirely in terms of the given data (x_i, y_i) $(i = 1,2,...,n)$.

Specifically, in the case of the data appearing in Table 2, we find that the coefficients of the system (6) may be evaluated as follows:

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1) = \frac{10(11)(21)}{6} = 385$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} i = \frac{1}{2}n(n+1) = \frac{10(11)}{2} = 55$$

$$n = 10$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} i\, y_i = y_1 + 2y_2 + 3y_3 + ... + 10y_{10} = 1139.27$$

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + ... + y_{10} = 170.34$$

so that we need only require that

$$385a + 55b = 1139.27$$
$$55a + 10b = 170.34$$

for which the (unique) solution is

$$a = \frac{2034}{825} \qquad b = \frac{292105}{82500}$$

or approximately

$$a = 2.453333333\,, \qquad b = 3.540666667\,.$$

Thus, the "best-fitting" least-squares linear function for the data in Table 2 is given by

(7) $$y(x) = 2.453333333\,x + 3.540666667,$$

in which the coefficients are accurate to 10 significant figures.

Based upon the above least-squares linear function (7) , the calculated values (to 3 decimal places only) of the dependent variable y corresponding to the values exhibited in Table 2 are shown in Table 3 below.

x	y(x) [calculated to 3 decimal places]	y [given]
1	5.994	6.05
2	8.447	8.32
3	10.901	10.74
4	13.354	13.43
5	15.807	15.90
6	18.261	18.38
7	20.714	20.93
8	23.167	23.32
9	25.621	24.91
10	28.074	28.36

TABLE 3 : APPROXIMATE CALCULATED VALUES, USING A LEAST-SQUARES LINEAR FUNCTION, FOR THE DATA IN TABLE 2

Remark: The results of Table 3 are presented solely to exhibit the "fit" of the least-squares linear function (7) to the given data of Table 2 . Since their only purpose is one of comparison, we have arbitrarily chosen to display 3 decimal places. **It should be noted that these values should not be used in any subsequent calculations, because they are merely approximations to the actual values as calculated through the use of (7) . Whenever additional calculations are to be performed, the desired values are calculated directly from (7) , with the full limits of the computing device employed to avoid the introduction of spurious errors.**

For completeness, the graph of the least-squares linear function (7) , approximating the data of Table 2 , has also been included on Figure 2 .

As a measure of the "fit" of the "least-squares" linear function (7) to the data given in Table 2 , it is easily shown that the "**residual sum of squares**" S , as defined by (4) , for the linear function (7) is

$$(8) \qquad S(a,b) = \sum_{i=1}^{n} (2.453333333\, x_i + 3.540666667 - y_i)^2 \quad (\text{with } n = 10)$$

$$= .8661733333 \quad [\text{to 10 significant figures}] .$$

Various comments concerning the significance of the value of this quantity will be made as further examples are considered. However, it can be noted, even at this early stage, that at least theoretically, there is no linear function that can provide a smaller value to S , than that exhibited in (8) , provided that all calculations are performed to infinite precision. Thus, it is essential that all calculations be made, to the full limits of

the calculator, and be based on actual calculated values in relation to (7) rather than on the approximate values as shown in Table 3 .

2.4 - THE METHOD OF LEAST-SQUARES FOR POLYNOMIAL FUNCTIONS OF HIGHER DEGREE

Consider the data exhibited in Table 4 below:

x	y
3.00	31.5
3.25	30.4
3.50	29.2
3.75	28.1
4.00	26.9
4.25	26.4
4.50	25.3
4.75	25.2
5.00	25.1
5.25	25.2
5.50	25.4
5.75	26.3
6.00	27.0
6.25	28.2
6.50	29.3
6.75	29.9

TABLE 4 : GIVEN DATA

If this data is plotted as in Figure 3 , it is immediately evident that there is no linear function which could be used to approximate this data at all reasonably; however, it does appear that the trend of this data could be represented, to a reasonable degree of accuracy, by a quadratic function of the form

(9) $$y(x) = ax^2 + bx + c,$$

in which a , b and c are unknown constants which must yet be determined.

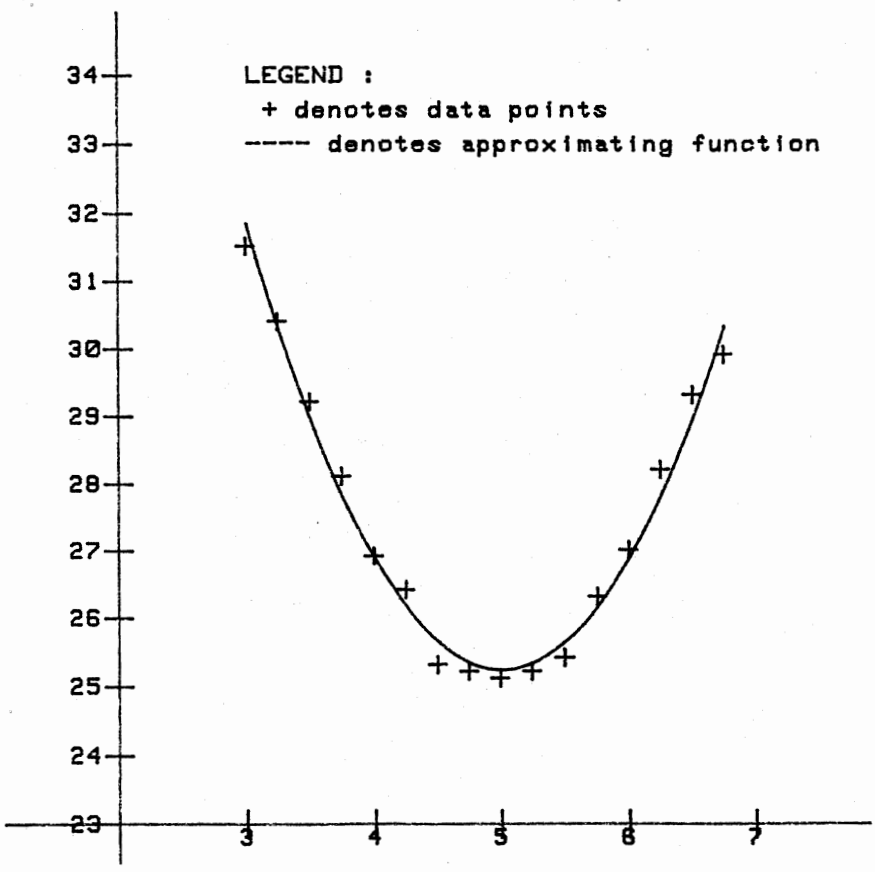


Figure 3: A QUADRATIC LEAST-SQUARES FIT

To determine the most appropriate choice of the unknown constants a , b and c appearing in (9) , we merely adopt the procedure of the last section, namely:

For n given data points (x_j, y_j) ($j = 1,2,...,n$; with n a positive integer) we denote the corresponding calculated values of y by

$$y(x_j) = ax_j^2 + bx_j + c$$

and construct the **"residual sum of squares"** (of the differences between the calculated values and observed values of y) by

(10)
$$S = \sum_{i=1}^{n} (y(x_j) - y_j)^2$$

$$= \sum_{i=1}^{n} (ax_j^2 + bx_j + c - y_j)^2$$

where, as in the previous case, it is noted that

(i) $S = S(a,b,c)$

(ii) $S(a,b,c) = 0$ if $y(x_j) = y_j$ for all $j = 1,2,...,n$

while

(iii) $S(a,b,c) > 0$ otherwise.

Thus for the **optimal choice** of a , b and c (and hence the "best-fitting" quadratic function for the data of Table 4) we adopt the criterion that

S(a,b,c) assume a minimum value for appropriate choices of a , b and c .

Hence NECESSARY conditions for optimality are

$$\frac{\partial S}{\partial a} = 0 \ , \ \frac{\partial S}{\partial b} = 0 \ \text{ and } \ \frac{\partial S}{\partial c} = 0 \, ,$$

or equivalently, by differentiation of (10) ,

$$2 \sum_{j=1}^{n} (ax_j^2 + bx_j + c - y_j)\, x_j^2 = 0$$

$$2 \sum_{j=1}^{n} (ax_j^2 + bx_j + c - y_j)\, x_j = 0$$

$$2 \sum_{j=1}^{n} (ax_j^2 + bx_j + c - y_j) = 0 \ .$$

Thus, for the "best-fitting" quadratic function (9) , it is necessary that a , b and c be a solution of the set of linear equations

(11)
$$\left(\sum_{j=1}^{n} x_j^4 \right) a + \left(\sum_{j=1}^{n} x_j^3 \right) b + \left(\sum_{j=1}^{n} x_j^2 \right) c = \sum_{j=1}^{n} x_j^2 y_j$$

$$\left(\sum_{j=1}^{n} x_j^3 \right) a + \left(\sum_{j=1}^{n} x_j^2 \right) b + \left(\sum_{j=1}^{n} x_j \right) c = \sum_{j=1}^{n} x_j y_j$$

$$\left(\sum_{j=1}^{n} x_j^2 \right) a + \left(\sum_{j=1}^{n} x_j \right) b + n\, c = \sum_{j=1}^{n} y_j \ ,$$

in which the coefficients are determined simply as the sum of powers and products of the given data x_j and y_j for $j = 1,2,...,n$.

In particular, in the case of the data exhibited in Table 4 , these coefficients may be evaluated directly to obtain

(12)
$$\sum_{j=1}^{n} x_j^4 = 12117.53125$$

$$\sum_{j=1}^{n} x_j^3 = 2164.5$$

$$\sum_{j=1}^{n} x_j^2 = 401.5$$

$$\sum_{j=1}^{n} x_j = 78.0$$

$$n = 16$$

$$\sum_{j=1}^{n} x_j^2 y_j = 10979.38125$$

$$\sum_{j=1}^{n} x_j y_j = 2133.475$$

$$\sum_{j=1}^{n} y_j = 439.4$$

Remarks:

1. As indicated earlier, when some computational device is used, it is essential to calculate these values to as high a degree of precision as possible in order to preserve the integrity of the method of least-squares.
2. Some of the above results can be calculated with less reliance on a calculator by noting that

$$x_1 = 3 = \frac{12}{4} = \frac{(1+11)}{4}$$

$$x_2 = \frac{13}{4} = \frac{(2+11)}{4}$$

$$x_3 = \frac{14}{4} = \frac{(3+11)}{4}$$

.
.
.

$$x_{16} = \frac{27}{4} = \frac{(16+11)}{4}$$

so that

$$x_j = \frac{(j+11)}{4} \text{ for } j = 1,2,...,16,$$

or equivalently

$$x_j = \frac{i}{4} \quad \text{where } i = j+11 \text{ , for } j = 1,2,...,16$$

with corresponding values of i being given by

$$i = 12,13,...,27.$$

This observation allows us to employ the following well-known results for the sums of the powers of the first n positive integers:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2},$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6},$$

$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4},$$

$$\sum_{i=1}^{n} i^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}.$$

In the present example, since j runs over the range j = 1, 2,...,16 , we may therefore write

$$\sum_{j=1}^{n} x_j = \frac{1}{4}\left(\sum_{i=1}^{27} i - \sum_{i=1}^{11} i\right)$$

$$= \frac{1}{4}\left(\frac{(27)(28)}{2} - \frac{(11)(12)}{2}\right) = \frac{624}{8} = 78$$

$$\sum_{j=1}^{n} x_j^2 = (\tfrac{1}{4})^2 \left(\sum_{i=1}^{27} i^2 - \sum_{i=1}^{11} i^2 \right)$$

$$= \frac{1}{16}\left(\frac{(27)(28)(55)}{6} - \frac{(11)(12)(23)}{6} \right) = \frac{38544}{96} = 401.5$$

$$\sum_{j=1}^{n} x_j^3 = (\tfrac{1}{4})^3 \left(\sum_{i=1}^{27} i^3 - \sum_{i=1}^{11} i^3 \right)$$

$$= \frac{1}{64}\left(\frac{(27)^2(28)^2}{4} - \frac{(11)^2(12)^2}{4} \right) = \frac{554112}{256} = 2164.5$$

and finally

$$\sum_{j=1}^{n} x_j^4 = (\tfrac{1}{4})^4 \left(\sum_{i=1}^{27} i^4 - \sum_{i=1}^{11} i^4 \right)$$

$$= \frac{1}{256}\left(\frac{(27)(28)(55)(2267)}{30} - \frac{(11)(12)(23)(395)}{30} \right)$$

$$= \frac{93062640}{7680} = 12117.53125$$

in agreement with the preceding calculations.

The linear system (11) , with coefficients given by (12) possesses the (unique) solution [to 10 significant figures]

$$a = 1.659943975$$

$$b = -16.58915964$$

$$c = 66.68043412$$

so that the least-squares quadratic estimating function for the data of Table 4 is

(13) $$y(x) = 1.659943975x^2 - 16.58915964x + 66.68043412 \ .$$

For comparison purposes the approximate calculated values y(x) , (to 3 decimal places) based on (13) , corresponding to the data of Table 4 , are shown in Table 5 , and the graph of (13) is also shown on Figure 3.

x	y(x) [calculated, approximate]	y [given]
3.00	31.852	31.5
3.25	30.299	30.4
3.50	28.953	29.2
3.75	27.814	28.1
4.00	26.883	26.9
4.25	26.159	26.4
4.50	25.643	25.3
4.75	25.334	25.2
5.00	25.233	25.1
5.25	25.340	25.2
5.50	25.653	25.4
5.75	26.175	26.3
6.00	26.903	27.0
6.25	27.840	28.2
6.50	28.984	29.3
6.75	30.335	29.9

TABLE 5 : APPROXIMATE CALCULATED VALUES, USING THE LEAST-SQUARES QUADRATIC APPROXIMATING FUNCTION (13), FOR THE DATA OF TABLE 4.

Finally, for the least-squares quadratic function (13) , the residual sum of squares (10) has the value

$$(14) \qquad S(a,b,c) = \sum_{j=1}^{n} (1.659943975x_j^2 - 16.58915964x_j + 66.68043412 - y_j)^2$$

$$= 1.016854342$$

to 10 significant digits.

Although no absolute significance has been placed on the value of the resulting residual sum of squares S by the method of least-squares, (8) and (14) suggest that the linear function (7) provides a better least-squares approximation to the data of Table 2 than the quadratic function (13) does to the data of Table 4 .

Thus, the question naturally arises as to whether or not a cubic polynomial of the form

(15) $$y(x) = ax^3 + bx^2 + cx + d,$$

for some constants a , b , c and d will provide a better fit to the data of Table 4 than the least-squares quadratic estimate (13) . Using the previous two cases as our guide, it may be shown that for the residual sum of squares

(16) $$S(a,b,c,d) = \sum_{j=1}^{n} (y(x_j) - y_j)^2$$

$$= \sum_{j=1}^{n} (ax_j^3 + bx_j^2 + cx_j + d - y_j)^2$$

[where

$$y(x_j) = ax_j^3 + bx_j^2 + cx_j + d$$

is the assumed cubic function approximation to the observed value y_j], to attain a minimum value for appropriate choices of a , b , c and d , it is necessary that

$$\frac{\partial S}{\partial a} = 0 \ , \ \frac{\partial S}{\partial b} = 0 \ , \ \frac{\partial S}{\partial c} = 0 \ , \ \text{and} \ \frac{\partial S}{\partial d} = 0 \, ,$$

or equivalently,

(17) $$(\sum_{j=1}^{n} x_j^6)\, a + (\sum_{j=1}^{n} x_j^5)\, b + (\sum_{j=1}^{n} x_j^4)\, c + (\sum_{j=1}^{n} x_j^3)\, d = \sum_{j=1}^{n} x_j^3 y_j$$

$$(\sum_{j=1}^{n} x_j^5)\, a + (\sum_{j=1}^{n} x_j^4)\, b + (\sum_{j=1}^{n} x_j^3)\, c + (\sum_{j=1}^{n} x_j^2)\, d = \sum_{j=1}^{n} x_j^2 y_j$$

$$(\sum_{j=1}^{n} x_j^4)\, a + (\sum_{j=1}^{n} x_j^3)\, b + (\sum_{j=1}^{n} x_j^2)\, c + (\sum_{j=1}^{n} x_j)\, d = \sum_{j=1}^{n} x_j y_j$$

$$(\sum_{j=1}^{n} x_j^3)\, a + (\sum_{j=1}^{n} x_j^2)\, b + (\sum_{j=1}^{n} x_j)\, c + n\, d = \sum_{j=1}^{n} y_j \, .$$

In the case of the data of Table 4 , it may be shown that this system becomes

$$\begin{aligned}
412955.9629a + \ & 69906.28125b + \ 12117.53125c + \ 2164.5d = \ 59351.01719 \\
69906.28125a + \ & 12117.53125b + \ 2164.5c + \ 401.5d = \ 10979.38125 \\
12117.53125a + \ & 2164.5b + \ 401.5c + \ 78.0d = \ 2133.475 \\
2164.5a + \ & 401.5b + \ 78.0c + \ 16d = \ 439.4
\end{aligned}$$

in which a number of the coefficients have previously been evaluated and the remaining coefficients may be evaluated by similar techniques.

This system again possesses a unique solution for a , b , c and d , and provides the least-squares cubic approximating function

(18) $y(x) = 0.0165768333x^3 + 1.417509135x^2 - 15.44671814x + 64.95209635$

for the data of Table 4 . The approximate calculated values of y , based on the cubic function (18) are shown in Table 6 below

x	y(x) [calculated, approximate]	y [given]
3.00	31.817	31.5
3.25	30.292	30.4
3.50	28.964	29.2
3.75	27.835	28.1
4.00	26.906	26.9
4.25	26.180	26.4
4.50	25.657	25.3
4.75	25.339	25.2
5.00	25.228	25.1
5.25	25.326	25.2
5.50	25.633	25.4
5.75	26.151	26.3
6.00	26.883	27.0
6.25	27.829	28.2
6.50	28.991	29.3
6.75	30.370	29.9

TABLE 6: APPROXIMATE CALCULATED VALUES, USING THE LEAST-SQUARES CUBIC APPROXIMATING FUNCTION (18), FOR THE DATA OF TABLE 4.

Remarks:

1. The very small value of the leading coefficient in (18) suggests that a quadratic function could possibly provide as good an approximation for this data as (18) does. Indeed, a casual comparison of the results of Tables 5 and 6 indicates that in half the cases the quadratic least-squares estimate (13) provides an approximation to the given data which is at least as good as that provided by the cubic least-squares estimate (18) , although in the remaining cases it provides a worse approximation.

   Thus, by virtue of the simplicity principle, it appears desirable to employ (13) rather than (18) as the least-squares estimate for the data of Table 4 .

2. For the least-squares cubic function (18) , the residual sum of squares (16) has value

   (19) $$S(a,b,c,d) = 1.010755120.$$

   A comparison of (19) and (14) indicates that (18) does indeed provide over-all a better fit to the given data than (13) , although the marginal improvement in the value of S is to a large extent counter-balanced by the significant increase in the amount of computation necessary to determine (18) . This observation therefore supports the above conclusion.

3. The analysis of the preceding three examples can easily be repeated to determine the set of necessary conditions to be satisfied by the coefficients of a polynomial of prescribed degree in order to provide a minimum value to the residual sum of squares S associated with the assumed class of polynomials (of prescribed degree). However, with the previous three cases as our guide, it is a simple matter to write down (by induction) the appropriate linear system of equations (analogous to the systems (6) , (11) and (17)) to be satisfied by the coefficients of this assumed class of polynomials.

4. The above remarks indicate that it is highly desirable to develop some procedure which will provide, under the assumption that a given set of data may be approximated by a polynomial of some degree, the appropriate degree to be used for the desired least-squares polynomial approximatilng function. We shall not address this problem now, but will return to it after we have discussed the method of least-squares for other (non-polynomial) functions.

5. The given data and the corresponding least-squares cubic function (18) are shown in Figure 4 , even though it is almost impossible (on the scale presented) to distinguish this graph from that of Figure 3 .

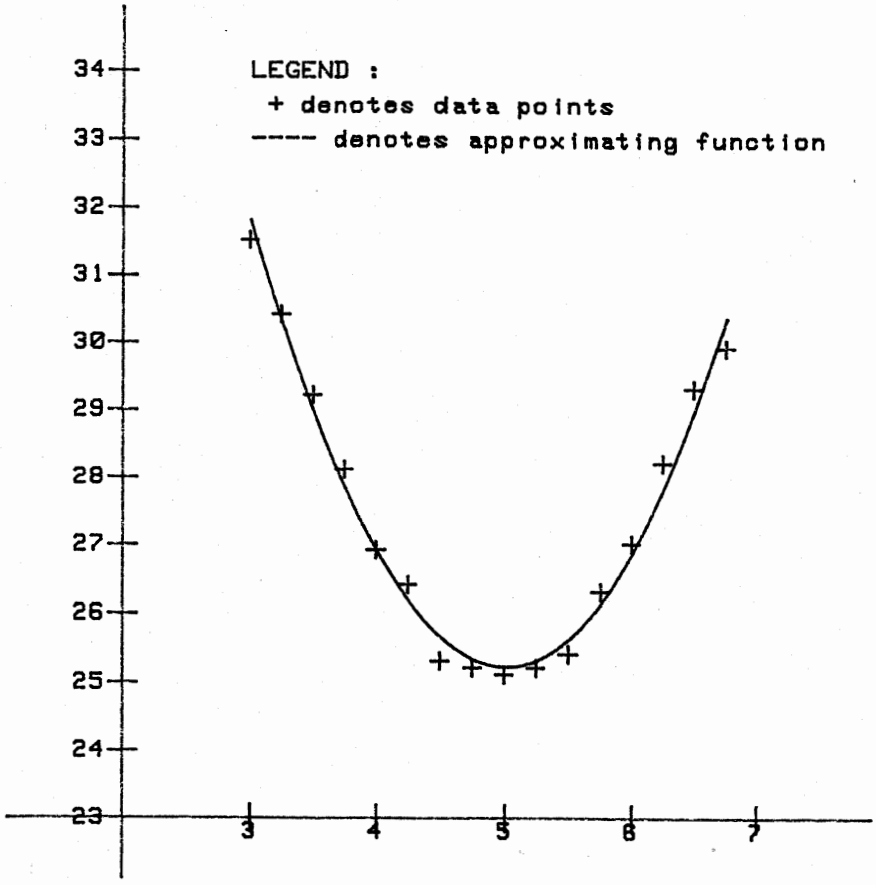


Figure 4 : A CUBIC LEAST-SQUARES FIT

2.5 - THE METHOD OF LEAST-SQUARES FOR OTHER CLASSES OF FUNCTIONS

As indicated in the previous sections, the method of least-squares is always based on an initial ASSUMPTION concerning the type of function to be used to replicate the given data. The purpose of the present section is to demonstrate that this initial assumption need not be restricted solely to polynomial functions. In particular, we shall demonstrate procedures to be used when the method of least-squares is applied to the class of exponential functions or to the class of power functions. Additional examples, in which the method of least-squares is applied to other classes of functions, will be found in the assignments associated with these notes.

2.5.1 - A METHOD OF LEAST-SQUARES FOR EXPONENTIAL FUNCTIONS

Consider the data of Table 7 , shown graphically in Figure 5 :

x	y
0.5	140
1.0	180
1.5	230
2.0	290
2.5	365
3.0	455
3.5	565
4.0	670
4.5	785
5.0	1000
5.5	1230

TABLE 7 : GIVEN DATA

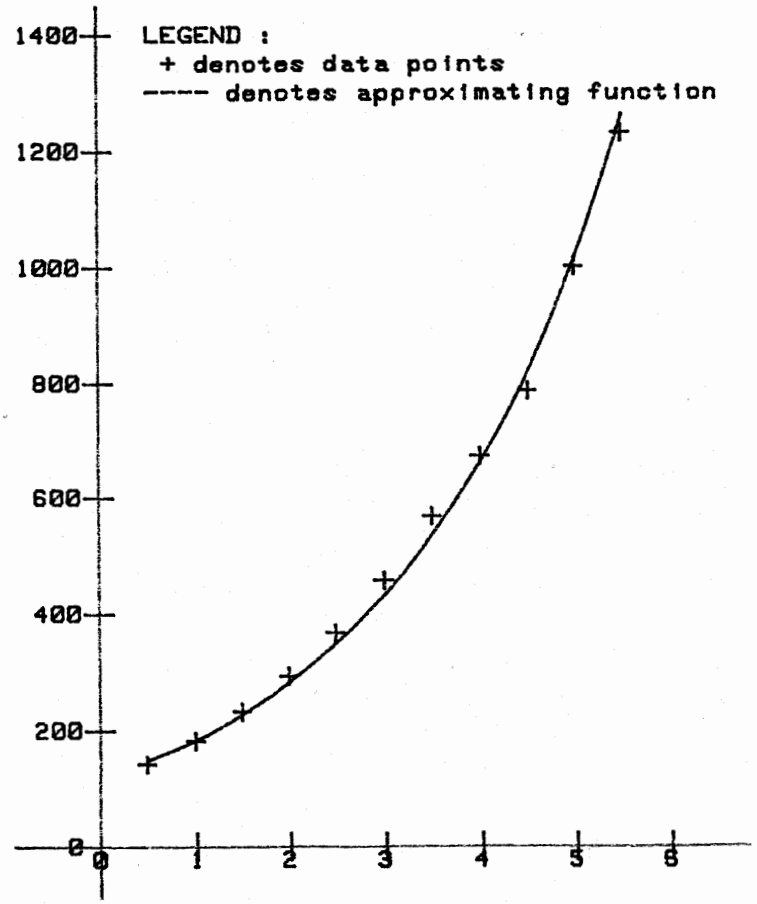


Figure 5 : AN EXPONENTIAL LEAST-SQUARES FIT

The immediate question that arises concerns the class of functions which should be used to approximate this data. However, the observations that

(i) for $x > 0$, $y > 0$,

and

(ii) as x increases, y tends to increase at an increasing rate,

suggest that possibly this data may be replicated by an exponential function of the form

$$y(x) = K e^{\ell x}, \quad (K > 0, \ell > 0). \tag{20}$$

In addition, we note that under this assumption, we have

$$\begin{aligned} \ell n y(x) &= \ell n K + \ell x \\ &= k + \ell x, \quad (k = \ell n K, \text{ or equivalently } K = e^{k}) \end{aligned} \tag{21}$$

so that if the data may be approximated by (20) then the graph of $\ell n y$ vs. x should be approximated by a straight line. For the given data a graph of $\ell n y$ vs. x is shown in Figure 6 .

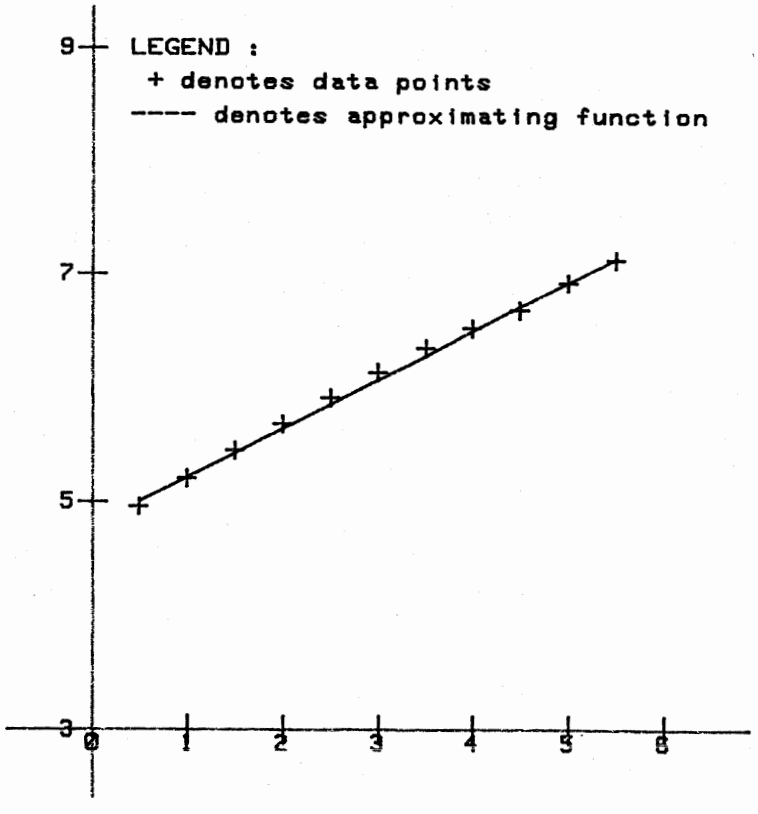


Figure 6 : A GRAPH OF $\ell n y$ vs. x FOR THE DATA OF TABLE 7

The linearity of the graph of ℓny vs. x in Figure 6 is striking and supports the assumption that the given data may be approximated by (20) .

In an effort to determine K and ℓ in (20) , we apply the method of least-squares directly to (20) to obtain the corresponding residual sum of squares

(22) $$S(K,\ell) = \sum_{i=1}^{n} (y(x_i) - y_i)^2$$

$$= \sum_{i=1}^{n} (Ke^{\ell x_i} - y_i)^2$$

with <u>necessary</u> optimality conditions

$$\frac{\partial S}{\partial K} = 2 \sum_{i=1}^{n} (Ke^{\ell x_i} - y_i)\, e^{\ell x_i} = 0$$

and

$$\frac{\partial S}{\partial \ell} = 2 \sum_{i=1}^{n} (Ke^{\ell x_i} - y_i)\, Ke^{\ell x_i} x_i = 0 .$$

The last two conditions are equivalent to

(23) $$(\sum_{i=1}^{n} e^{2\ell x_i})\, K = \sum_{i=1}^{n} y_i e^{\ell x_i}$$

$$(\sum_{i=1}^{n} x_i e^{2\ell x_i})\, K^2 = (\sum_{i=1}^{n} x_i y_i e^{\ell x_i})\, K .$$

Although this is a system of two equations in K and ℓ , it is a **non-linear** system and therefore, at best, is difficult to solve for K and ℓ . Thus, we are inclined to abandon this procedure.

Nevertheless, this does not mean that we must abandon our search for a least-squares exponential estimate (20) for the given data. As an alternative to the above procedure, we simply observe that (21) may be written as

(24) $$Y(x) = k + \ell x , \quad [\text{ with } Y(x) = \ell ny(x)]$$

which is a <u>linear</u> function in x , and consider the (logarithmic) residual sum of squares

(25) $$S_{\ell og}(k,\ell) = \sum_{i=1}^{n} (Y(x_i) - Y_i)^2$$

$$= \sum_{i=1}^{n} (\ell ny(x_i) - \ell ny_i)^2$$

$$= \sum_{i=1}^{n} (\ell x_i + k - \ell ny_i)^2$$

where

(26) $$Y_i = \ell ny_i \qquad (i = 1,2,....,n)$$

denotes the logarithm of the given value y_i appearing in Table 7 , and

(27) $$Y(x_i) = \ell ny(x_i) = k + \ell x_i \qquad (i = 1,2,..., n)$$

denotes the logarithm of the predicted value $y(x_i)$ of y_i corresponding to the value x_i .

Moreover, using the method of least-squares for linear functions as developed previously, we may conclude that it is necessary that k and ℓ satisfy the system of linear equations (6), which in the present notation provide

(28) $$(\sum_{i=1}^{n} x_i^2)\, \ell + (\sum_{i=1}^{n} x_i)\, k = \sum_{i=1}^{n} x_i Y_i$$

$$= \sum_{i=1}^{n} x_i\, \ell ny_i$$

$$(\sum_{i=1}^{n} x_i)\, \ell + n\, k = \sum_{i=1}^{n} Y_i$$

$$= \sum_{i=1}^{n} \ell ny_i \, .$$

In the case of the data of Table 7 , this system becomes

$$126.5\ell + 33k = 212.1448805$$
$$33\ell + 11k = 66.79506617$$

and possesses the unique solution

$$\ell = .4276247996$$

$$k = 4.789404344 \, .$$

Thus, with the use of this logarithmic-based method of least-squares, we find that the best-fitting exponential function for the data of Table 7 is given by

(29) $$y(x) = K\, e^{\ell x}$$

$$= e^{k}\, e^{\ell x}$$

$$= 120.2297318\; e^{\,.4276247996x}\,.$$

In Table 8 below, a comparison is shown between the given data of Table 7 and the corresponding predicted data, calculated according to (29) :

x_i	$y(x_i)$ [calculated, approximate]	y_i
0.5	148.89	140
1.0	184.39	180
1.5	228.34	230
2.0	282.78	290
2.5	350.19	365
3.0	433.67	455
3.5	537.05	565
4.0	665.08	670
4.5	832.63	785
5.0	1019.97	1000
5.5	1263.12	1230

TABLE 8 : APPROXIMATE CALCULATED VALUES, USING THE LOGARITHMIC-BASED LEAST-SQUARES METHOD FOR AN EXPONENTIAL APPROXIMATING FUNCTION, FOR THE DATA OF TABLE 7.

Observations and Remarks:

1. Although the predicted values $y(x_i)$ of y_i, exhibited in Table 8, are all too high near the ends of the range $x \in [0.5, 5.5]$, and are all too low in the centre of this range, there is no exponential function of the form (20) which will provide a better fit to the data of Table 7 than that exhibited in (29), provided the least-squares technique used is that based on the logarithmic residual sum of squares (25). Indeed, it can be shown that in this case (25) provides the value

$$S_{\ell og}(k,\ell) = .0151144591 .$$

2. However, this does not necessarily imply that there is no exponential function which will fit the data better than (29). If, instead of using the logarithmic-based residual sum of squares (25), we had chosen the standard residual sum of squares (22) as our optimization function, we would find that the exponential function (29) provides a (standard) residual sum of squares having value

$$S(K,\ell) = 4620.87533 .$$

   The magnitude of this quantity suggests that it would be possible to find a better-fitting exponential function of the form (20), by requiring that K and ℓ satisfy (23). However, as discussed earlier, solution of this system is much more difficult than that of the system (28), based on the logarithmic residual sum of squares (25), and hence has been abandoned. It should carefully be noted that the two procedures are not identical, and in all likelihood will not result in the same "least-squares approximating exponential function", simply because the optimality functions, and hence optimality criteria, are not the same.

3. In summary, it is therefore very important, whenever a least-squares technique is used, to be very careful to **exhibit clearly the optimality criterion being employed (as well as the class of functions to which it is to be applied)** when drawing conclusions as to the best-fitting replicating function for the given data.

2.5.2 - A METHOD OF LEAST-SQUARES FOR POWER FUNCTIONS

Let us suppose that a set of data (x_i, y_i) $(i = 1,2,...n)$ is given and furthermore that a graph of $\ell n y_i$ vs $\ell n x_i$ is very nearly linear, so that the modified data $(\ell n x_i, \ell n y_i)$ may be approximated by a linear function of the form

$$\ell n y_i = a\, \ell n x_i + b.$$

Under these conditions, it is clear, through exponentiation of both sides of this equation, that

$$y_i = e^b e^{a \ell n x_i}$$

$$= B\, e^{a \ell n x_i} \quad (B = e^b, \text{ or equivalently } b = \ell n B)$$

$$= B\, x_i^a ,$$

from which we may conclude that the given data may be approximated by a power function of the form

(30) $$y(x) = B\, x^a \quad (B > 0),$$

for a and B unknown constants, whose values must yet be determined.

As in the previous case, a direct least-squares technique based on the standard sum of residual squares

$$S(a,B) = \sum_{i=1}^{n} (y(x_i) - y_i)^2$$

$$= \sum_{i=1}^{n} (B\, x_i^a - y_i)^2$$

provides a non-linear system of equations to be satisfied by a and B . (The actual form of this system is itself not important; the important point being that the system is non-linear and hence difficult, if not impossible, to solve.)

Thus, this direct least-squares procedure is abandoned in favour of an indirect least-squares technique, in this case based on a (double logarithmic) residual sum of squares

(31) $$S^*_{\ell og}(a,b) = \sum_{i=1}^{n} (Y(X_i) - Y_i)^2$$

$$= \sum_{i=1}^{n} (a\, \ell n x_i + b - \ell n y_i)^2$$

in which

(32) $$X_i = \ell n x_i$$

(33) $$Y_i = \ell n y_i$$

(34) $$X = \ell n x$$

and

(35) $$Y(X) = \ell n y(x)$$

$$= a\,\ell nx + b$$

$$= a\,X + b\,.$$

Then, under the assumption that the linear function (35) will provide a suitable approximating function for the data $(X_i, Y_i) = (\ell nx_i, \ell ny_i)$ $(i = 1,2,..,n)$, with optimality being attained when (31) has a minimum value, the method of least-squares provides the system of linear equations

$$(\sum_{i=1}^{n} X_i^2)\,a + (\sum_{i=1}^{n} X_i)\,b = \sum_{i=1}^{n} X_i Y_i$$

$$(\sum_{i=1}^{n} X_i)\,a + n\,b = \sum_{i=1}^{n} Y_i\ ,$$

which must be satisfied by a and b. This system may be rewritten in terms of the original data (x_i, y_i), by virtue of (32) and (33), as

(36)
$$(\sum_{i=1}^{n} (\ell nx_i)^2)\,a + (\sum_{i=1}^{n} \ell nx_i)\,b = \sum_{i=1}^{n} ((\ell nx_i)(\ell ny_i))$$

$$(\sum_{i=1}^{n} \ell nx_i)\,a + n\,b = \sum_{i=1}^{n} \ell ny_i\ .$$

Solving the latter system for a and b, noting that

$$B = e^b\ ,$$

provides a least-squares approximating power function of the form (30) for the given data, relative to the least-squares logarithmic residual sum of squares (31).

2.6 - FINITE DIFFERENCES AND THEIR APPLICATION TO THE METHOD OF LEAST SQUARES

As indicated at the end of section 2.4, in using the method of least-squares to fit a curve to some given data (x_i, y_i) , $i = 1,2,...,n$, two problems commonly occur:

1. can the given data be approximated by means of a polynomial

$$y(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_1 x + a_0$$

of degree k ?

and

2. if so, what is the approximate choice for the degree k of this polynomial?

It should be noted that both of these questions must be answered before implementing the method of least-squares, because the precise form of the function to be used to replicate the data is required as input (as an assumption) for the technique. The method of least-squares can only provide the particular function of the assumed class of functions which best fits the given set of data.

To assist us in answering these questions, it is often convenient to make use of finite differences, as described below: Let x_i , for $i = 1,2,..,n$, (n a positive integer), be n equally spaced values of the independent variable x , arranged in ascending order, so that we may write

(37) $$x_i = x_{i-1} + \Delta x \quad (i = 2,3,...,n)$$

where

$$\Delta x = x_i - x_{i-1} \quad (i = 1,2,...,n)$$

is the common (constant) difference interval between successive values of x . For a given function $y = y(x)$, let

(38) $$y_i = y(x_i) .$$

Then, the first (backward) difference of y_i ($i = 1,2,...,n$), denoted by Δy_i , is defined by

(39) $$\Delta y_i = y_i - y_{i-1} ,$$

with higher order (backward) differences ($\Delta^2 y_i$, $\Delta^3 y_i$, . . .) being defined recursively by

(40) $$\Delta^{\ell} y_i = \Delta(\Delta^{\ell-1} y_i), \quad \ell = 2,3....$$

Remarks:

1. The operator Δ, as defined by (39), is a linear operator, in the sense that if $y(x)$ and $y^*(x)$ are any two functions of x, with y_i and y^*_i defined by (38), and c is any real number, then

$$\Delta(y_i + y^*_i) = (y_i + y^*_i) - (y_{i-1} + y^*_{i-1})$$

$$= (y_i - y_{i-1}) + (y^*_i - y^*_{i-1})$$

$$= \Delta y_i + \Delta y^*_i$$

and

$$\Delta(cy_i) = cy_i - cy_{i-1}$$

$$= c(y_i - y_{i-1})$$

$$= c\Delta y_i .$$

2. Thus, each of the operators Δ^2, Δ^3, . . . is also a linear operator and the effect of Δ^{ℓ} on a tabulated function $\{(x_i, y_i) \mid i = 1,2,...,n\}$ may be written as a linear combination of the values y_j ($j = 1,2,...,n$) of the dependent variable. For example:

$$\Delta^2 y_i = \Delta(\Delta y_i)$$

$$= \Delta(y_i - y_{i-1})$$

$$= \Delta y_i - \Delta y_{i-1}$$

$$= (y_i - y_{i-1}) - (y_{i-1} - y_{i-2})$$

$$= y_i - 2y_{i-1} + y_{i-2}$$

$$\Delta^3 y_i = \Delta(\Delta^2 y_i)$$

$$= \Delta(y_i - 2y_{i-1} + y_{i-2})$$

$$= \Delta y_i - 2\Delta y_{i-1} + \Delta y_{i-2}$$

$$= (y_i - y_{i-1}) - 2\,(y_{i-1} - y_{i-2}) + (y_{i-2} - y_{i-3})$$

$$= y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}$$

etc.

Throughout the remainder of this section we shall be primarily concerned with finite differences for polynomial functions of the form

(4l) $$y(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_1 x + a_0$$

where k is a positive integer, and a_k, a_{k-1}, . . . , a_1, a_0 are real numbers.

Examples:

(a) If $y(x) = a$ (constant): $y_i = a$

$$\Delta y_i = y_i - y_{i-1}$$

$$= a - a$$

$$= 0 \ .$$

(b) If $y(x) = x$: $y_i = x_i$

$$\Delta y_i = \Delta x_i = x_i - x_{i-1}$$

$$= \Delta x \quad \text{(constant)}$$

$$\Delta^2 y_i = \Delta^2 x_i = \Delta(\Delta x) = 0 \ .$$

(c) If $y(x) = x^2$: $y_i = x_i^2$

$$\Delta y_i = \Delta x_i^2 = x_i^2 - x_{i-1}^2$$

$$= (x_i + x_{i-1})(x_i - x_{i-1})$$

$$= (x_i + x_{i-1})\Delta x$$

$$\Delta^2 y_i = \Delta^2 x_i^2 = \Delta((x_i + x_{i-1})\Delta x)$$

$$= \Delta(x_i + x_{i-1})\Delta x$$

$$= (\Delta x_i + \Delta x_{i-1})\Delta x$$

$$= (\Delta x + \Delta x)\Delta x$$

$$= 2(\Delta x)^2 \quad \text{(constant)}$$

$$\Delta^3 y_i = \Delta^3 x_i^2 = \Delta(2(\Delta x)^2) = 0 \ .$$

(d) If $y(x) = x^3$: $y_i = x_i^3$

$$\Delta y_i = \Delta x_i^3 = x_i^3 - x_{i-1}^3$$

$$= (x_i^2 + x_i x_{i-1} + x_{i-1}^2)(x_i - x_{i-1})$$

$$= (x_i^2 + x_i x_{i-1} + x_{i-1}^2)\Delta x$$

$$\Delta^2 y_i = \Delta^2 x_i^3 = \Delta((x_i^2 + x_i x_{i-1} + x_{i-1}^2)\Delta x)$$

$$= \Delta(x_i^2 + x_i x_{i-1} + x_{i-1}^2)\Delta x$$

$$= ((x_i + x_{i-1})\Delta x + (x_i x_{i-1} - x_{i-1} x_{i-2}) + (x_{i-1} + x_{i-2})\Delta x)\Delta x$$

$$= ((x_i + 2x_{i-1} + x_{i-2})\Delta x + x_{i-1}(x_i - x_{i-2}))\Delta x$$

$$= ((x_i + 2x_{i-1} + x_{i-2})\Delta x + x_{i-1}(2\Delta x))\Delta x$$

$$= (x_i + 4x_{i-1} + x_{i-2})(\Delta x)^2$$

$$\Delta^3 y_i = \Delta^3 x_i^3 = \Delta((x_i + 4x_{i-1} + x_{i-2})(\Delta x)^2)$$

$$= \Delta(x_i + 4x_{i-1} + x_{i-2})(\Delta x)^2$$

$$= (\Delta x_i + 4\Delta x_{i-1} + \Delta x_{i-2})(\Delta x)^2$$

$$= (\Delta x + 4\Delta x + \Delta x)(\Delta x)^2$$

$$= 6(\Delta x)^3 \quad \text{(constant)}$$

$$\Delta^4 y_i = \Delta^4 x_i^3 = 0 \ .$$

<u>By induction</u>, we may establish that <u>if</u> $y = x^p$ (p a positive integer) , so that $y_i = x_i^p$ <u>then</u>

$$\Delta^p y_i = p!(\Delta x)^p \qquad \text{(constant)}$$

while

$$\Delta^{\ell} y_i = 0 \text{ , for } \ell > p \text{ .}$$

i.e., for p a positive integer:

$$\mathbf{\Delta^p x_i^p = p!(\Delta x)^p \qquad (constant)}$$

with

$$\mathbf{\Delta^{\ell} x_i^p = 0 \text{ , for } \ell > p \text{ .}}$$

More generally, since Δ^{ℓ} (ℓ a positive integer) is a linear operator, when y(x) is the polynomial (4l) of degree k, so that

(42) $$y_i = a_k x_i^k + a_{k-1} x_i^{k-1} + \ldots + a_1 x_i + a_0 \text{ ,}$$

then

(43) $$\Delta^k y_i = a_k \Delta^k x_i^k + a_{k-1} \Delta^k x_i^{k-1} + \ldots + a_1 \Delta^k x_i + a_0 \Delta^k 1$$

$$= a_k k!(\Delta x)^k \qquad \text{(constant)}$$

and

(44) $$\Delta^{\ell} y_i = 0 \text{ , for } \ell > k \text{ .}$$

The finite differences associated with a given function $y = y(x)$ are commonly written in the form of a difference table in which x_i and y_i are displayed in the first two columns, and the successive differences Δy_i, $\Delta^2 y_i$, . . . are displayed in the subsequent columns. As an illustration, the complete difference table for a tabulated function consisting of n data points $\{(x_i, y_i) \mid i = 1,2,..,n\}$ is shown in Figure 7 on the following page.

x	y	Δy	$\Delta^2 y$	. . .	$\Delta^r y$	$\Delta^{r+1} y$	. . .	$\Delta^{n-1} y$
x_1	y_1							
		Δy_2						
x_2	y_2		$\Delta^2 y_3$					
		Δy_3						
x_3	y_3		$\Delta^2 y_4$					
		Δy_4						
x_4	y_4							
					$\Delta^r y_{r+1}$			
						$\Delta^{r+1} y_{r+2}$		
					$\Delta^r y_{r+2}$			
x_{r-1}	y_{r-1}							
		Δy_r						
x_r	y_r		$\Delta^2 y_{r+1}$					
		Δy_{r+1}						$\Delta^{n-1} y_n$
x_{r+1}	y_{r+1}		$\Delta^2 y_{r+2}$					
		Δy_{r+2}						
x_{r+2}	y_{r+2}							
					$\Delta^r y_{n-1}$			
						$\Delta^{r+1} y_n$		
					$\Delta^r y_n$			
x_{n-3}	y_{n-3}							
		Δy_{n-2}						
x_{n-2}	y_{n-2}		$\Delta^2 y_{n-1}$					
		Δy_{n-1}						
x_{n-1}	y_{n-1}		$\Delta^2 y_n$					
		Δy_n						
x_n	y_n							

Figure 7: A COMPLETE DIFFERENCE TABLE FOR A DATA SET CONSISTING OF n POINTS $\{(x_i, y_i) \mid i = 1,2,...,n\}$

Observations concerning Figure 7 :

1. When the given data set consists of n data points (x_i, y_i) there are

(n-1)	first (backward) differences
(n-2)	second (backward) differences
.	
.	
.	
1	$(n-1)^{st}$ (backward) difference .

2. Since there is merely a single $(n-1)^{st}$ difference, namely $\Delta^{n-1}y_n$, the column corresponding to $\Delta^{n-1}y$ may be regarded as being a constant column, so that if we were to construct the column corresponding to $\Delta^n y$, this column would logically be a constant column consisting entirely of zeroes.

3. The above observation, with equations (42) and (43) being noted, suggests that the given set of data (x_i, y_i) for $i = 1,2,...,n$, may be represented by means of a polynomial of degree (n-1) . Indeed, this observation is in complete agreement with the results of section 2.2.

4. More specifically, it is observed that, if the column corresponding to $\Delta^r y$ $(r < n-1)$ consists of a constant string,

   i.e., $$\Delta^r y_{r+1} = \Delta^r y_{r+2} = \ldots = \Delta^r y_n ,$$

   so that the subsequent columns consist entirely of zeroes,

   i.e., $$\Delta^{\ell} y_i = 0 \quad \text{for } i = \ell+1, \ldots, n ; \ \ell = r+1, \ldots, n-1$$

   then equations (42) and (43) suggest that the tabulated function may be represented by a polynomial of degree r .

Example: The tabulated data appearing in the first two columns of the difference table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$	...
1	7.93						
		2.12					
2	10.05		.49				
		2.61		.03			
3	12.66		.52		0		
		3.13		.03		0	
4	15.79		.55		0		
		3.68		.03		0	
5	19.47		.58		0		
		4.26		.03		0	
6	23.73		.61		0		
		4.87		.03			
7	28.60		.64				
		5.51					
8	34.11						

consists of eight data points (x_i, y_i) $i = 1,2,...,8$ and hence can be reproduced exactly by a polynomial of degree 7 using the techniques of section 2.2. However, since the column of third order differences $\Delta^3 y$ is constant, the above observation suggests that this data can also be represented precisely by a cubic polynomial (degree 3) of the form

(45) $$y(x) = ax^3 + bx^2 + cx + d \ ,$$

at least in the range of values $1 \le x \le 8$ exhibited in the table. Indeed, if this suggestion is correct, there are various ways in which the coefficients a , b , c and d of this polynomial may be evaluated as suggested below:

Method 1: Since the suggestion is that the cubic function (45) reproduces this set of data precisely, we may chose any four of the given data points and find a , b , c and d so that (45) passes through the four chosen points. In particular, if we choose the first four points, we obtain the system of four linear equations

$$\begin{aligned} a + b + c + d &= 7.93 \\ 8a + 4b + 2c + d &= 10.05 \\ 27a + 9b + 3c + d &= 12.66 \\ 64a + 16b + 4c + d &= 15.79 \end{aligned}$$

which possesses the (unique) solution

(46) $$a = .005 \ , \ b = .215 \ , \ c = 1.44 \ , \ d = 6.27 \ .$$

Thus the desired cubic polynomial is simply

(47) $$y(x) = .005\,x^3 + .215\,x^2 + 1.44x + 6.27\ .$$

As suspected, when this polynomial is evaluated at $x = 1,2,\ldots,8$ we find the table of values

x	y
1	7.93
2	10.05
3	12.66
4	15.79
5	19.47
6	23.73
7	28.60
8	34.11

from which it is clear that (47) provides an exact reproduction of the given data.

Method 2: As an alternative to the above procedure, we may again assume that the first four points (or any four chosen points) must lie on the cubic polynomial (45) and use the Lagrange Interpolation formula (3) (with $n = 4$) to obtain the values of a , b , c and d . Not surprisingly, this procedure also provides (47) and hence provides an exact reproduction of the given data.

Method 3: As a second alternative, let us assume merely that the above set of data (x_i, y_i) for $i = 1,2,\ldots,8$ may be approximated by a cubic polynomial of the form (45) and apply the method of least-squares for this particular class of functions. Under this assumption, we have previously shown that it is necessary for a , b , c and d to be solutions of the system of linear equations (17) , where by direct calculation it may be shown that

$$\sum_{i=1}^{n} x_i^6 = 446964$$

$$\sum_{i=1}^{n} x_i^5 = 61776$$

$$\sum_{i=1}^{n} x_i^4 = 8772$$

$$\sum_{i=1}^{n} x_i^3 = 1296$$

$$\sum_{i=1}^{n} x_i^2 = 204$$

$$\sum_{i=1}^{n} x_i = 36$$

$$n = 8$$

$$\sum_{i=1}^{n} x_i^3 y_i = 36274.26$$

$$\sum_{i=1}^{n} x_i^2 y_i = 5340.18$$

$$\sum_{i=1}^{n} x_i y_i = 841.98$$

$$\sum_{i=1}^{n} y_i = 152.34 \ .$$

When the resulting system (17) is solved, using a Texas Instruments TI-59 calculator, the unique solution [to 10 significant figures] is

(48)
$$\begin{aligned} a &= .0049999996 \\ b &= .2150000061 \\ c &= 1.439999977 \\ d &= 6.270000023 \ . \end{aligned}$$

It is immediately observed that (48) are not precisely the same as (46) ; however when (48) are rounded to 3 significant figures we again obtain (46) and hence (47) . Apparently, these (relatively small) discrepancies in the values of a , b , c and d occur as the result of round-off errors accumulating in the internal program in the TI-59 calculator.

The above procedure may be extended to the case when the column in the difference table corresponding to $\Delta^r y$ ($r < n-1$) is "relatively constant" by noting the effect of small errors in the recorded values y_i of a tabulated function.

For example, suppose that the recorded value $y_5 = 19.47$ (corresponding to $x_5 = 5$) appearing in the preceding table is subjected to a small recording error ε and hence is recorded as $19.47 + \varepsilon$ rather than the "actual value" 19.47 . [Normally we would not know the value of ε , but have in this case shown ε explicitly merely for the purposes of demonstration]. With this small error introduced into the data, the correspondilng difference table becomes

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$	...
1	7.93						
		2.12					
2	10.05		.49				
		2.61		.03			
3	12.66		.52		ε		
		3.13		$.03+\varepsilon$		-5ε	
4	15.79		$.55+\varepsilon$		-4ε		
		$3.68+\varepsilon$		$.03-3\varepsilon$		10ε	
5	$19.47+\varepsilon$		$.58-2\varepsilon$		6ε		
		$4.26-\varepsilon$		$.03+3\varepsilon$		-10ε	
6	23.73		$.61+\varepsilon$		-4ε		
		4.87		$.03-\varepsilon$			
7	28.60		.64				
		5.51					
8	34.11						

From this difference table, it can be seen that not only is the small recording error ε propogated through the successive columns of the table, but is moreover magnified as we proceed from one column to the next.

For example a small recording error, say $\varepsilon = .01$, which is very small in relation to the "actual value" 19.47 of the entry y_5 , gives rise to errors in the third order differences which can be as large as the "actual third order differences" themselves. Indeed in this case, rather than having higher order differences which are all zero, we obtain successive columns of differences which are diverging from the "actual" values (which are in this case all 0).

This situation is further compounded when more than one recorded value of y_i is subject to a "recording error". Nonetheless, it is not unreasonable to expect that even with such small recording errors present, the above set of data could reasonably be approximated by a cubic polynomial of the form (45) with coefficients being very close to the values (46) . However, since it is no longer acceptable to require that the given data points lie on (46) , we are faced with the problem of finding the best-fitting cubic polynomial which approximates the given (error-containing) data. This, of course, involves the application of the method of least-squares as demonstrated in method 3 of the previous example.

The above discussion suggests the following:

For a given set of n data points (x_i, y_i) $(i = 1,2,...,n)$, **if the ranges of values of the successive columns of differences** $\Delta^{\ell}y$ $(\ell = 1, 2,...,k)$ **decrease "fairly rapidly" and "approach zero" and if the column** $\Delta^{k}y$ **is "relatively constant", then the given set of data may be approximated (at least within the range of values of the independent variable** x) **"reasonably well" by a polynomial of degree k of the form (4l)** .

Once the above suggestion has been used as a guide in choosing the appropriate degree of the class of polynomials to be used as the initial assumption, then the method of least-squares may be applied to determine the most appropriate choices of the coefficients for this polynomial to provide the best-fitting polynomial (of this assumed degree) which approximates the given data.

Illustrative Example: Consider the data of Table 4 of section 2.4 . For this data we construct the difference table shown below:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
3.00	31.5			
		-1.1		
3.25	30.4		-.1	
		-1.2		.2
3.50	29.2		.1	
		-1.1		-.2
3.75	28.1		-.1	
		-1.2		.8
4.00	26.9		.7	
		-.5		-1.3
4.25	26.4		-.6	
		-1.1		1.6
4.50	25.3		1.0	
		-.1		-1.0
4.75	25.2		0	
		-.1		.2
5.00	25.1		.2	
		.1		-.1
5.25	25.2		.1	
		.2		.6
5.50	25.4		.7	
		.9		-.9
5.75	26.3		-.2	
		.7		.7
6.00	27.0		.5	
		1.2		-.6
6.25	28.2		-.1	
		1.1		.6
6.50	29.3		.5	
		.6		
6.75	29.9			

It is observed that the ranges of values for the successive differences Δy , $\Delta^2 y$, $\Delta^3 y$ are given by the inequalities

$$-1.2 \leq \Delta y \leq 1.2$$

$$-0.6 \leq \Delta^2 y \leq 1.0$$

$$-1.3 \leq \Delta^3 y \leq 1.6 \ .$$

Since the ranges of values of the successive differences $\Delta^{\ell}y$ decrease "quite rapidly toward zero", at least as ℓ increases from 1 to 2 (although they subsequently increase as ℓ increases further) and since $\Delta^2 y$ is "relatively constant", we are lead to consider the suggestion that this set of data may be approximated by a quadratic polynomial (i.e., degree 2).

In particular, we have previously applied the method of least-squares, with the above assumption, to this data to obtain the best-fitting least-squares quadratic function for this data, namely

$$y(x) = 1.659943975x^2 - 16.58915964x + 66.68043412$$

with the residual sum of squares

$$S = 1.016854342 \ .$$

As indicated earlier, this small value of S indicates that this quadratic function provides a "reasonable" approximating function for the given data.

CHAPTER 3 - SOME SIMPLE DIFFERENTIAL EQUATIONS COMMONLY OCCURING IN THE CONSTRUCTION OF MATHEMATICAL MODELS

3.1 - INTRODUCTION

In this short chapter we shall briefly discuss some results from the mathematical theory of differential equations which commonly appear in the development of simple mathematical models of real-world phenomena. Although the material presented is intended to be a review for most students in this course, it is acknowledged that for some students this material may be new. Thus, greater space has been allocated to this discussion than may be necessary, in order to present those features of most importance to our intended use.

3.2 - THE GENERAL FIRST ORDER DIFFERENTIAL EQUATION

The general first order differential equation, having (unknown) dependent variable y and independent variable x, may be written (not necessarily in a unique way) in the general form

$$\frac{dy}{dx} = f(x,y), \tag{1}$$

in which $f(x,y)$ is some given (hence known) function of the variables x and y.

The possible non-uniqueness of the above form for a first order differential equation may be demonstrated through the following example.

Example: The first order differential equation

$$(y')^2 + xy' - y = 0$$

(where ' denotes differentiation with respect to x) may be regarded as being a quadratic equation in y', with coefficients being functions of x and y, and may thus be rewritten, by means of the quadratic formula, in either of the following two forms:

$$y' = \frac{(-x + (x^2 + 4y)^{1/2})}{2}$$

or

$$y' = \frac{(-x - (x^2 + 4y)^{1/2})}{2}$$

each of which is of the form (1) . Thus, there are two possible choices for the function f(x,y) in (1) in this case, and the given differential equation must be regarded as being equivalent to both of the latter two equations.

In addition, a function $y = y(x)$, for x in some interval I , is said to be a solution of (1) on I if $y'(x)$ exists throughout I , and if

$$y'(x) = f(x,y(x)) \quad \text{for all } x \in I ,$$

i.e., if substitution of $y = y(x)$ into (1) produces an identity in x on I .

The mathematical theory of differential equations is concerned with the following three questions:

1. THE PROBLEM OF EXISTENCE:
   Does a given differential equation possess a solution?
2. THE PROBLEM OF DETERMINATION:
   If so, how do we find such a solution?
3. THE PROBLEM OF UNIQUENESS?
   If a solution exists, is it unique?

The first and last of these questions are theoretical in nature, although they do have some very practical aspects associated with them, while the second question is itself largely a practical matter. Below is given a basic existence-uniqueness theorem for initial-value problems associated with the differential equation (1) , and the remainder of this chapter is concerned with the problem of determination of solutions of first order differential equations and their associated initial-value problems.

3.3 - A BASIC EXISTENCE-UNIQUENESS THEOREM

The following well-known theorem provides sufficient conditions for the first-order initial-value problem

$$(2) \qquad \frac{dy}{dx} = f(x,y) , \text{ subject to } y_0 = y(x_0) ,$$

to possess a unique local solution (i.e., in some neighbourhood of the point (x_0, y_0) in the xy-plane) .

THEOREM: Let a and b be positive real numbers. Suppose that $f(x,y)$ and $\frac{\partial f}{\partial y}$ are continuous on some rectangle R , of width $2a$ and height $2b$, centred at (x_0, y_0) ,

i.e., $$R : |x - x_0| \le a , \; |y - y_0| \le b .$$

Then, there exists a positive real number h $(0 < h \le a)$ and a differentiable function $y = y(x)$ defined on the interval I (on the x-axis) of width $2h$, centred at x_0 ,

i.e., $\mathbf{I: \ |x - x_0| \le h \ , \ (0 < h \le a)}$

satisfying the following four conditions:

(i) $\mathbf{y = y(x)}$ **is a solution of (1) on I ,**

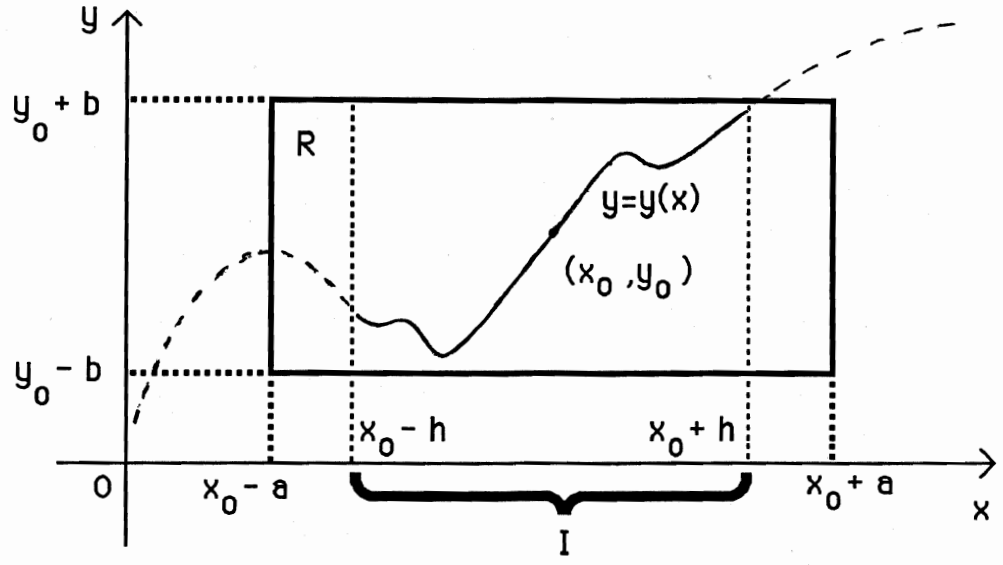
(ii) $\mathbf{y = y(x)}$ **passes through** $\mathbf{(x_0, y_0)}$ **,**

(iii) $\mathbf{y = y(x)}$ **satisfies the inequality** $\mathbf{|y(x) - y_0| \le b}$ **on I ,**

and

(iv) $\mathbf{y = y(x)}$ **is <u>unique</u> on I , in the sense that it is the only function satisfying conditions (i) - (iii) .**

Graphically we may represent the above properties of the solution of this initial-value problem as follows:



Remarks:

1. As indicated by the above theorem statement, the interval I upon which the solution $y = y(x)$ exists and is unique may be, but need not be, the whole of the interval $|x - x_0| \le a$ appearing in the definition of the rectangle R .

2. The above theorem provides <u>sufficient</u> conditions for the existence and uniqueness of a solution of the given initial-value problem (2) . These conditions are <u>not necessary</u> conditions, so that even when they are not valid it is possible for the given initial-value problem to possess a unique solution, although this is in general not the case. There are more general existence-uniqueness theorems which provide weaker

(sufficient) conditions than those exhibited above. However, we shall not discuss these results.

3.4 - SOME TECHNIQUES FOR SOLVING SIMPLE FIRST ORDER DIFFERENTIAL EQUATIONS

There are a number of techniques available for solving first order differential equations, each of which is dependent on some particular property of the function f(x,y) appearing on the right-hand side of (1) . Below we shall discuss two such techniques and exhibit their application to specific examples.

3.4.1 - SEPARABLE DIFFERENTIAL EQUATIONS

If f(x,y) may be factored into a product of two functions, each of which is a function only of one of the variables x or y , then (1) may be rewritten in the form

$$A(x)\,dx + B(y)\,dy = 0\,, \tag{3}$$

in which the variables have been separated into distinct terms. In this case we say that the differential equation is a separable differential equation.

In an effort to solve (3) , we note that if $y = y(x)$ is a solution of (3) on some interval I , then

$$A(x)\,dx + B(y(x))\,y'(x)\,dx = 0 \quad \text{for all } x \text{ in } I\,.$$

Since the latter result is an identity on I , we may integrate it on I to obtain

$$\int A(x)\,dx + \int B(y(x))\,y'(x)\,dx = c\,.$$

However, since the integration in the second term depends upon the desired unknown function $y = y(x)$ on I , it is impossible to perform this integration explicitly. To circumvent this difficulty, we observe that the unknown function $y = y(x)$ may be regarded as defining an (unspecified) change of variable in this integral, so that we may write that, under this transformation of variables,

$$\int B(y(x))\,y'(x)\,dx = \int B(y)\,dy\,.$$

Therefore, it appears that the relation

$$\int A(x)\,dx + \int B(y)\,dy = c \tag{4}$$

defines a solution $y = y(x)$ of (3) implicitly on some interval I .

To verify this suspicion, we note that if (4) defines a function $y = y(x)$ on some interval I , then by implicit differentiation with respect to x on I ,

$$A(x)\,dx + B(y(x))\,y'(x)\,dx = 0 \quad \text{for all } x \text{ in } I\,,$$

so that this implicitly-defined function is indeed a solution of (3) on I .

EXAMPLE: Solve: $\frac{dy}{dx} = \frac{2xy}{1-x^2}$ $(x \neq \pm 1)$

Separate variables: $\frac{1}{y}dy = \frac{2x}{1-x^2}\,dx$ $(y \neq 0, x \neq \pm 1)$

Integrate: $\ln|y| = -\ln|1-x^2| + c$

i.e., $|y| = \frac{K}{|1-x^2|}$ $(K > 0)$

i.e., $y = \frac{L}{(1-x^2)}$ $(L \neq 0)$.

[OBSERVATION: as is the usual case, it is easily verified that the above function is indeed a solution of the given differential equation, on any of the three intervals $x < -1$ or $-1 < x < 1$ or $1 < x$. Indeed, if

$$y = \frac{L}{(1-x^2)}$$

then

$$y' = \frac{(-1)L(-2x)}{(1-x^2)^2}$$

$$= \frac{2xL}{(1-x^2)^2}$$

$$= \frac{2xy}{(1-x^2)}$$ as required .]

3.4.2 - LINEAR FIRST ORDER DIFFERENTIAL EQUATIONS

When f(x,y) is a linear function in y [with coefficients being any given functions of x] of the form

$$f(x,y) = Q(x) - P(x)\, y ,$$

(1) may be written in the form

$$(5) \qquad \frac{dy}{dx} + P(x)\, y = Q(x) ,$$

which is then said to be a linear first order differential equation.

In order to solve (5), suppose there exists a non-zero function $\mu = \mu(x)$, defined on some interval I, satisfying the condition that

$$(6) \qquad \mu(x)\frac{dy}{dx} + \mu(x)\, P(x)\, y = \frac{d\,[\mu(x)\, y]}{dx} \quad \text{on } I ,$$

so that, upon multiplication by $\mu(x)$, (5) becomes

$$(7) \qquad \frac{d\,[\mu(x)\, y]}{dx} = \mu(x)\, Q(x) \quad \text{on } I ,$$

which is now a first order differential equation which is immediately integrable. Thus, we may integrate the last equation on I to obtain

$$\mu(x)\, y = \int \mu(x)\, Q(x)\, dx + c$$

or equivalently, since $\mu(x) \neq 0$ on I,

$$y = \frac{1}{\mu(x)} \int \mu(x)\, Q(x)\, dx + \frac{c}{\mu(x)}$$

Hence, if there exists such a non-zero function $\mu(x)$, upon multiplication of (5) by this function, we obtain an equivalent differential equation of the form (7) which is immediately integrable. Thus, whenever there exists such a non-zero function $\mu(x)$, it is said to be an integrating factor for the first order linear differential equation (5).

To determine whether or not an integrating factor for (5) exists, we return to the requirement that $\mu(x)$ must satisfy condition (6), namely

$$\mu(x)\frac{dy}{dx} + \mu(x)\, P(x)\, y = \frac{d\mu(x)}{dx}\, y + \mu(x)\frac{dy}{dx}$$

or equivalently, upon cancellation of various terms,

$$y\left[\frac{d\mu(x)}{dx} - P(x)\,\mu(x)\right] = 0\,,$$

i.e., since $y = 0$ is <u>not</u> in general a solution of (5) , we must have

$$\frac{d\mu(x)}{dx} = P(x)\,\mu(x) \tag{8}$$

But (8) is a first order <u>separable</u> differential equation for the desired integrating factor $\mu(x)$, since the function $P(x)$ is known. To solve (8) , we separate variables to obtain

$$\frac{1}{\mu(x)}\,d\mu = P(x)\,dx\,,$$

and following the general procedure, we integrate this equation to obtain

$$\ln|\mu| = \int P(x)\,dx + c$$

$$\text{i.e.,}\quad |\mu| = K\,e^{\int P(x)\,dx} \qquad (K > 0)$$

$$\text{i.e.,}\quad \mu(x) = L\,e^{\int P(x)\,dx} \qquad (L \neq 0)\,. \tag{9}$$

Thus, if (5) has an integrating factor, it must be of the form (9) , in which $P(x)$ denotes the coefficient of y in the given differential equation (5) . However, (9) defines an infinite number of possible integrating factors for (5) , and since we only need one such integrating factor, we may choose L to be any appropriate constant other than zero. Typically, L is chosen to be "1" , although any other non-zero value is acceptable.

EXAMPLE: Solve: $x\,\frac{dy}{dx} + 2y = x\sin x$.

Rewrite in standard form:

(10) $$\frac{dy}{dx} + \frac{2}{x}y = \sin x \qquad (x \neq 0)\,.$$

Integrating factor: $\mu(x) = L\, e^{\int P(x)\,dx}$

$$= L\, e^{\int (2/x)\,dx}$$

$$= L\, e^{2\ln|x|}$$

$$= L\, x^2$$

so an appropriate choice for the integrating factor is

$$\mu(x) = x^2 \,.$$

Multiply (10) by x^2 : $x^2\frac{dy}{dx} + 2xy = x^2 \sin x$.

Identify form (7) : $\frac{d[x^2 y]}{dx} = x^2 \sin x$.

Integrate:

(11) $$x^2\, y = \int x^2 \sin x\, dx + c \,.$$

To evaluate the integral on the right-hand side of (11) , we employ the formula for integration by parts, namely

$$\int u\, dv = u\, v - \int v\, du\,,$$

with

$$u = x^2 \,,\ du = 2x\, dx \,,\ dv = \sin x\, dx \,,\ v = -\cos x \,.$$

Thus:

(12) $$\int x^2 \sin x\, dx = -x^2 \cos x - \int (-\cos x)(2x)\, dx$$

$$= -x^2 \cos x + 2\int x \cos x\, dx \,.$$

To evaluate the latter integral, we again employ the formula for integration by parts with

$$u = x \ , \ du = dx \ , \ dv = \cos x \, dx \ , \ v = \sin x$$

to obtain

(13) $$\int x \cos x \, dx = x \sin x - \int \sin x \, dx$$

$$= x \sin x + \cos x \ .$$

Substitute (13) into (12), and then (12) into (11):

$$x^2 \, y = - x^2 \cos x + 2x \sin x + 2 \cos x + c$$

$$= (2 - x^2) \cos x + 2x \sin x + c \ ,$$

i.e.,

(14) $$y(x) = \left(\frac{2}{x^2} - 1 \right) \cos x + \frac{2}{x} \sin x + \frac{c}{x^2} \qquad (x \neq 0)$$

[As was pointed out in the previous example, we may as usual check that the above function is a solution of the given differential equation (10), by direct differentiation of (14).]