PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A Purification Method for Computing the Latent Columns of Numerical Matrices and Some Integrals of Differential Equations

L. F. Richardson

Phil. Trans. R. Soc. Lond. A 1950 **242**, 439-491
doi: 10.1098/rsta.1950.0007

A PURIFICATION METHOD FOR COMPUTING THE LATENT COLUMNS OF NUMERICAL MATRICES AND SOME INTEGRALS OF DIFFERENTIAL EQUATIONS

By L. F. RICHARDSON, F.R.S.

Hillside House, Kilmun, Argyll, Britain

(*Received* 18 *June* 1949—*Revised* 29 *July* 1949)

CONTENTS

The emphasis is on practical arithmetical methods suitable for application to matrices of so many rows and columns that their determinants are intractable. The arithmetic begins from an arbitrary, or guessed, column of numbers. This may be regarded as a raw material composed of the desired constituent plus an unknown assortment of impurities. Although the impurities are not known, their behaviour when subjected to certain computing operations can be foreseen sufficiently to allow them to be cleaned away gradually. These operations depend on all the latent roots. There is an initial stage of groping-about to find roughly where the latent roots are located. This leads on to a final stage of rapid routine in which roughly known latent roots are used to purify the column associated with the desired root. It is shown that the cleaning operation spreads over a considerable region of the Argand diagram; so that very imperfect information is, nevertheless, often effective. The proof of the numerical result is by comparison of independent estimates, and not by theorems about infinitely many operations.

A closely similar purification method was applied in 1910 to a wide class of problems (so-called jury problems) involving differential equations, such as Laplace's, or the equation of plane stress. In the present paper the choice of cleaning agents for such purposes is made more systematic by employing Legendre's principle of least squares.

1. Introduction, notation and easy examples

There are many iterative methods for finding the latent roots of a matrix, and here is another: namely, premultiplication of a suitable column by a matrix formed from the given matrix by subtracting a suitable number from each of the elements in its principal diagonal. I have several reasons for thinking that this method may be worthy of separate publication: (1) It is not mentioned by Hotelling in his valuable reviews of such methods (1943 *a*, *b*), where he gave altogether thirty-eight references. (2) It is not mentioned by Frazer, Duncan

& Collar in a book (1938) where they devote much space to numerical methods. (3) Although Aitken in 1937 alluded to it in his concluding remarks (p. 303), he did so only briefly, and his main emphasis was quite different. (4) I arrived at the present method independently by thinking about the iterations to the solutions of partial differential equations which I had published in 1910. (5) The method gave satisfactory results on unsymmetrical square matrices of nine or ten rows which occurred in the theory of arms-races (Richardson 1939, 1947). (6) It is more powerful than the process of premultiplication by the given matrix, which, according to Holzinger & Harman (1948), was first published in 1933 by Hotelling, which was used by Duncan & Collar (1934) and was described by Hotelling (1943*a*) as fundamental. (7) The present method affords an outlet for the computer's skill and discretion and thus may make his toil less of a dull routine and more like a game. (8) My two previous accounts of the method (1939, 1947) can now be much improved, by alliance with the great science of matrix algebra; a coalescence suggested to me in 1948 by Mr Ardie Lubin.

Let K denote the given matrix of n rows and n columns of real numbers. It is not otherwise restricted, and in particular need not be symmetrical. Let I denote, as usual, the so-called 'unit-matrix' of n rows and n columns which has unity in each place on its principal diagonal and zero elsewhere:

$$K = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots\dots\dots\dots\dots\dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{bmatrix}, \tag{1}$$

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots\dots\dots\dots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{2}$$

The determinantal equation

$$|K-\lambda I| = 0 \tag{3}$$

has n roots which will be denoted by $\lambda_1, \lambda_2, \dots, \lambda_n$. For the present let them be arranged in any order.

Let A denote an arbitrary column of n numbers, real or complex. Let $\mathbf{0}$ denote a column of n zeroes. Then in matrix algebra

$$(K-\lambda_1 I)(K-\lambda_2 I)\dots(K-\lambda_n I)A = \mathbf{0}. \tag{4}$$

to n brackets

This is an immediate consequence of the Cayley-Hamilton theorem, that marvel of generality and simplicity. A fairly easy proof of it is given by Bôcher (1935, pp. 296–297).

In order to explain the computing operations clearly it may be best to mention first, not what the computer actually does, but what he would like to do if he could. Suppose that he knew all the latent roots and that they were distinct from one another, but did not know their latent columns. Then he could compute a column P_3 defined thus, by omitting $(K-\lambda_3 I)$

$$P_3 = (K-\lambda_1 I)(K-\lambda_2 I)(K-\lambda_4 I)\dots(K-\lambda_n I)A. \tag{5}$$

to $(n-1)$ brackets.

From (4) and (5) it follows that $$(K-\lambda_3 I)\,P_3 = \mathbf{0}. \tag{6}$$

So λ_3 can be obtained from any element of KP_3 by dividing it by the corresponding element of P_3. There are n such quotients, and their equality is the proof that P_3 and λ_3 have been correctly computed. The column is called the 'latent column associated with λ_3'. Similarly for any other distinct λ. The peculiarities connected with equal roots will be discussed later.

The sum of all the latent roots is readily available, for

$$\lambda_1+\lambda_2+\ldots+\lambda_n = k_{11}+k_{22}+\ldots+k_{nn}. \tag{7}$$

So the computer can hardly be said to know all the roots but one. He may, however, happen to know all the roots but two, say λ_3 and λ_4. If so, taking an arbitrary column A he can compute

$$B = (K-\lambda_1 I)\,(K-\lambda_2 I)\,(K-\lambda_5 I)\,(K-\lambda_6 I)\ldots(K-\lambda_n I)\,A. \tag{8}$$
to $(n-2)$ brackets

Then the Cayley-Hamilton theorem assures him that

$$(K-\lambda_3 I)\,(K-\lambda_4 I)\,B = \mathbf{0},$$

that is $$K^2B-(\lambda_3+\lambda_4)\,KB+\lambda_3\lambda_4 B = \mathbf{0}. \tag{9}$$

This is a set of n simultaneous equations, and in each $\lambda_3\lambda_4$ is the only unknown. So the computer obtains n independent values of $\lambda_3\lambda_4$, and their agreement is the proof that his work is correct. Then from $\lambda_3\lambda_4$ and the known $\lambda_3+\lambda_4$ the roots λ_3 and λ_4 are obtained separately.

It is only the ratios of the elements of a latent column P_ν that are specified by

$$(K-\lambda_\nu I)\,P_\nu = \mathbf{0}.$$

Various devices for fixing the arbitrary common factor are in use. The $P_1, P_2, \ldots, P_n$ are sometimes called 'latent vectors'; but in view of the objection by Jeffreys & Jeffreys (1946, p. 107) they will here be called 'latent columns'. The latter name is more specific because there are also latent rows. Frazer *et al.* (1938) call the P 'modal columns'. The theory and practice of finding latent rows is the same as for the latent columns, except that the brackets $(K-\lambda I)$ must then postmultiply rows instead of premultiplying columns.

When, for economy of paper, what is really a column is set as a row, it will be enclosed in curly brackets, after the manner of Turnbull & Aitken (1932). Square brackets are reserved for matrices set in their proper orientation.

In order to save time and printing-space the examples are small and simple, but they are intended to illustrate methods which would be applicable to matrices of many rows and columns. Thus the examples are like a game played according to the rule that forbids any operation that would be too troublesome with a matrix of ten rows and ten columns; in particular, the player is forbidden to evaluate a determinant.

It is necessary to have affixes 0, 1, 2, 3, ... for the serial number of successive approximation. As the upper site is occupied by indices of powers and the lower site is occupied by the reference numbers of the latent roots and columns, I propose *to set the iteration number at a middle level*. Thus X^0, X^1, X^2, X^3, ... denotes a sequence of approximations. It is difficult to do this with small letters; and so Aitken's otherwise convenient custom of using small letters for columns has been regretfully abandoned.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Transposition of the rows and columns of a matrix will be indicated by the superscript $\sim$ after the manner of Jeffreys & Jeffreys (1946, p. 108). It is more distinctive than the customary accent.

It is frequently necessary to divide each element of a column KX by the corresponding element of the column X so as to obtain a column of quotients. The result will be written $KX \div X$. The justification for this notation is that the expression cannot be written in standard matrix algebra; and that the division sign $\div$, though still used in school algebra, is hardly ever seen in modern higher mathematics.

To anyone who has purified chemicals, the scheme of operations is suggested by calling A a raw material and P a pure extract from it. In the later stages of computation each of the λ may be roughly known; and, if so, an approximation to the ideal operations (5) and (8) can be performed. This may be called 'fine purification'. The initial stage is much less obvious. It may be described as groping-about in search for rough estimates of roots.

Example (1). The matrix

$$K = \begin{bmatrix} -1 & 1 & 3 \\ 3 & -2 & 2 \\ 1 & 2 & -3 \end{bmatrix}$$

occurs, let us suppose, in a problem about the stability of the relations between three armed nations (Richardson 1939, 1947). The chief question is to find the latent root which has the greatest real part; for if it is positive, there will be a drift towards war. As K is unsymmetrical, it is not obvious whether complex latent roots are present or not. The mean of the latent roots is obviously -2. If they are not all equal, there must be one or two which have real parts less than -2. The computer does not know their values. But as an experiment he decides to premultiply the column A repeatedly by $K+4I$, for that would tend to clean away the associates of any root in the neighbourhood of -4. Many authors take the arbitrary initial column to be simply $\{1, 1, 1\}$. It is easy to think of. On this occasion the computer, having had some experience with matrices rather like the given K, fancies that the latent column for which he is seeking will resemble the sums along the rows of K omitting the principal diagonal, that is, $\{4, 5, 3\}$. If he has guessed wisely, the subsequent purifications will be shortened, and vice versa. Accordingly he computes the sequence

$$K^{s+1} = (K+4I)\,X^s, \quad \text{starting from} \quad X^0 = \{4, 5, 3\}.$$

It goes as follows:

$$X^0 = \begin{bmatrix} 4 \\ 5 \\ 3 \end{bmatrix}, \quad X^1 = \begin{bmatrix} 26 \\ 28 \\ 17 \end{bmatrix}, \quad X^2 = \begin{bmatrix} 157 \\ 168 \\ 99 \end{bmatrix}, \quad X^3 = \begin{bmatrix} 936 \\ 1005 \\ 592 \end{bmatrix}, \quad X^4 = \begin{bmatrix} 5589 \\ 6002 \\ 3538 \end{bmatrix}.$$

In order to compare these columns with one another, he divides each by its first element after the manner of Duncan & Collar (1934), obtaining respectively

$$\begin{bmatrix} 1{\cdot}00 \\ 1{\cdot}25 \\ 0{\cdot}75 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0769 \\ 0{\cdot}6538 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0701 \\ 0{\cdot}6306 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0737 \\ 0{\cdot}6325 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0739 \\ 0{\cdot}6330 \end{bmatrix}.$$

When thus divided the successive columns appear to be converging. It is well known in connexion with the infinite series $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\ldots+\frac{1}{n}+\ldots$ that an appearance of

convergence may be deceptive. If the convergence had been final, in the sense that X^4 had been precisely equal to $5589X^3/936$, we could have argued thus: $X^4 = KX^3 + 4X^3$ by definition; therefore $KX^3 = X^3(5589/936 - 4) = 1{\cdot}9712X^3$; therefore the latent root is 1·9712, and X^3 is its latent column, by the definition of a latent root. It is, however, neater and more satisfactory to regard X^4 as the purified column, and to test it in a manner which disregards entirely the processes by which it was obtained, and which uses its elements equally. Let each element of KX^4 be divided by the corresponding element of X^4. The result of this 'division-test' is

$$\left\{\frac{11027}{5589},\ \frac{11839}{6002},\ \frac{6979}{3538}\right\} = \{1{\cdot}9730,\quad 1{\cdot}9725,\quad 1{\cdot}9726\}.$$

If these quotients had been precisely equal, there would have been no doubt that they equalled a latent root λ, and that X^4 was its latent column, directly from the definition. Actually the imperfect equality leaves open the question as to how the true λ is related to the estimates. The answer, for symmetrical K, is that the true λ lies near to Rayleigh's mean $\tilde{X}KX/\tilde{X}X$, which is discussed on pp. 449 to 450. For unsymmetrical K a simple answer is unfortunately lacking. In the example of § 5 one of the true λ is found to lie outside the range of its estimates. This peculiarity is explained in § 8 by the almost proportionality between standard columns which cannot be proportional. It is probably a rare phenomenon.

In the above example it may seem strange that a purification which began with various surmises has led rapidly to a result which looks fairly definite. The coefficient ρ in the multiplier $K - \rho I$ was taken to be $\rho = -4$ on very slight evidence. In general, ρ is to be regarded as an estimate of an unwanted latent root. Further explanation will be found on p. 482.

From the point of view of purification it is reasonable, now that one constituent of the raw material has been found, to clean that one away before looking for anything else. Accordingly, if the other latent roots are required, the computer should work out

$$(K - 1{\cdot}973I)\{1, 1, 1\} \quad \text{which comes to} \quad \{1{\cdot}027,\ 1{\cdot}027,\ -1{\cdot}973\} = Z^0, \text{ say.}$$

If the matrix K had many rows it might now be necessary to make further iterations starting from Z^0. But actually all but two of the latent roots are already known, so that the following brief procedure is available. From the Cayley-Hamilton theorem

$$(K - \lambda_2 I)(K - \lambda_3 I)Z^0 = \mathbf{0} \quad \text{or nearly so.}$$

But
$$\lambda_2 + \lambda_3 = -6 - \lambda_1 = -7{\cdot}973.$$

So
$$K^2Z^0 + 7{\cdot}973KZ^0 = -\lambda_2\lambda_3 Z^0.$$

In this there is only one unknown, namely, $\lambda_2\lambda_3$. Computation gives

$$K^2Z^0 + 7{\cdot}973KZ^0 = \{-17{\cdot}192,\ -17{\cdot}192,\ 33{\cdot}000\}.$$

On dividing each element of this column by the corresponding element of $-Z^0$ one obtains

$$\lambda_2\lambda_3 = \{16{\cdot}740,\ 16{\cdot}740,\ 16{\cdot}726\}.$$

These three estimates agree fairly well. From $\lambda_2 + \lambda_3$ and $\lambda_2\lambda_3$ the separate latent roots are then found to be

$$\lambda_2 = -3{\cdot}986_5 + 0{\cdot}918i, \quad \lambda_3 = -3{\cdot}986_5 - 0{\cdot}918i.$$

These are the roots of greatest modulus. For the application to arms-races the desired root was that of greatest real part, namely, λ_1. That is the end of the groping operations, for all the latent roots are known approximately.

Now fine purification can begin. If any latent root is required to be known with greater accuracy, this can be achieved by taking the best previous estimate of its latent column, and then cleaning that with the aid of the best estimates of the other two latent roots. For example, suppose that $\lambda_1 = 1{\cdot}973$ is not sufficiently accurate. The best previous estimate of its latent column is

$$X^4 = \{5589,\ 6002,\ 3538\}.$$

The premultiplier for cleaning X^4 should ideally be $(K-\lambda_3 I)\,(K-\lambda_2 I)$, the best approximation to which at the present stage is $K^2+7{\cdot}973K+16{\cdot}735I$. By computation in which all digits were retained

$$(K^2+7{\cdot}973K+16{\cdot}735I)\begin{bmatrix}5589\\6002\\3538\end{bmatrix} \text{ was found to be } \begin{bmatrix}203199{\cdot}186\\218196{\cdot}817\\128619{\cdot}997\end{bmatrix} = W^0, \text{ say.}$$

Next KW^0 was computed and its elements were divided by the corresponding elements of W^0 to give respectively

$$\lambda_1 = 1{\cdot}97273242 \quad \text{or} \quad 1{\cdot}97273234 \quad \text{or} \quad 1{\cdot}97273235.$$

These agree to 7 significant digits. If desired still greater accuracy could be obtained by computing first

$$W^1 = (K^2+7{\cdot}973K+16{\cdot}735I)\,W^0$$

and thence KW^1, and finally by dividing each element of KW^1 by the corresponding element of W^1.

The above example is typical of the method to be discussed more fully in the rest of this paper. It may be contrasted with a method which Hotelling in 1943 described as 'the fundamental iterative method for finding latent roots and vectors', namely, that in which an arbitrary column is premultiplied repeatedly by the given matrix. This corresponds to $\rho = 0$. The same method has been extensively used by Duncan & Collar (1934) and by Aitken (1937). This so-called fundamental method has been applied to the matrix of the immediately previous example by computing sequences $Y^{s+1} = KY^s$. Two initial columns have been taken; one sequence begins with $Y^0 = \{4, 5, 3\}$ so as to allow strict comparison between the effects of $\rho = 0$ and those of $\rho = -4$; the other sequence begins with $\{1, 1, 1\}$, because that has often been taken by Duncan & Collar, by Aitken, and by Hotelling. The sequences run

$$\begin{bmatrix}4\\5\\3\end{bmatrix},\ \begin{bmatrix}10\\8\\5\end{bmatrix},\ \begin{bmatrix}13\\24\\11\end{bmatrix},\ \begin{bmatrix}44\\13\\28\end{bmatrix},\ \begin{bmatrix}53\\162\\-14\end{bmatrix};$$

$$\begin{bmatrix}1\\1\\1\end{bmatrix},\ \begin{bmatrix}3\\3\\0\end{bmatrix},\ \begin{bmatrix}0\\3\\9\end{bmatrix},\ \begin{bmatrix}-81\\24\\117\end{bmatrix},\ \begin{bmatrix}456\\-57\\-384\end{bmatrix},\ \begin{bmatrix}-1665\\714\\1494\end{bmatrix},\ \begin{bmatrix}6861\\-3435\\-4719\end{bmatrix},\ \begin{bmatrix}-24453\\18015\\14148\end{bmatrix}.$$

Neither of them has so far converged. I do not say that they are useless, but only that they are not leading to the desired latent column. Aitken (1937) has shown how to extract information from such apparently irregular sequences by his 'δ^2-process'. The procedure of repeated multiplication by K finds the latent root, or roots, of greatest modulus, though not necessarily by the best process. This root may, or may not, be what one wishes to know. Frazer *et al.* (1938, p. 143) show how the remaining roots can be obtained successively in the order of their moduli. The present method of repeated multiplication by $K-\rho I$ allows the computer more freedom to pursue the desired latent root whichever it may be.

Example 2. Required to find the middle latent root of

$$K = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & 5 \\ 2 & 5 & 11 \end{bmatrix},$$

a matrix studied by Duncan & Collar (1934). A computer, who does not know Duncan & Collar's results, may reasonably be supposed to think and act as follows. The mean of the roots is 6; so λ_2 may be near 6. Next, as an experiment, he sums along the rows of K obtaining

$$K\{1, 1, 1\} = \{6, 12, 18\};$$

so λ_1 may be near 12, and if so λ_3 will be near zero. Rayleigh's principle would, however, hardly justify the averaging of estimates of λ_1 so discordant as 6, 12, 18; therefore the computer tries to remove some of the P_2 and P_3 from $\{1, 1, 1\}$ by multiplying it by $(K-6I)\,(K-0I)$. The result is

$$(K-6I)\,K\{1, 1, 1\} = (K-6I)\,\{6, 12, 18\} = \{36, 90, 162\} = X^2, \quad \text{say.}$$

Then $KX^2 = \{576, 972, 2304\}$. The ratios of the elements of KX^2 to the corresponding elements of X^2 are 16, 10·8, 14·2. These ratios are less discordant than 6, 12, 18, and their mean is 13·7. The present state of information is that λ_1 is about 14, λ_2 perhaps near 6, and if so λ_3 is near -2. If so $(K-14I)\,(K+2I)$ should clean away P_1 and P_3 so as to leave P_2. The computer wishes that he had some easy way of approximating to λ_3. However, none occurs to him, so he works out

$$(K-14I)\,(K+2I) = K^2-12K-28I = \begin{bmatrix} -40 & 0 & 12 \\ 0 & -34 & 24 \\ 12 & 24 & -10 \end{bmatrix} = 2J, \quad \text{say.}$$

He has no idea of the shape of the desired P_2, and so he takes the initial column to be $Y^0 = \{1, 1, 1\}$, doubting its suitability. He multiplies repeatedly by J; but in order to watch for convergence he divides each successive column by its first member. Thus the sequence is defined by $Y'^s = JY^{s-1}$, and then Y^s equals Y'^s divided by its first member. The sequence Y^s goes thus:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 0{\cdot}3571 \\ -0{\cdot}9286 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 0{\cdot}6732 \\ -0{\cdot}5838 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 0{\cdot}7850 \\ -0{\cdot}7232 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 0{\cdot}9049 \\ -0{\cdot}7821 \end{bmatrix}, \quad \begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0031 \\ -0{\cdot}8411 \end{bmatrix}.$$

There is an appearance of slow convergence. The computer thinks that he might hasten it if he had any estimate of λ_2 better than 6. So he breaks off the Y^s sequence, takes its last member which is Y^5 and computes KY^5 thus:

$$KY^5 = \{2{\cdot}3240, 2{\cdot}8100, -2{\cdot}2366\}.$$

On division by the corresponding elements of Y^5 these give $\lambda_2 = 2{\cdot}32, 2{\cdot}80, 2{\cdot}66$. Evidently his first notion that λ_2 was 6 must now be rejected. The simple mean of the three estimates gives $\lambda_2 = 2{\cdot}56$ and Rayleigh's type of mean is $\tilde{Y}^5KY^5/\tilde{Y}^5Y^5 = 2{\cdot}59$. At this stage the best estimates are that λ_1 is about 14, λ_2 is $2{\cdot}5_9$, and therefore λ_3 is about unity instead of -2 as formerly supposed. The cleaning away of P_3 can now be speeded by taking $(K-14I)\,(K-I)$

as the multiplier instead of $(K-14I)(K+2I)$. The initial column for this new sequence is the best available estimate of P_2, namely, Y^5. The sequence goes

$$\begin{bmatrix} 1{\cdot}0000 \\ 1{\cdot}0031 \\ -0{\cdot}8411 \end{bmatrix}, \begin{bmatrix} 1{\cdot}00000 \\ 1{\cdot}36683 \\ -1{\cdot}05335 \end{bmatrix}, \begin{bmatrix} 1{\cdot}00000 \\ 1{\cdot}35240 \\ -1{\cdot}04502 \end{bmatrix}, \begin{bmatrix} 1{\cdot}00000 \\ 1{\cdot}35288 \\ -1{\cdot}04526 \end{bmatrix}, \begin{bmatrix} 1{\cdot}000000 \\ 1{\cdot}352859 \\ -1{\cdot}045263 \end{bmatrix}, \begin{bmatrix} 1{\cdot}0000000 \\ 1{\cdot}3528600 \\ -1{\cdot}0452593 \end{bmatrix}.$$

The sequence now appears to be converging very rapidly. The proof of the result is, however, not any appearance of convergence, but the agreement between three values of λ_2, obtained from the last column, Y^{10}, thus:

$$KY^{10} \div Y^{10} = \{2{\cdot}615201,\ 2{\cdot}615203,\ 2{\cdot}615191\}.$$

The required latent root is thus determined to one part in 200,000.

The purpose of this example has been to show that a middle root can be obtained accurately with the aid of rough estimates of the other roots. Their true values are

$$\lambda_1 = 14{\cdot}43, \quad \lambda_3 = 0{\cdot}9539_3.$$

Looking back, the computer wishes that he had thought of making this simple experiment at the outset:

$$K\{1,\ 1,\ -1\} = \{2,\ 2,\ -4\};$$

for this would have given estimates $\lambda = 2, 2, 4$ of which Rayleigh's mean is 2·67, quite a useful approximation to λ_2. But how could the computer have thought of that?

When one has become accustomed to the ideal of multiplication by $n-1$ of the n brackets

$$(K-\lambda_1 I)(K-\lambda_2 I)\ldots(K-\lambda_n I),$$

and to the expedient of repeated multiplication by $K-\rho I$, and when one looks back from this point of view at the repeated multiplications by K which are recommended elsewhere, they appear to involve the assumption that the unwanted latent roots are clustered near zero; and for that assumption there is in general no warrant.

2. Preliminary estimates

The iterations have to begin with a trial column and an estimate of an unwanted latent root.

Bounds to the latent roots of any square matrix

A theorem due to Hirsch about matrices with complex elements reduces to the following when the elements are real. Any real matrix K can be expressed as

$$K = H+S, \quad \text{where} \quad \tilde{H} = H \quad \text{and} \quad \tilde{S} = -S. \tag{10}$$

Let k, h, s denote the maxima of the absolute values of the elements of K, H, S respectively. Any latent root λ of K is expressible as $\lambda = \alpha+i\beta$, where α and β are real. Hirsch's theorem is then that

$$|\lambda| \leqslant nk, \quad |\alpha| \leqslant nh, \quad |\beta| \leqslant ns, \tag{11}$$

where n is the number of rows in K. These inequalities have the great merit of having been definitely proved (MacDuffee 1933, pp. 25–26); yet the bounds are often so much in excess of the actual value as to be of little use. Although I do not know of any simple infallible

formula suitable for estimating the roots in all cases, yet there certainly are easy self-checking procedures, which can be stated in words, that usually lead rapidly to estimates much nearer to an actual real root than is Hirsch's bound. One makes a few experiments with trial columns in the following manner.

Trial columns

It is an easy experiment to try a column such as

$$X = \{1, 1, 1\} \quad \text{or} \quad X = \{1, -1, 1, -1\} \quad \text{or} \quad X = \{0, 1, 2, 3, \ldots, n\},$$

or other sets of small whole numbers. It may happen that the n estimates $KX \div X$ are very discordant. If so that experiment fails; yet only a little time has been wasted. On the contrary, it may happen that the n estimates $KX \div X$ are of the same order of magnitude; if so a useful hint is obtained.

Example 3. Given

$$K = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Try

$$X = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{finding} \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 32 \\ 50 \end{bmatrix}.$$

So $KX \div X = \{14, 16, 16\frac{2}{3}\}$, and this suggests a latent root near 16. A first attempt at improvement was to take $Y = \{12, 21, 30\}$. This leads to $KY \div Y = \{12, 15{\cdot}9, 17{\cdot}4\}$. That has increased the gradient of the three estimates. So throw Y away, and make a move from X in the opposite direction to $W = \{8, 19, 30\}$. This leads to $KW \div W = \{17, 16{\cdot}2, 15{\cdot}9\}$, and confirms that there is a latent root near 16. Formal purifications could now begin. Hirsch's procedure, on the contrary, is

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 5 & 7 \\ 5 & 7 & 9 \end{bmatrix} + \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix},$$

so that $h = 9$, $s = 2$, whence $|\alpha| \leqslant 27$, $|\beta| \leqslant 6$.

The next two examples show that after a trial column has located a latent root, there remains the question as to which of the roots it is.

Example 4.

$$\begin{bmatrix} 1 & 9 & 9 \\ 9 & 1 & 9 \\ 9 & 9 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 19 \\ 19 \\ 19 \end{bmatrix}.$$

Therefore $\lambda = 19$. The other two latent roots are -8, -8; so 19 is the greatest. Hirsch's bound is $|\lambda| \leqslant 27$.

Example 5.

$$\begin{bmatrix} 1 & -9 & -9 \\ -9 & 1 & -9 \\ -9 & -9 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -17 \\ -17 \\ -17 \end{bmatrix}.$$

Therefore $\lambda = -17$. The other two roots are 10, 10; so -17 is the least. Hirsch's bound is $|\lambda| \leqslant 27$.

Further examples of trial columns, used in conjunction with Rayleigh's mean, are given in the section on symmetrical matrices.

Guessing is usually considered to be beneath the dignity of scientific research. In the wrong place guessing would, of course, be a contemptible form of careless inaccuracy. Duncan & Collar (1934), Aitken (1937) and Hotelling (1943*a*) began their iterations with what they called the 'arbitrary initial vector'; and they often took it to be $\{1, 1, \ldots, 1\}$ without stating any reason for this choice. From the point of view of Harold Jeffreys's *Theory of probability* (1939) the column $\{1, 1, 1, \ldots, 1\}$ may be regarded as the correct expression of complete ignorance. As such it is most inappropriate to the learned authors named above. An example will be given in § 5 where $Y^0 = \{1, 1, 1, 1\}$ is an unsuitable beginning, because it contains only a small proportion of the desired constituent. So, if the computer has any experience of matrices at all like the given one, I would advise him to use that experience by making a *guess* at the initial column. A bad initial guess cannot lead to a wrong result, because the ultimate consequences of the guess are subjected to a test of achievement. But a bad guess wastes, and a good guess saves, arithmetical toil. The chemical analogue of a good guess is a raw material rich in the desired constituent.

3. Symmetrical matrices

Symmetrical matrices deserve early mention because of their occurrence in correlation analysis* and because of their peculiar simplicities. They do not, however, suffice for all applications of social interest; for example, in the theory of arms-races, love and hate may not be quite mutual, and the matrix may be unsymmetrical. The peculiar simplicities of symmetrical matrices include the following:

(i) The latent roots are all real.

(ii) Although equal latent roots may occur, yet the associated elementary divisors are all linear.

(iii) The latent columns are mutually orthogonal if the latent roots are distinct. If equal roots occur, their associated columns can be chosen orthogonal to each other and to the rest.

Proofs of (i) go back to Lagrange and are given in many texts.

Proof of (ii) is attributed by Bromwich (1906, Art. 26) to Weierstrass.

Proof of (iii) appears in H. & B. Jeffreys (1946, Arts. 4·031 and 4·083).

The relation of the property (ii) to computing operations may be illustrated as follows:

Example 6. $K = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$. It has a pair of equal latent roots $\lambda_1 = \lambda_2 = 3$. The premultiplier $K - 3I$ annihilates the latent columns associated with both roots, because

$$(K-3I)\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for any a and b.

Contrast example 10 on p. 470.

Example 7. In connexion with the following K, consider the three trial columns X, Y, Z:

$$K = \begin{bmatrix} \alpha & \beta & \beta \\ \beta & \alpha & \beta \\ \beta & \beta & \alpha \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad Y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

* *Note added* 17 *January* 1950. The application of the purification method to psychological correlations is illustrated elsewhere (Richardson 1950).

Multiplication shows that

$$KX = (\alpha+2\beta)\,X, \quad KY = (\alpha-\beta)\,Y, \quad KZ = (\alpha-\beta)\,Z.$$

So X, Y, Z are three latent columns of K. They are mutually orthogonal. The linear premultiplier $K-(\alpha+\beta)\,I$ annihilates both Y and Z. Any mixture such as $bY+cZ$, where b and c are scalars, also satisfies

$$(K-(\alpha-\beta)\,I)\,(bY+cZ) = \mathbf{0},$$

so that the latent columns for this double root may be chosen variously. Such choices are discussed by H. & B. Jeffreys (1946, Art. 4·083). To find two linearly independent latent columns belonging to the same double root it would be necessary to begin iterations with different trial columns.

Although the property (ii) has long been known, perhaps a proof of the related computing operations may be acceptable here. Let $\lambda_1 = \lambda_2$. Let A be an arbitrary column and let Q be defined by

$$Q = (K-\lambda_3 I)\,(K-\lambda_4 I)\,\ldots\,(K-\lambda_n I)\,A.$$

Then by the Cayley-Hamilton theorem

$$(K-\lambda_1 I)\,(K-\lambda_1 I)\,Q = \mathbf{0}.$$

Define P by

$$(K-\lambda_1 I)\,Q = P.$$

Then

$$(K-\lambda_1 I)\,P = \mathbf{0}.$$

So far this is general for a pair of equal roots of any sort of matrix K. When, however, $K = \tilde{K}$, the transposition of $KP = \lambda_1 P$ gives $\lambda_1\tilde{P} = (KP)^{\sim} = \tilde{P}\tilde{K} = \tilde{P}K$. Postmultiply both sides by Q obtaining $\lambda_1\tilde{P}Q = \tilde{P}KQ$. Substitute for KQ its equal $\lambda_1 Q+P$ obtaining $\lambda_1\tilde{P}Q = \tilde{P}(\lambda_1 Q+P)$; whence $\tilde{P}P = 0$. This entails that P is a column of zeroes. Accordingly the general formula $(K-\lambda_1 I)\,Q = P$ simplifies to $(K-\lambda_1 I)\,Q = \mathbf{0}$ when $K = \tilde{K}$. That is to say, *the columns associated with a double root are annihilated by a linear premultiplier when the matrix is symmetrical*, although otherwise a quadratic premultiplier might be necessary.

Rayleigh's mean

Let P_1 denote any one of the n latent columns. Suppose that we have a column X consisting mostly of P_1 but slightly contaminated by admixture with $P_2, P_3, \ldots, P_n$. Then $KX \div X$ gives n slightly unequal estimates of λ_1. But

$$\frac{\tilde{X}KX}{\tilde{X}X} = \mu, \quad \text{say,} \tag{12}$$

gives a single estimate. It may suitably be called 'Rayleigh's mean', because Rayleigh pointed out that this type of averaging by way of two quadratic forms has a purifying effect (Rayleigh 1894, p. 110; H. & B. Jeffreys 1946, Art. 4·084; Temple & Bickley 1933). A usual proof is to expand X thus

$$X = \sum_{\nu=1}^{\nu=n} a_\nu P_\nu, \tag{13}$$

in which the n columns P_ν are mutually orthogonal and normalized. It then follows that

$$\frac{\tilde{X}KX}{\tilde{X}X} = \frac{\Sigma a_\nu^2 \lambda_\nu}{\Sigma a_\nu^2}. \tag{14}$$

Orthogonality is available if K is symmetrical, but not in general. When there are non-linear elementary divisors Rayleigh's mean is not much use (see p. 479), but for symmetrical matrices it is often superbly efficient.

Example 8 (from Thurstone 1935, pp. 125–127).

$$K = \begin{bmatrix} 1{\cdot}95 & -1{\cdot}07 & 0{\cdot}69 \\ -1{\cdot}07 & 0{\cdot}75 & 0{\cdot}11 \\ 0{\cdot}69 & 0{\cdot}11 & 1{\cdot}71 \end{bmatrix}.$$

Try $X = \{1, 1, 1\}$. It follows that

$$KX \div X = \{1{\cdot}57, -0{\cdot}21, 2{\cdot}51\}.$$

These estimates of a root λ are too discordant, so that the experiment fails. Instead try $Y = \{1, -1, 1\}$. It follows that $KY = \{3{\cdot}71, -1{\cdot}71, 2{\cdot}29\}$; so that the $KY \div Y$ are all of one sign. Rayleigh's mean is $\mu = 2{\cdot}57$, and this is a useful first approximation to a latent root $\lambda_1 = 2{\cdot}849690$ given by Thurstone. In interpreting Thurstone's statements it must be noticed that his β is minus our λ. Hirsch's bound is $|\lambda| \leqslant 5{\cdot}85$, and is much in excess. The other latent roots according to Thurstone are $\lambda_2 = 1{\cdot}560310$ and $\lambda_3 = 0$.

Example 9. Thurstone (1935, pp. 131–133) discussed the following matrix:

$$K = \begin{bmatrix} 5{\cdot}011317 & 0{\cdot}165747 & 0{\cdot}076365 & 0{\cdot}015854 \\ 0{\cdot}165747 & 1{\cdot}183860 & -0{\cdot}053695 & -0{\cdot}040502 \\ 0{\cdot}076365 & -0{\cdot}053695 & 0{\cdot}428619 & -0{\cdot}044782 \\ 0{\cdot}015854 & -0{\cdot}040502 & -0{\cdot}044782 & 0{\cdot}341573 \end{bmatrix}.$$

It is noticeable that the elements on the principal diagonal have larger moduli than the others. If the other elements were zero then the latent columns would be

$$\{1, 0, 0, 0\}, \quad \{0, 1, 0, 0\}, \quad \{0, 0, 1, 0\}, \quad \{0, 0, 0, 1\}.$$

It is suitable therefore to take these columns for trial. The first of them gives

$$K\{1, 0, 0, 0\} = \{5{\cdot}011317, 0{\cdot}165747, 0{\cdot}076365, 0{\cdot}015854\}.$$

The division test gives

$$K\{1, 0, 0, 0\} \div \{1, 0, 0, 0\} = \{5{\cdot}011317, \infty, \infty, \infty\}.$$

These estimates appear hopelessly discordant. Yet Rayleigh's mean gives

$$\mu = \frac{[1, 0, 0, 0]\, K\{1, 0, 0, 0\}}{[1, 0, 0, 0]\, \{1, 0, 0, 0\}} = 5{\cdot}011317, \quad \text{or say } 5{\cdot}01.$$

In like manner Rayleigh's means for the other three trial columns are equal to diagonal elements of K. So the four latent roots are estimated as 5·01, 1·18, 0·43, 0·34. Thurstone (p. 133, putting $\lambda = -\beta$) states a quartic equation for λ, and its roots as 5·0197117, 1·182688, 0·444971, 0·317999. The ease with which Rayleigh's mean has led to useful approximations is remarkable. Hirsch's bound to the latent roots is $|\lambda| \leqslant 20{\cdot}04$, which is true but useless. It appears from this example that ∞ should be disregarded when the concordance of $KX \div X$ is considered; for infinity arises from a zero element in a column; and this element may be slightly in error.

The choice of cleaning operations for symmetrical matrices

In the final stage of purification all the latent roots may be roughly known, and the best procedure is then to take ρ in the premultiplier $K-\rho I$ equal in turn to each root as nearly as may be.

The following theory provides a useful guide at an earlier stage when the greatest and least of the latent roots are beginning to be located, but little is known about the others. An arbitrary column A^0 can be expanded in the form

$$A^0 = \sum_{\nu=1}^{\nu=n} a_\nu P_\nu, \tag{15}$$

in which a_ν is a scalar and P_ν is the latent column satisfying

$$KP_\nu = \lambda_\nu P_\nu. \tag{16}$$

It is here that the symmetry of K guarantees simplicity. For matrices in general there is an expansion, but the terms may behave differently, so that the theory of this section may not apply.

The present theory should be kept in the background of the computer's mind. He does not need to work out the numerical values of the coefficients a_ν. The mere knowledge that such an expansion exists gives him much useful guidance. Or, in a large organization, the following theory might suitably be known to the mathematician-in-charge, but the routine-computers could be excused from knowing it.

The expansion is possible because the n columns $P_1, P_2, \ldots, P_n$ are orthogonal and therefore linearly independent. Now let ρ be any real number and let the computer premultiply A^0 by $K-\rho I$ to obtain A^1. Although he is unaware of the details of the expansion, his arithmetic corresponds to

$$A^1 = (K-\rho I)\, A^0 = \sum_{\nu=1}^{\nu=n} (\lambda_n-\rho)\, a_\nu P_\nu. \tag{17}$$

Suppose that the computer desires to find a latent root λ_s and its P_s by cleaning away all the other P. Here s takes a meaning different from that in §1. To keep the coefficient of P_s constant, so that it may serve as a standard, divide through by $\lambda_s-\rho$, obtaining

$$\frac{A^1}{\lambda_s-\rho} = \sum_{\nu=1}^{\nu=n} \left(\frac{\lambda_\nu-\rho}{\lambda_s-\rho}\right) a_\nu P_\nu. \tag{18}$$

In order that the desired cleaning may occur, ρ must be such as to make

$$-1 < \left(\frac{\lambda_\nu-\rho}{\lambda_s-\rho}\right) < 1 \tag{19}$$

for every ν except s. We may think of the λ and ρ as $n+1$ points on the real axis. The condition can then be expressed by saying that *ρ must be farther from the desired root than from any of the other roots.*

For definiteness let the roots, being real, be arranged in order of magnitude so that

$$\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \ldots \geqslant \lambda_{n-1} \geqslant \lambda_n. \tag{20}$$

At the groping stage, the computer does not know where the unwanted latent roots λ_ν are, nor what coefficients a_ν are attached to their latent columns P_ν in the expansion (15) of his initial column A^0. A suitable expression for his ignorance can be arranged as follows. First regard a_ν as a mass attached to the λ-axis at the point $\lambda = \lambda_\nu$. Then, as both λ_ν and a_ν are unknown, regard these discrete masses as spread out uniformly along the λ-axis throughout the range $\lambda_n \leqslant \lambda \leqslant \lambda_1$. This picture may suitably be called a uniform density of probability expressing ignorance (Jeffreys 1939). It is proper to omit the suffix ν to λ_ν which now becomes the continuous co-ordinate λ. Let

$$f = (\lambda - \rho)\,(\lambda_s - \rho)^{-1}. \tag{21}$$

A suitable name for f is the 'ratio of purification'. For trustworthy purification ρ must be chosen so that throughout the range of λ where any unwanted root can lie

$$-1 < f < 1. \tag{22}$$

It is further desirable that f should be as near to zero as possible throughout the same range. This requirement can be expressed, in accordance with Legendre's principle of least squares, by making

$$G^2 = \frac{1}{\lambda_1 - \lambda_n} \int_{\lambda_n}^{\lambda_1} f^2 \, d\lambda \tag{23}$$

a minimum for variations of ρ. Here G may be called the 'root mean square ratio of purification'. It is conceivable that a choice of ρ which made the average of f^2 a minimum might yet allow f to stray outside the necessary bound for a particular value of λ; so both conditions (22) and (23) should be retained.

The pure mathematician will notice a slight ambiguity which runs throughout this section: the same symbol, say λ_1, may stand either for the accurate value which the computer is seeking, or for the rough approximation which he has already found. The latter could be distinguished by an accent as λ_1'; but the distinction is hardly worth the encumbrance; for computers are accustomed to use many standard symbols such as $\sqrt{2}$, π and e somewhat ambiguously for various approximations, without becoming confused as to their exact meanings.

When the desired root λ_s is the greatest root λ_1, then the condition $-1 < f < 1$ is satisfied by making

$$\rho < \tfrac{1}{2}(\lambda_1 + \lambda_n). \tag{24}$$

The semi-infinity of the range to which ρ is thus confined explains why very different iterations tend to a multiple of P_1 if carried far enough. The relations are illustrated in figure 1. The abscissa is λ, the ordinate is f. The effect, good or bad, of any multiplication by $K - \rho I$ depends on the function $f = (\lambda - \rho)/(\lambda_1 - \rho)$; and this is represented by any straight line which passes through the point $\lambda = \lambda_1, f = 1$. The value of ρ is readily visible as the point where the sloping line cuts the middle horizontal, because $\lambda = \rho$ where $f = 0$. Only some of the radii through the point $\lambda = \lambda_1$, $f = 1$ represent reliable cleaning operations, namely those for which the portion of the sloping line between the vertical lines lies entirely between the upper and lower horizontals. Five such sloping lines are shown. They intersect the axis $f = 0$ at points for which $\rho < \frac{1}{2}(\lambda_n + \lambda_1)$. Although any one of these represents an infallible cleaner, some are more efficient than others. It is desirable that the ordinate of the sloping line should be as near to zero as possible within the range between the verticals at λ_n and λ_1.

The two sloping lines which would, if produced, cut the axis $f = 0$ at $\rho < \lambda_n$ obviously represent cleanings less efficient than those for which $\rho \geqslant \lambda_n$. The practical range of ρ is thus narrowed to

$$\lambda_n \leqslant \rho < \tfrac{1}{2}(\lambda_n + \lambda_1). \tag{25}$$

The thick line, for which $\lambda_1 - \rho = 2(\rho - \lambda_n)$, represents the purification which is best on the average, in the sense that it secures the minimum of G^2 in (23). The proof is as follows. From figure 1 it is evident that the position of any point on the horizontal axis should be reckoned as a fraction of the range from λ_1 to λ_n. Accordingly, let a co-ordinate x and a constant α be defined by

$$x = \frac{\lambda_1 - \lambda}{\lambda_1 - \lambda_n}, \tag{26}$$

$$\alpha = \frac{\lambda_1 - \rho}{\lambda_1 - \lambda_n}, \tag{27}$$

so that

$$\lambda - \rho = (\alpha - x)(\lambda_1 - \lambda_n). \tag{28}$$

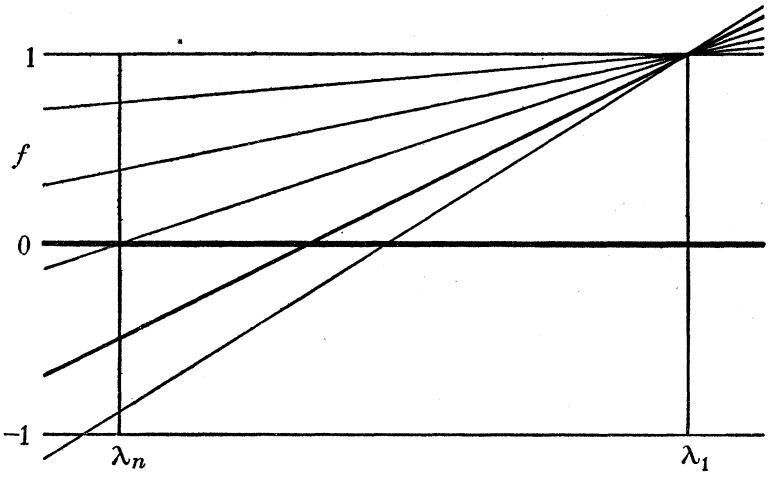

FIGURE 1. The latent roots are points on the abscissa. The ordinate is the ratio of purification. Each sloping line shows the effect of multiplication by $(K - \rho I)/(\lambda_1 - \rho)$ for a particular value of ρ. The thick sloping line shows the best choice of ρ when the greatest root λ_1 is required.

Then from (21), in which λ_s is here λ_1, the ratio of purification simplifies to

$$f = 1 - \frac{x}{\alpha}. \tag{29}$$

The mean square of f is

$$G^2 = \frac{1}{\lambda_1 - \lambda_n}\int_{\lambda_n}^{\lambda_1} f^2 d\lambda = \int_0^1 \left(1 - \frac{x}{\alpha}\right)^2 dx = 1 - \frac{1}{\alpha} + \frac{1}{3\alpha^2}. \tag{30}$$

Because $d\alpha/d\rho$ is constant, the minimum of G^2 with respect to ρ is the same as that with respect to α. The conditions that G^2 should be a minimum are therefore

$$0 = \frac{1}{\alpha^2} - \frac{2}{3\alpha^3} \tag{31}$$

and

$$-\frac{2}{\alpha^3} + \frac{2}{\alpha^4} > 0. \tag{32}$$

Although $\alpha \to \infty$ would satisfy (31), it would violate (32). Figure 1 shows that α cannot be zero. The conditions (31) and (32), after multiplication respectively by the finite positive quantities α^3 and α^4, become

$$\alpha = \tfrac{2}{3} \tag{33}$$

and

$$1-\alpha > 0, \tag{34}$$

which are consistent.

These are both satisfied by

$$\rho = \tfrac{1}{3}(2\lambda_n + \lambda_1). \tag{35}$$

Moreover, this best-on-the-average choice of ρ is reliable everywhere, because it is consistent with $\rho < \frac{1}{2}(\lambda_n + \lambda_1)$.

It is noticeable that the point $\lambda = 0$ is not marked on figure 1. That point would be irrelevant. The corresponding ρ would be zero and the corresponding operation would be multiplication simply by K, as practised by several other authors. It is thus evident that multiplication by K may happen to be a suitable operation for particular matrices and particular latent roots, but that it cannot be suitable for all symmetrical matrices when the greatest root is sought. It finds the root of greatest modulus, and that may be the least root.

When the desired root is the least root λ_n, a closely similar theory applies. It is necessary for reliability everywhere that $\rho > \frac{1}{2}(\lambda_1 + \lambda_n)$. It is disadvantageous to make $\rho > \lambda_1$. The best choice on the average is

$$\rho = \tfrac{1}{3}(2\lambda_1 + \lambda_n). \tag{36}$$

When the desired root λ_s *is neither the greatest nor the least*, but is suspected to lie at $\lambda = s$, no linear premultiplier is competent to find it. Example 2 illustrated this situation. We must then consider instead a quadratic premultiplier with ρ and σ, one on either side of the desired root λ_s. Thus define A^1 by

$$A^1 = (K-\rho I)(K-\sigma I) A^0. \tag{37}$$

Then

$$\frac{A^1}{(\lambda_s-\rho)(\lambda_s-\sigma)} = \sum_{\nu=1}^{\nu=n} \frac{(\lambda_\nu-\rho)(\lambda_\nu-\sigma)}{(\lambda_s-\rho)(\lambda_s-\sigma)} a_\nu P_\nu. \tag{38}$$

The coefficient of P_s is the same in this expansion as it was in that of A^0. It is desired that the coefficients of all the other P should move towards zero. Let g be defined by

$$g = \frac{(\lambda-\rho)(\lambda-\sigma)}{(s-\rho)(s-\sigma)}, \tag{39}$$

in which the unknown λ_s has been replaced by its suspected value s. Let g be called the 'ratio of purification'. The suffix ν no longer appears in g because the roots λ_ν are suspected, but not known. The desired cleaning will occur if ρ and σ can be chosen so as to make $|g| < 1$ for all λ in the two ranges

$$\lambda_n \leqslant \lambda < s \quad \text{and} \quad s < \lambda \leqslant \lambda_1.$$

At $\lambda = s$, $g = 1$. To make unity the maximum of g it is necessary that $dg/d\lambda$ should vanish at $\lambda = s$. But $dg/d\lambda$ vanishes at $\lambda = \frac{1}{2}(\rho+\sigma)$. Therefore ρ and σ must be chosen equidistant from s. Let the origin be shifted to s by defining y and r thus:

$$y = \lambda - s, \tag{40}$$

$$\rho = s - r, \tag{41}$$

$$\sigma = s + r. \tag{42}$$

Then from (39)
$$g = \frac{(y+r)(y-r)}{-r^2} = 1-\frac{y^2}{r^2}. \tag{43}$$

In order that the purification may be reliable, we must have $-1<g<1$ for all y in the range $\lambda_n-s\leqslant y\leqslant\lambda_1-s$, except at $y=0$. If this condition were broken anywhere, it would be by g becoming less than -1 at an extreme of the range. The critical condition is $y^2/r^2=2$, and so r must be such that
$$r^2>\tfrac{1}{2}(\lambda_1-s)^2 \tag{44}$$
and
$$r^2>\tfrac{1}{2}(\lambda_n-s)^2. \tag{45}$$

If these inequalities are both satisfied, then repeated multiplication by $(K-\rho I)(K-\sigma I)$ will tend to give the desired P_s if its λ_s lies near enough to s; but how fast or slowly is a further question.

It will be noticed that the formulae for choosing a purifier to find λ_s involve s which is an estimate of λ_s, after the manner of successive approximations. The computer must take a risk by inserting his suspected value of λ_s to see whether it leads on, as it usually should, to a more accurate value of λ_s. The risk which he must take is that of wasting his time; there is no risk of a wrong result, because the result will be checked by a test of achievement (§§ 4 and 7).

The best choice of r on the average will be one which makes
$$0 = \frac{d}{dr}\int_{\lambda_n}^{\lambda_1} g^2\,d\lambda \tag{46}$$
and
$$\frac{d^2}{dr^2}\int_{\lambda_n}^{\lambda_1} g^2\,d\lambda>0. \tag{47}$$

The integral depends on the relative position of the desired root λ_s in the range from λ_n to λ_1. The subsequent expressions are simplified by introducing l, p, q defined thus as measure of the suspected position s:
$$l = \lambda_1-\lambda_n, \tag{48}$$
$$\lambda_1-s = lp, \tag{49}$$
$$s-\lambda_n = lq; \tag{50}$$
for then each of l, p, q is positive, and
$$p+q = 1. \tag{51}$$

Next
$$\int_{\lambda_n}^{\lambda_1} g^2\,d\lambda = \int_{-lq}^{lp}\left(1-\frac{2y^2}{r^2}+\frac{y^4}{r^4}\right)dy = l-\frac{2}{3}\frac{l^3}{r^2}(p^3+q^3)+\frac{1}{5}\frac{l^5}{r^4}(p^5+q^5). \tag{52}$$

The conditions that this should be a minimum are
$$0 = \tfrac{1}{3}l^3r^{-3}(p^3+q^3)-\tfrac{1}{5}l^5r^{-5}(p^5+q^5), \tag{53}$$
$$-l^3r^{-4}(p^3+q^3)+l^5r^{-6}(p^5+q^5)>0. \tag{54}$$

Although infinite r satisfies the former, it violates the latter, and corresponds to a maximum of the integral. Infinite r must therefore be excluded.

Let the first condition be multiplied by r^5l^{-5} which is finite, and let the second be multiplied by r^6l^{-5} which is finite and positive, so that
$$0 = \tfrac{1}{3}r^2l^{-2}(p^3+q^3)-\tfrac{1}{5}(p^5+q^5), \tag{55}$$
$$-r^2l^{-2}(p^3+q^3)+p^5+q^5>0. \tag{56}$$

Both these conditions are satisfied by

$$\frac{r^2}{l^2} = \frac{3}{5}\left(\frac{p^5+q^5}{p^3+q^3}\right). \tag{57}$$

That is the best choice of r on the average. The corresponding g is

$$g = 1 - \left(\frac{\lambda - s}{\lambda_1 - \lambda_n}\right)^2 \frac{5}{3}\left(\frac{p^3+q^3}{p^5+q^5}\right). \tag{58}$$

For picking out the r suitable to a given p, formula (57) provides the following numerical scale; in subsequent work $r/(\lambda_1 - \lambda_n)$ is called β:

$p =$	0·5	0·55	0·6	0·65	0·7	0·75	0·8	0·85	0·9	0·95	1·0
$\beta =$	0·387	0·400	0·434	0·479	0·526	0·572	0·615	0·657	0·697	0·736	0·775

When $p < 0{\cdot}5$ the r is that appropriate to $1-p$ on the above scale. These satisfy (44) and (45). Figure 2 is of the same type as figure 1 except that only the best choices fixed by (57) are shown; and except that these are shown for three positions of the desired root λ_s, namely, those specified by $p = \frac{1}{2}, \frac{2}{3}, \frac{5}{6}$. The desired root can easily be seen, because it is the abscissa of the maximum of the curve. The necessary condition, $-1 < g < 1$ except at $\lambda = \lambda_s$, is obviously satisfied.

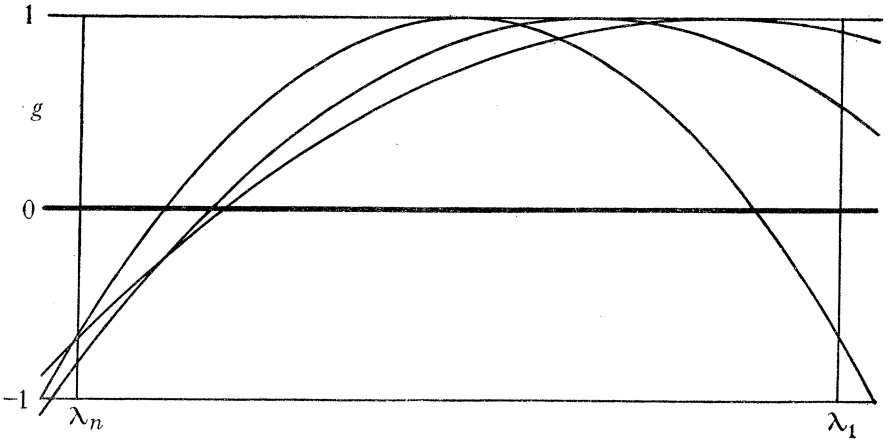

FIGURE 2. Latent roots are points on the abscissa. The ordinate is the ratio of purification. Each parabola corresponds to the multiplication by that $(K-\rho I)/(K-\sigma I)$ which is the best for finding the latent root corresponding to the vertex.

Many purifications

From the figures it is evident that there is a considerable range of fairly good ρ; so that to insist on having quite the best ρ might be like refusing to set foot in a motor-car until assured that it was of the very best make. Nevertheless, those who plan large-scale purifications are sure to wish to know what choice of ρ is the best. One can choose the multiplier already shown to be of the best for a single purification and go on using it repeatedly. To do so is not very wrong, yet it is not the best choice. For the basis of the theorem about the optimum was that the probability of a_ν was uniformly spread over the range $\lambda_n \leqslant \lambda \leqslant \lambda_1$. That was an expression of the computer's initial ignorance. During the progress of his arithmetic he becomes of necessity somewhat knowledgeable. To put it another way, if a uniform probability of a_ν is

a suitable assumption concerning the raw column X^0, it cannot also be suitable for the partially purified column $X^1 = (K-\rho I)\,X^0$, because multiplication by $K-\rho I$ changes any distribution of a_ν. Apart from the requirements of practice, the theory has some rather subtle features of intellectual interest. The best $\rho_1, \rho_2, \rho_3, \ldots, \rho_m$ are found to depend on how far ahead the computer plans. He may not be able to foresee how many factors

$$(K-\rho_1 I)\,(K-\rho_2 I)\,(K-\rho_3 I)\ldots$$

will be needed, and accordingly he may take for ρ_1 the value that is best for a single factor, then for ρ_2 the value that is best after that ρ_1 has been used, and so on. This may be called 'successive adjustment'. On the contrary, the computer may expect that at least m factors will be needed, and accordingly he may at the outset choose $\rho_1, \rho_2, \ldots, \rho_m$ to give the best result when operated jointly. This may be called the 'comprehensive initial plan'. The ρ which are best for one system are not the same as the ρ which are best for the other.

Many purifications for finding the greatest, or least, latent root

For definiteness the statements below all relate to finding the greatest root λ_1. If instead the least root λ_n is desired, all one needs to do is to take the results for λ_1 and rotate the range $\lambda_n \leqslant \lambda \leqslant \lambda_1$ about its midpoint so as to interchange its ends.

As a preliminary it is desirable to get rid of irrelevant considerations. From figure 1 it is evident that the position of a point should be reckoned as a fraction of the range from λ_1 to λ_n. Accordingly, let a co-ordinate x, and a set of constants $\alpha_1, \alpha_2, \ldots, \alpha_j, \ldots, \alpha_m$ be defined by

$$x = \frac{\lambda_1-\lambda}{\lambda_1-\lambda_n}, \tag{59}$$

$$\alpha_j = \frac{\lambda_1-\rho_j}{\lambda_1-\lambda_n}. \tag{60}$$

Beginning with an arbitrary column A, let a set of m multiplications be made so as to obtain a purified column X, thus:

$$X = (K-\rho_1 I)\,(K-\rho_2 I)\ldots(K-\rho_m I)\,A. \tag{61}$$

The following theory is an extension of that already given for $m = 1$. The coefficient of the desired column P_1, in the expansion of A, is the same as in the expansion of

$$X(\lambda_1-\rho_1)^{-1}\,(\lambda_1-\rho_2)^{-1}\ldots(\lambda_1-\rho_m)^{-1}.$$

In the latter quantity the coefficient of any latent column P belonging to any suspected root λ is altered from its coefficient in A in the ratio f_m defined by

$$f_m = \frac{(\lambda-\rho_1)\,(\lambda-\rho_2)\ldots(\lambda-\rho_m)}{(\lambda_1-\rho_1)\,(\lambda_1-\rho_2)\ldots(\lambda_1-\rho_m)}. \tag{62}$$

In view of (59) and (60)

$$f_m = \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\ldots\left(1-\frac{x}{\alpha_m}\right). \tag{63}$$

According to the principle of least squares G^2, defined by

$$G^2 = \frac{1}{\lambda_1-\lambda_n}\int_{\lambda_n}^{\lambda_1} f_m^2\,d\lambda, \tag{64}$$

is to be made a minimum for variations of $\rho_1, \rho_2, \ldots, \rho_m$. But, on introduction of (59),

$$G^2 = \int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2 \left(1-\frac{x}{\alpha_2}\right)^2 \ldots \left(1-\frac{x}{\alpha_m}\right)^2 dx; \tag{65}$$

and in this form G^2 has to be made a minimum for variations of $\alpha_1, \alpha_2, \ldots, \alpha_m$. None of the α vanishes, so that the integrand in (65) has continuous α-derivatives of all orders. The expression (65) will be applied to different plans of operation.

Purifiers adjusted in succession

The needs of the computer who says, with Cardinal Newman,

'I do not ask to see
The distant scene; one step enough for me'

are easy to satisfy, in so far as the minimum occurs for one independent variable at a time. That is to say, in (65) $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}$ are already known from previous steps, and only α_m remains to be determined. From (65),

$$\frac{\partial G^2}{\partial \alpha_m} = \frac{2}{\alpha_m^2}\int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2 \left(1-\frac{x}{\alpha_2}\right)^2 \ldots \left(1-\frac{x}{\alpha_{m-1}}\right)^2 \left(1-\frac{x}{\alpha_m}\right) x\,dx, \tag{66}$$

$$\frac{\partial^2 G^2}{\partial \alpha_m^2} = -\frac{2}{\alpha_m}\frac{\partial G^2}{\partial \alpha_m} + \frac{2}{\alpha_m^4}\int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2 \left(1-\frac{x}{\alpha_2}\right)^2 \ldots \left(1-\frac{x}{\alpha_{m-1}}\right)^2 x^2\,dx. \tag{67}$$

The integrand in (67) is positive, so that:

$$\text{If } \frac{\partial G^2}{\partial \alpha_m} = 0, \quad \text{then } \frac{\partial^2 G^2}{\partial \alpha_m^2} > 0, \quad \text{and so } G^2 \text{ is a minimum.} \tag{68}$$

As none of the α can be zero, it is convenient to clear (66) of fractions. The condition for a minimum of G^2 then becomes

$$0 = \int_0^1 (\alpha_1 - x)^2 (\alpha_2 - x)^2 \ldots (\alpha_{m-1} - x)^2 (\alpha_m - x)\, x\,dx. \tag{69}$$

This can be solved for α_m giving

$$\alpha_m = \frac{\int_0^1 (\alpha_1 - x)^2 (\alpha_2 - x)^2 \ldots (\alpha_{m-1} - x)^2 x^2\,dx}{\int_0^1 (\alpha_1 - x)^2 (\alpha_2 - x)^2 \ldots (\alpha_{m-1} - x)^2 x\,dx}. \tag{70}$$

Formula (70) is to be applied in succession for $m = 1, 2, 3, \ldots$. First

$$\alpha_1 = \int_0^1 x^2\,dx \Big/ \int_0^1 x\,dx = \frac{2}{3}, \quad \text{as already found.} \tag{71}$$

Next
$$\alpha_2 = \int_0^1 (\tfrac{2}{3} - x)^2 x^2\,dx \Big/ \int_0^1 (\tfrac{2}{3} - x)^2 x\,dx = \frac{8}{15} = 0{\cdot}5333\ldots. \tag{72}$$

Then
$$\alpha_3 = \frac{\int_0^1 (\frac{2}{3} - x)^2 (\frac{8}{15} - x)^2 x^2\,dx}{\int_0^1 (\frac{2}{3} - x)^2 (\frac{8}{15} - x)^2 x\,dx} = \frac{802}{1365} = 0{\cdot}5875\ldots. \tag{73}$$

The numbers α_1, α_2, α_3 go down and up again. This peculiar oscillation has been carefully verified.

It is perhaps hardly necessary to pursue further the theme of purifiers adjusted one step ahead; for it will be shown, in table 1 below, that more foresight is advantageous.

Plans for more than one step in advance

It is now necessary to vary several of the purifiers together; the equations to make G^2 stationary are a simultaneous set; and, as to whether G^2 is a minimum, the decision depends on a quadratic form having the second derivatives for its coefficients. As the subsequent general formulae for m steps in advance are rather complicated, the case $m = 2$ may serve as an agreeable introduction.

Two purifiers α_1 and α_2

The α_1-derivative of (65) is

$$\frac{\partial G^2}{\partial \alpha_1} = \frac{2}{\alpha_1^2}\int_0^1 x\left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)^2 dx. \tag{74}$$

This has to vanish, and so the factor $2/\alpha_1^2$ can be omitted. Then the factor $1-x/\alpha_1$ allows the integral to be split into two parts thus:

$$\int_0^1 x\left(1-\frac{x}{\alpha_2}\right)^2 dx = \frac{1}{\alpha_1}\int_0^1 x^2\left(1-\frac{x}{\alpha_2}\right)^2 dx. \tag{75}$$

On working out the integrals it is found that

$$\alpha_1\left(\frac{\alpha_2^2}{2}-\frac{2\alpha_2}{3}+\frac{1}{4}\right) = \frac{\alpha_2^2}{3}-\frac{2\alpha_2}{4}+\frac{1}{5}. \tag{76}$$

Interchange of the suffices gives, as simultaneous with (76),

$$\alpha_2\left(\frac{\alpha_1^2}{2}-\frac{2\alpha_1}{3}+\frac{1}{4}\right) = \frac{\alpha_1^2}{3}-\frac{2\alpha_1}{4}+\frac{1}{5}. \tag{77}$$

There are now two cases. If $\alpha_1 = \alpha_2$, the equations (76) and (77) become identical; each is a cubic. By iterations a root was found to be

$$\alpha_1 = \alpha_2 = 0{\cdot}6227\ldots. \tag{78}$$

In the other case $\alpha_1 \neq \alpha_2$. Then subtraction of (77) from (76), and cancellation of a common factor $\alpha_1-\alpha_2$, leads to

$$3+6\alpha_1\alpha_2 = 4(\alpha_1+\alpha_2). \tag{79}$$

Instead α_1 times (76) minus α_2 times (77) leads to

$$12+20\alpha_1\alpha_2 = 15(\alpha_1+\alpha_2). \tag{80}$$

From (79) and (80) it follows that

$$\alpha_1+\alpha_2 = \tfrac{6}{5}, \quad \alpha_1\alpha_2 = \tfrac{3}{10}. \tag{81}$$

So α_1 and α_2 are the roots of the quadratic $0 = \frac{3}{10}-\frac{6}{5}\alpha+\alpha^2$. Whence

$$\alpha_1 = 0{\cdot}844949\ldots, \quad \alpha_2 = 0{\cdot}355051\ldots. \tag{82}$$

Thus (78) and (82) are two distinct pairs of purifiers, either of which makes G^2 stationary.

It remains to be seen whether either pair makes G^2 a minimum. From (74)

$$\frac{\partial^2 G^2}{\partial \alpha_1^2} = -\frac{2}{\alpha_1}\frac{\partial G^2}{\partial \alpha_1} + \frac{2}{\alpha_1^4}\int_0^1 x^2\left(1-\frac{x}{\alpha_2}\right)^2 dx. \tag{83}$$

In the second member, the first term vanishes if G^2 is stationary, and the integrand in the second term is positive. Therefore:

$$\text{If} \quad \frac{\partial G^2}{\partial \alpha_1} = 0, \quad \text{then} \quad \frac{\partial^2 G^2}{\partial \alpha_1^2} > 0. \tag{84}$$

The corresponding statement for α_2 follows by interchange of suffixes.

The α_2-derivative of (74) is

$$\frac{\partial^2 G^2}{\partial \alpha_2 \partial \alpha_1} = \frac{4}{\alpha_1^2 \alpha_2^2}\int_0^1 x^2\left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right) dx. \tag{85}$$

The integrand in (85) becomes equal to that in (83) when $\alpha_1 = \alpha_2$. Consequently

$$\text{At} \quad \alpha_1 = \alpha_2 = 0{\cdot}6227\ldots, \quad \frac{\partial^2 G^2}{\partial \alpha_1 \partial \alpha_2} = 2\frac{\partial^2 G^2}{\partial \alpha_1^2} = 2\frac{\partial^2 G^2}{\partial \alpha_2^2}. \tag{86}$$

The theory of the maxima and minima of functions of two variables (Goursat 1924, **1**, 107–115) shows that G^2 will be an extreme where it is stationary if

$$\left(\frac{\partial^2 G^2}{\partial \alpha_1 \partial \alpha_2}\right)^2 < \frac{\partial^2 G^2}{\partial \alpha_1^2}\frac{\partial^2 G^2}{\partial \alpha_2^2}, \tag{87}$$

but not if this inequality is reversed. Comparison with (86) shows that G^2 is neither maximum nor minimum at $\alpha_1 = \alpha_2 = 0{\cdot}6227\ldots$. This may be called a saddle-point.

It remains to examine the other pair of purifiers which make G^2 stationary. Here a peculiar relation comes to our aid. From (74) and (85)

$$\begin{aligned}\frac{\alpha_1^2}{2}\frac{\partial G^2}{\partial \alpha_1} - \frac{\alpha_2^2}{2}\frac{\partial G^2}{\partial \alpha_2} &= \int_0^1 x\left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\left\{1-\frac{x}{\alpha_2}-\left(1-\frac{x}{\alpha_1}\right)\right\} dx \\ &= \left(\frac{1}{\alpha_1}-\frac{1}{\alpha_2}\right)\int_0^1 x^2\left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right) dx = \left(\frac{1}{\alpha_1}-\frac{1}{\alpha_2}\right)\frac{\alpha_1^2\alpha_2^2}{4}\frac{\partial^2 G^2}{\partial \alpha_2 \partial \alpha_1}.\end{aligned} \tag{88}$$

That is to say:

$$\text{If} \quad \frac{\partial G^2}{\partial \alpha_1} = 0 \quad \text{and} \quad \frac{\partial G^2}{\partial \alpha_2} = 0, \quad \text{but} \quad \alpha_1 \neq \alpha_2, \quad \text{then} \quad \frac{\partial^2 G^2}{\partial \alpha_2 \partial \alpha_1} = 0. \tag{89}$$

According to (82), the condition (89) is satisfied for $\alpha_1 = 0{\cdot}8449\ldots$, $\alpha_2 = 0{\cdot}3550\ldots$; so that the inequality (87) then occurs, and G^2 is an extreme. Then (84) shows that G^2 is a minimum. *Thus* (82) *specifies the best choice of a pair of purifiers.*

These formulae have been tested by working out G^2 explicitly. From (65)

$$G^2 = 1 - \left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right) + \frac{1}{3}\left(\frac{1}{\alpha_1^2}+\frac{4}{\alpha_1\alpha_2}+\frac{1}{\alpha_2^2}\right) - \frac{1}{2\alpha_1\alpha_2}\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right) + \frac{1}{5\alpha_1^2\alpha_2^2}. \tag{90}$$

From (90), G^2 has been computed for an assortment of values of α_1 and α_2 with the results shown in table 1. These numbers support the theory. The plan for successive adjustment is not quite so good as the plan for two steps ahead. Yet the range of G^2 between the saddle-

point and the minimum is not large enough to matter much in practical purifying operations. The last column shows that too much separation does harm. The minimum value of G, the root-mean-square ratio of purification, is $\frac{1}{3}$ exactly.

TABLE 1

	saddle-point	successive adjustment			minimum	
$\alpha_1 =$	0·6228	$\frac{2}{3}$	0·75	0·85	0·8499	1·0
$\alpha_2 =$	0·6228	$\frac{8}{15}$	0·5	0·4	0·3551	0·25
$G^2 =$	0·135	0·133	0·126	0·113	0·1111	0·200

The general case of m purifiers $\alpha_1, \alpha_2, \ldots, \alpha_m$

The general theory has been carried as far as equation (65), which is now our starting-point. The usual procedure as to minima is to find first the co-ordinates which make the function stationary, and afterwards to inquire whether it is a minimum there. In the present problem it is easier to proceed in the opposite sense, by discussing first whether a stationary G^2, if it could be found, would be a minimum or not. In the case $m = 2$, the more obvious stationary G^2 was found to be neither minimum nor maximum.

On variation of α_j alone, if j is neither 1, 2, nor m,

$$\frac{\partial G^2}{\partial \alpha_j} = \frac{2}{\alpha_j^2}\int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2\left(1-\frac{x}{\alpha_2}\right)^2 \ldots \left(1-\frac{x}{\alpha_j}\right) x \ldots \left(1-\frac{x}{\alpha_m}\right)^2 dx. \tag{91}$$

It is tempting for brevity and generality to write this as

$$\frac{\partial G^2}{\partial \alpha_j} = \frac{2}{\alpha_j^2}\int_0^1 \frac{x f_m^2\, dx}{1-x/\alpha_j} \qquad (j = 1, 2, \ldots, m); \tag{92}$$

but, if the form (92) is used, one must bear in mind that the integrand is a continuous function of x, as is plainly shown by (91); because in (92) the denominator $1-x/\alpha_j$, which may become zero, cancels with a factor of f_m^2. Similar remarks apply to subsequent integrals. From (91)

$$\begin{aligned}\frac{\partial^2 G^2}{\partial \alpha_j^2} &= -\frac{2}{\alpha_j}\frac{\partial G^2}{\partial \alpha_j} + \frac{2}{\alpha_j^4}\int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2\left(1-\frac{x}{\alpha_2}\right)^2 \ldots x^2 \ldots \left(1-\frac{x}{\alpha_m}\right)^2 dx \\ &= -\frac{2}{\alpha_j}\frac{\partial G^2}{\partial \alpha_j} + \frac{2}{\alpha_j^4}\int_0^1 \frac{x^2 f_m^2\, dx}{(1-x/\alpha_j)^2} \qquad (j = 1, 2, \ldots, m). \end{aligned} \tag{93}$$

In this the first term vanishes if G^2 is stationary. The integrand in the second term is positive, so that:

$$\text{If} \quad \frac{\partial G^2}{\partial \alpha_j} = 0, \quad \text{then} \quad \frac{\partial^2 G^2}{\partial \alpha_j^2} > 0 \qquad (j = 1, 2, \ldots, m). \tag{94}$$

Again, from (91)

$$\begin{aligned}\frac{\partial^2 G^2}{\partial \alpha_j \partial \alpha_k} &= \frac{4}{\alpha_j^2 \alpha_k^2}\int_0^1 \left(1-\frac{x}{\alpha_1}\right)^2\left(1-\frac{x}{\alpha_2}\right)^2 \ldots \left(1-\frac{x}{\alpha_j}\right) x \left(1-\frac{x}{\alpha_k}\right) x \ldots \left(1-\frac{x}{\alpha_m}\right)^2 dx \\ &= \frac{4}{\alpha_j^2 \alpha_k^2}\int_0^1 \frac{x^2 f_m^2\, dx}{(1-x/\alpha_j)\,(1-x/\alpha_k)} \qquad (j, k = 1, 2, \ldots, m). \end{aligned} \tag{95}$$

The sign of the integrand in (95) is not obvious. However, there is a remarkably simple relation between this mutual second derivative and a weighted difference of the corresponding first derivatives. Thus from (92) and (95)

$$\frac{\alpha_j^2}{2}\frac{\partial G^2}{\partial \alpha_j}-\frac{\alpha_k^2}{2}\frac{\partial G^2}{\partial \alpha_k}=\int_0^1 x f_m^2\left\{\frac{1}{1-x/\alpha_j}-\frac{1}{1-x/\alpha_k}\right\}dx$$

$$=\left(\frac{1}{\alpha_j}-\frac{1}{\alpha_k}\right)\int_0^1\frac{x^2 f_m^2\,dx}{(1-x/\alpha_j)(1-x/\alpha_k)}=\left(\frac{1}{\alpha_j}-\frac{1}{\alpha_k}\right)\frac{\alpha_j^2\alpha_k^2}{4}\frac{\partial^2 G^2}{\partial\alpha_j\partial\alpha_k}\quad(j,k=1,2,\ldots,m).\tag{96}$$

Whence it follows that:

$$\text{If}\quad\frac{\partial G^2}{\partial\alpha_j}=0\quad\text{and}\quad\frac{\partial G^2}{\partial\alpha_k}=0,\quad\text{but}\quad\alpha_j\neq\alpha_k,\quad\text{then must}\quad\frac{\partial^2 G^2}{\partial\alpha_j\partial\alpha_k}=0.\tag{97}$$

This fact makes a most welcome simplification of the theory. For the decision whether G^2 is a minimum, or not, depends on the expansion of G^2 by Taylor's theorem in the variations, $\delta\alpha_j$, of the m parameters. When the first derivatives vanish, one expects in general to have to study the quadratic form in m variables involving the second derivatives; and that may be bothersome. But here the quadratic form is already reduced to a sum of squares. Moreover, by (94), each of the squares is positive. It follows that:

If $\alpha_1,\alpha_2,\ldots,\alpha_m$, no two of which are equal, make G^2 stationary, they make it a minimum. (98)

It will next be shown that G^2 cannot be a minimum if any two of the α are equal. For if α_j is put equal to α_k *after* the differentiations, then the integrands in (93) and (95) become equal, so that:

$$\text{If}\quad\frac{\partial G^2}{\partial\alpha_j}=0\quad\text{and}\quad\frac{\partial G^2}{\partial\alpha_k}=0\quad\text{and}\quad\alpha_j=\alpha_k,\quad\text{then}\quad\frac{\partial^2 G^2}{\partial\alpha_j\partial\alpha_k}=2\frac{\partial^2 G^2}{\partial\alpha_j^2}=2\frac{\partial^2 G^2}{\partial\alpha_k^2}.\tag{99}$$

Let $\delta\alpha$ denote a variation of α from the value which makes G^2 stationary. It suffices to consider $\delta\alpha_j$ and $\delta\alpha_k$, all the other $\delta\alpha$ being zero. For if G^2 is not a minimum in this special case, G^2 cannot be a minimum for all possible $\delta\alpha$. The second-order terms in the Taylor expansion of G^2 are

$$\frac{\partial^2 G^2}{\partial\alpha_j^2}(\delta\alpha_j)^2+2\frac{\partial^2 G^2}{\partial\alpha_j\partial\alpha_k}\delta\alpha_j\delta\alpha_k+\frac{\partial^2 G^2}{\partial\alpha_k^2}(\delta\alpha_k)^2,\tag{100}$$

and under the conditions (99) these simplify to

$$\frac{\partial^2 G^2}{\partial\alpha_j^2}\{(\delta\alpha_j)^2+4\delta\alpha_j\delta\alpha_k+(\delta\alpha_k)^2\}.\tag{101}$$

The quadratic form in (101) is capable of either sign because $4^2-4>0$. Therefore:

G^2 cannot be a minimum for all possible variations of $\alpha_1,\alpha_2,\ldots,\alpha_m$ in the neighbourhood of $\alpha_j=\alpha_k$. (102)

In other words there is some advantage to be gained by a suitable scatter of the purifiers. The late A. E. H. Love, F.R.S., pointed this advantage out to me in a personal letter about 40 years ago (Richardson 1910, p. 321 footnote); but I do not remember that he gave any reason for the suggestion. My own opinion wandered long among belief, doubt, and error, until in June 1949 I made the theorems (96) and (99), which settle the question in favour of scattering.

It remains to find the unequal values of $\alpha_1, \alpha_2, \ldots, \alpha_m$ which make G^2 a minimum. In view of (92) the α are determined by the set of m simultaneous equations

$$0 = \int_0^1 \frac{x f_m^2 dx}{1 - x/\alpha_j} \qquad (j = 1, 2, \ldots, m). \tag{103}$$

The difficulty of solving them can be decreased by taking differences, either simple or weighted. The highest common factor of all the integrands is xf_m, where f_m is given by (63). The other factors in any integrand are $m-1$ brackets of the type $(1-x/\alpha_j)$. Let two of the integrals be selected which agree except as to one bracket in each integrand, and take their simple difference. For example, m being 4,

$$0 = \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\left(1-\frac{x}{\alpha_3}\right) x f_4\, dx - \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\left(1-\frac{x}{\alpha_4}\right) x f_4\, dx.$$

Therefore, as $\alpha_3 \neq \alpha_4$,

$$0 = \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right) x^2 f_4\, dx, \tag{104}$$

which contains one bracket less, but an extra factor x. Instead let each of the given integrals be weighted by the α corresponding to its peculiar bracket. Thus the last example, when so modified, becomes

$$0 = \alpha_3 \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\left(1-\frac{x}{\alpha_3}\right) x f_4\, dx - \alpha_4 \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\left(1-\frac{x}{\alpha_4}\right) x f_4\, dx;$$

and therefore, as $\alpha_3 \neq \alpha_4$,

$$0 = \int_0^1 \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right) x f_4\, dx, \tag{105}$$

which agrees with the previous result in the loss of a bracket, but differs from it in that the power of x is not augmented. Such a reduction of the number of brackets can be continued by successive differences, and at each stage the exponent of x can either be increased by unity, or left unchanged, according as the difference is plain or suitably weighted. Thus finally all the brackets, except those in the highest common factor xf_m, are done away; and the exponent of x is increased by any positive integer not exceeding $m-1$. The set of simultaneous equations (103) is replaced by the set

$$0 = \int_0^1 x f_m\, dx, \quad 0 = \int_0^1 x^2 f_m\, dx, \quad \ldots, \quad 0 = \int_0^1 x^{m-1} f_m\, dx, \quad 0 = \int_0^1 x^m f_m\, dx, \tag{106}$$

and this is much more manageable. In order to perform the integrations, let f_m be imagined as multiplied out, so that

$$f_m = \left(1-\frac{x}{\alpha_1}\right)\left(1-\frac{x}{\alpha_2}\right)\ldots\left(1-\frac{x}{\alpha_m}\right) = 1 - b_1 x + b_2 x^2 + \ldots + b_m (-x)^m. \tag{107}$$

The b are symmetric functions of the α, but there is no need to state the relations explicitly. The further working out will be illustrated by the case $m = 3$.

The integrations in (106) lead to

$$\left.\begin{aligned} 0 &= \tfrac{1}{2} - \tfrac{1}{3} b_1 + \tfrac{1}{4} b_2 - \tfrac{1}{5} b_3, \\ 0 &= \tfrac{1}{3} - \tfrac{1}{4} b_1 + \tfrac{1}{5} b_2 - \tfrac{1}{6} b_3, \\ 0 &= \tfrac{1}{4} - \tfrac{1}{5} b_1 + \tfrac{1}{6} b_2 - \tfrac{1}{7} b_3. \end{aligned}\right\} \tag{108}$$

These linear simultaneous equations have a distinctive pattern which would extend obviously to any greater m. They are satisfied by

$$b_1 = \tfrac{15}{2}, \quad b_2 = 15, \quad b_3 = \tfrac{35}{4}. \tag{109}$$

Therefore the ratio of purification, f_m, specializes to

$$f_3 = 1 - \tfrac{15}{2}x + 15x^2 - \tfrac{35}{4}x^3, \tag{110}$$

when the α are chosen in the best manner. But by (107), f_m vanishes when x becomes equal to any one of the α. So $\alpha_1, \alpha_2, \alpha_3$ were found as the roots of the cubic $f_3 = 0$. As it was known that the roots were real and distinct, the trigonometrical method of solution was used (Hobson 1921, p. 153). It gave

$$\alpha_1 = 0{\cdot}9119, \quad \alpha_2 = 0{\cdot}5905, \quad \alpha_3 = 0{\cdot}2123. \tag{111}$$

In like manner it was found that the best choice of a set of four purifiers was that given by the roots of the quartic, obtained by putting $f_4 = 0$ in

$$f_4 = 1 - 12x + 42x^2 - 56x^3 + \tfrac{126}{5}x^4. \tag{112}$$

The quartic was solved by Horner's method; and its roots were found to be

$$\alpha_1 = 0{\cdot}9429, \quad \alpha_2 = 0{\cdot}7232, \quad \alpha_3 = 0{\cdot}4164, \quad \alpha_4 = 0{\cdot}1398. \tag{113}$$

Many purifications for finding an intermediate latent root

The desired root is suspected to lie near s such that $\lambda_n < s < \lambda_1$. It has already been shown that, when an intermediate root is sought, the premultiplier $K - \rho I$ needs to be replaced by $(K - \rho I)(K - \sigma I)$, where

$$\rho = s - r, \quad \sigma = s + r. \tag{114}$$

The problem is to choose the best value of r so as to make a rapid improvement in the estimate of the desired root. The formulae differ in detail from those appropriate to the search for the greatest or least latent root. For conciseness one might yearn after generalized formulae that would include both kinds; but for clarity they are better stated separately. Yet similar questions arise; and some parts of the argument are settled by mere allusion to the preceding theory. Various related notions are held together by adjacent letters: f and g, G and H, α and β, b and c. Suppose that the computer plans to use m such pairs ρ, σ in a combined operation leading from an arbitrary column A to a partially purified column X thus:

$$X = (K - \rho_1 I)(K - \sigma_1 I)(K - \rho_2 I)(K - \sigma_2 I) \ldots (K - \rho_m I)(K - \sigma_m I) A. \tag{115}$$

As before in equation (15),

$$A = \sum_{\nu=1}^{\nu=n} a_\nu P_\nu.$$

Also as before let us compare A, not with X, but with X divided by

$$(s - \rho_1)(s - \sigma_1)(s - \rho_2)(s - \sigma_2) \ldots (s - \rho_m)(s - \sigma_m).$$

The coefficient of the latent column belonging to any latent root λ is altered in the ratio g_m, say, where

$$g_m = \frac{(\lambda - \rho_1)(\lambda - \sigma_1)(\lambda - \rho_2)(\lambda - \sigma_2) \ldots (\lambda - \rho_m)(\lambda - \sigma_m)}{(s - \rho_1)(s - \sigma_1)(s - \rho_2)(s - \sigma_2) \ldots (s - \rho_m)(s - \sigma_m)}. \tag{116}$$

Here g_m, like f_m, may be called the ratio of purification. At $\lambda = s$, $g_m = 1$. The premultiplier $(K - \rho_1 I)(K - \sigma_1 I) \ldots (K - \rho_m I)(K - \sigma_m I)$ will be a reliable cleaner provided that $g_m^2 < 1$ for

all λ in the range $\lambda_n \leqslant \lambda \leqslant \lambda_1$ except s. Keeping that necessary but simple condition in reserve, let us go on to inquire what choice of $r_1, r_2, \ldots, r_m$ will be the best in the sense of making H^2, defined by

$$H^2 = \frac{1}{\lambda_1 - \lambda_n} \int_{\lambda_n}^{\lambda_1} g_m^2 \, d\lambda, \tag{117}$$

a minimum for variations of $r_1, r_2, \ldots, r_m$. H, like G, may be called the root-mean-square ratio of purification. The formulae are simplified by shifting the origin to s, and by taking $\lambda_1 - \lambda_n$ as the unit of length. That is to say let

$$x = (\lambda - s)/(\lambda_1 - \lambda_n), \tag{118}$$

$$\beta_j = r_j/(\lambda_1 - \lambda_n). \tag{119}$$

This x differs from the x of the previous section merely by a shift of origin from λ_1 to s. Then in view of (114)

$$(\lambda - \rho_j)(\lambda - \sigma_j) = (\lambda_1 - \lambda_n)^2 (x^2 - \beta_j^2);$$

and so from (116)

$$g_m = (1 - x^2/\beta_1^2)(1 - x^2/\beta_2^2) \ldots (1 - x^2/\beta_{m-1}^2)(1 - x^2/\beta_m^2). \tag{120}$$

As $\partial/\partial r_j = (\lambda_1 - \lambda_n)\, \partial/\partial \beta_j$ the minimum of H^2 with respect to $r_1, r_2, \ldots, r_m$ is also the minimum of H^2 with respect to $\beta_1, \beta_2, \ldots, \beta_m$. To simplify the termini of the integral, let p and q be defined by

$$p = (\lambda_1 - s)/(\lambda_1 - \lambda_n), \quad q = (s - \lambda_n)/(\lambda_1 - \lambda_n). \tag{121}$$

It follows that

$$p \geqslant 0, \quad q \geqslant 0 \quad \text{and} \quad p + q = 1. \tag{122}$$

Equation (117) now becomes

$$H^2 = \int_{-q} g_m^2 \, dx. \tag{123}$$

Its first derivatives are

$$\frac{\partial H^2}{\partial \beta_j} = \frac{4}{\beta_j^3} \int_q \frac{x^2 g_m^2 \, dx}{1 - x^2/\beta_j^2} \qquad (j = 1, 2, \ldots, m). \tag{124}$$

For brevity the integrand is written not in its lowest terms; the denominator cancels with a factor in g_m^2, so that an infinity does not occur. The same remark applies to many subsequent expressions.

The second derivatives with respect to the same β_j are

$$\frac{\partial^2 H^2}{\partial \beta_j^2} = -\frac{3}{\beta_j} \frac{\partial H^2}{\partial \beta_i} + \frac{8}{\beta_j^6} \int_{-q}^{p} \frac{x^4 g_m^2 \, dx}{(1 - x^2/\beta_j^2)^2} \qquad (j = 1, 2, \ldots, m). \tag{125}$$

So if $\partial H^2/\partial \beta_j = 0$, then

$$\partial^2 H^2/\partial \beta_j^2 > 0. \tag{126}$$

The mutual second derivatives are

$$\frac{\partial^2 H^2}{\partial \beta_j \partial \beta_k} = \frac{16}{\beta_j^3 \beta_k^3} \int_{-q}^{p} \frac{x^4 g_m^2 \, dx}{(1 - x^2/\beta_j^2)(1 - x^2/\beta_k^2)} \qquad (j, k = 1, 2, \ldots, m; j \neq k). \tag{127}$$

Comparison of (127) with (125) shows that:

$$\text{If} \quad \partial H^2/\partial \beta_j = 0 \quad \text{and} \quad \partial H^2/\partial \beta_k = 0 \quad \text{and} \quad \beta_j = \beta_k, \quad \text{then} \quad \frac{\partial^2 H^2}{\partial \beta_j \partial \beta_k} = 2 \frac{\partial^2 H^2}{\partial \beta_j^2} = 2 \frac{\partial^2 H^2}{\partial \beta_k^2}. \tag{128}$$

Hence it follows, by an argument which is the same as that applied to equation (99), that

H^2 cannot be a minimum where any two or more of the β are equal. (129)

On the contrary, from (124)

$$\frac{\beta_j^3}{4}\frac{\partial H^2}{\partial \beta_j}-\frac{\beta_k^3}{4}\frac{\partial H^2}{\partial \beta_k}=\int_{-q}^{p}x^2g_m^2\left\{\frac{1}{1-x^2/\beta_j^2}-\frac{1}{1-x^2/\beta_k^2}\right\}dx$$

$$=\left(\frac{1}{\beta_j^2}-\frac{1}{\beta_k^2}\right)\int_{-q}^{p}\frac{x^4g_m^2\,dx}{(1-x^2/\beta_j^2)\,(1-x^2/\beta_k^2)}. \qquad (130)$$

By comparison with (127) it follows that:

$$\text{If}\quad \frac{\partial H^2}{\partial \beta_j}=0\quad\text{and}\quad\frac{\partial H^2}{\partial \beta_k}=0,\quad\text{but}\quad\beta_j\neq\beta_k,\quad\text{then}\quad\frac{\partial^2 H^2}{\partial\beta_j\,\partial\beta_k}=0. \qquad (131)$$

Then by the same argument as that applied to equation (97) it follows that:

If $\beta_1,\beta_2,\ldots,\beta_m$, no two of which are equal, make H^2 stationary, they make it a minimum. (132)

To find the best β it is therefore necessary to solve the m simultaneous equations obtainable by putting $\partial H^2/\partial\beta_j=0$ in (124). Because no one of the β vanishes, a factor can be omitted leaving

$$0=\int_{-q}^{p}\frac{x^2g_m^2\,dx}{1-x^2/\beta_j^2}\qquad(j=1,2,\ldots,m). \qquad (133)$$

These can be simplified by a process like that applied to equations (103); except that x^2 now replaces x. The result, which resembles (106), is

$$0=\int_{-q}^{p}x^2g_m\,dx,\quad 0=\int_{-q}^{p}x^4g_m\,dx,\quad\ldots,\quad 0=\int_{-q}^{p}x^{2m}g_m\,dx. \qquad (134)$$

To facilitate the integrations, let coefficients c be defined by

$$g_m=(1-x^2/\beta_1^2)\,(1-x^2/\beta_2^2)\ldots(1-x^2/\beta_m^2)=1-c_2x^2+c_4x^4-c_6x^6+\ldots+c_{2m}(-x^2)^m. \qquad (135)$$

The c are thus all positive. The subsequent working-out is a modification of that already explained for an extreme latent root. The integrals (134), when evaluated, furnish a set of m linear simultaneous equations, which are solved for the c. These are substituted into the polynomial (135) in x^2. The required $\beta_1^2,\beta_2^2,\ldots,\beta_m^2$ are the values of x^2 which make g_m vanish. Difficulties accumulate as m increases. Let us consider in turn $m=1,2,3$.

For $m=1$:

$$0=\int_{-q}^{p}x^2g_1\,dx=\int_{-q}^{p}(x^2-c_2x^4)\,dx=\frac{p^3+q^3}{3}-c_2\frac{p^5+q^5}{5}.$$

And so

$$\beta_1^2=\frac{1}{c_2}=\frac{3}{5}\frac{p^5+q^5}{p^3+q^3}. \qquad (136)$$

This is merely a confirmation; for the same result was obtained in (57) by simpler methods.

For $m=2$: $\quad 0=\int_{-q}^{p}(x^2-c_2x^4+c_4x^6)\,dx,\quad 0=\int_{-q}^{p}(x^4-c_2x^6+c_4x^8)\,dx.$

As an abbreviation, introduce t defined thus:

$$t_{2k+1}=\int_{-q}^{p}x^{2k}\,dx=\frac{p^{2k+1}+q^{2k+1}}{2k+1}. \qquad (137)$$

The t are regarded as known, for they can be computed from the suspected position of the desired latent root. The simultaneous equations then become

$$0 = t_3 - c_2 t_5 + c_4 t_7, \quad 0 = t_5 - c_2 t_7 + c_4 t_9. \tag{138}$$

When these are solved for c_2, c_4 and the result substituted in (135), the latter becomes

$$g_2 = 1 - x^2 \frac{t_3 t_9 - t_5 t_7}{t_5 t_9 - t_7^2} + x^4 \frac{t_3 t_7 - t_5^2}{t_5 t_9 - t_7^2}, \tag{139}$$

and β_1^2, β_2^2 are the values of x^2 which satisfy the quadratic $g_2 = 0$. Numerical solutions are given in table 2 for five values of p. The change of variable from x to $y = x/p$ allows one to prove that in figure 3 the tangent at $p = 1$ cuts the curve at $p = \frac{1}{2}$ where the ordinate is half that at $p = 1$. For $m = 3$ the algebra becomes complicated; so that it is best to insert the arithmetical value of p into the linear simultaneous equations before solving them. Some results are given in table 3.

TABLE 2. FOR $m=2$

$p=$	$\frac{1}{2}$	$\frac{11}{20}$	$\frac{2}{3}$	$\frac{5}{6}$	1
$\pm\beta_2=$	0·4531	0·4868	0·6010	0·7540	0·9062
$\pm\beta_1=$	0·2692	0·2894	0·3266	0·4401	0·5385

TABLE 3. FOR $m=3$

$p=$	$\frac{1}{2}$	$\frac{11}{20}$	$\frac{2}{3}$	$\frac{5}{6}$	1
$\pm\beta_1=$	0·2029	0·2182	0·2523	0·3358	0·4058
$\pm\beta_2=$	0·3708	0·3964	0·4871	0·6235	0·7415
$\pm\beta_3=$	0·4746	0·5193	0·6316	0·7921	0·9491

Figure 3 shows all these β. Usefully good sets of purifiers can be selected by plotting g_m against λ for various trial values of β. In a former paper (Richardson 1910, figure 3) I did this in effect, but in a different notation, λ^2 there corresponding to λ here. For $m = 3$ and $p = \frac{1}{2}$ the set was $\pm\beta = 0{\cdot}2, 0{\cdot}4, 0{\cdot}5$. This is now seen to be not quite the best triplet.

Summary for the computer on the choice of cleaning operations when the given matrix K is symmetrical

The best procedure is stated here with unnecessary accuracy; for the minima which fix the best choices are only slightly curved. After attending to what is recommended below, you will usually do well to round off $\rho_1, \rho_2, \ldots$ to one or two significant digits, because the time saved by working with few digits is likely to be more than the time wasted by using imperfect purifiers. It is supposed that the greatest latent root λ_1 and the least λ_n are both roughly known. A raw column A, which may be $\{1, 1, 1, \ldots, 1\}$, or something else not specially appropriate to the given matrix K, is to be purified by forming

$$(K - \rho_1 I)(K - \rho_2 I) \ldots (K - \rho_m I) A. \tag{140}$$

The question is what values to choose for $\rho_1, \rho_2, \ldots, \rho_m$. There are two main cases according as the desired latent root is an extreme root, or an intermediate root.

To find the greatest root λ_1 more accurately. Let any one of $\rho_1, \rho_2, \ldots, \rho_m$ be denoted by ρ_j. The permanent feature of the situation is a constant α_j which specifies ρ_j; thus

$$\rho_j = \lambda_1 - \alpha_j(\lambda_1 - \lambda_n), \tag{141}$$

in which λ_1 and λ_n stand for such approximations as happen to be available at this stage.

The best choice of ρ_j depends on how many premultipliers like $(K-\rho_j I)$ you are going to use. A wrong forecast of this number will not lead to a wrong latent root, but it may slightly delay your arrival.

If only one premultiplier $(K-\rho_j I)$ is to be used, take $\alpha = \frac{2}{3}$.

If two, take $\alpha = 0{\cdot}3551,\ 0{\cdot}8449$.

If three, take $\alpha = 0{\cdot}2123,\ 0{\cdot}5905,\ 0{\cdot}9119$.

If four, take $\alpha = 0{\cdot}1398,\ 0{\cdot}4164,\ 0{\cdot}7232,\ 0{\cdot}9429$.

The calculation of the best set of five purifiers would be very tedious, and I have not done it. However, a graph showed (Richardson 1910, figure 1) that the following set of seven, namely,

$$\alpha = 0{\cdot}075,\quad 0{\cdot}2,\quad 0{\cdot}375,\quad 0{\cdot}55,\quad 0{\cdot}7,\quad 0{\cdot}85,\quad 1{\cdot}0$$

is usefully good, though it is probably not the best septuplet. Each of these sets is designed to be used as a whole. For example, to pick out from the quadruplet $\alpha = 0{\cdot}1398$, and to use it alone, would have a disastrous dirtying effect towards the end of the range of λ away from the desired latent root. There is no objection to using the whole sets repeatedly, and in any order; except that, for example, four operations with the best single α, or two operations with the best paired α, are not quite so efficient as one operation with the best set of four α.

To find the least latent root λ_n more accurately. Take

$$\rho_j = \lambda_n + \alpha_j(\lambda_1 - \lambda_n) \tag{142}$$

with the same α as for finding λ_1.

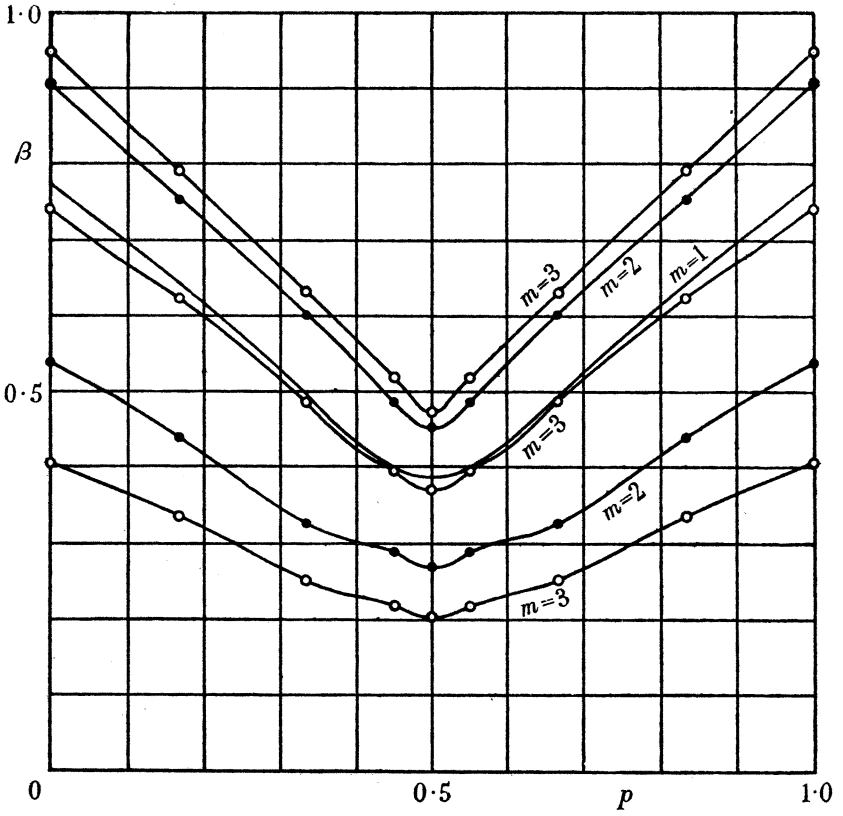

FIGURE 3. The best constants for finding intermediate latent roots.

To find an intermediate latent root which is suspected to be near $\lambda = s$, such that $\lambda_n < s < \lambda_1$. The linear premultipliers $K-\rho_j I$ have now to be used in pairs such as $(K-\rho_j I)(K-\sigma_j I)$, in which

$$\rho_j = s-\beta_j(\lambda_1-\lambda_n) \quad \text{and} \quad \sigma_j = s+\beta_j(\lambda_1-\lambda_n), \tag{143}$$

so that

$$(K-\rho_j I)(K-\sigma_j I) = (K-sI)^2-\beta_j^2(\lambda_1-\lambda_n)^2 I. \tag{144}$$

The question is about the choice of β; and β depends on the suspected position s of the desired root in the range from λ_n to λ_1. As a measure of this position, calculate p from

$$p = \frac{\lambda_1-s}{\lambda_1-\lambda_n}. \tag{145}$$

Then look at figure 3, select the abscissa p, and read the corresponding ordinate β on a curve. Several curves are drawn, because the best choice depends on how many quadratic premultipliers of the type (144) you are going to use:

If only one, then read β on the curve marked $m = 1$.

If two, then read β on the two curves marked $m = 2$.

If three, then read β on the three curves marked $m = 3$.

Each of these sets is designed to be used as a whole.

4. Tests of achievement, reconsidered more systematically

The division test (alias the linear annihilator)

The computer may arrive by various processes, some of which have already been illustrated, at a column P such that $KP = \lambda P$. The test which proves that P is of this form is to divide each element of KP by the corresponding element of P and to show that the n values of λ are equal. They are written $KP \div P$. This operation is one of the possible tests of achievement. It is also a very useful guide in the imperfect stages of computing, when instead of the latent column P the computer has only an approximation X thereto, and the n estimates of λ are not quite equal. Many authors have used this test; but here tests of achievement must be given special emphasis, because theorems about convergence are not used in support.

Test by simultaneous equations in two unknowns, alias the quadratic annihilator

The computer may not succeed in arriving at a column P which satisfies $KP = \lambda P$, but he may arrive at a column Q which satisfies

$$K^2Q+gKQ+hQ = \mathbf{0}. \tag{146}$$

This happens in a subsequent numerical example. The proof that he has done so is to write out the last statement as a set of n simultaneous equations for g and h, and to solve them in pairs. If each pair leads to the same values of g and h the test is passed. Just as for the division test, approximate agreement is a very useful guide to further operations. Next a proof is needed that for any two latent roots, λ_1 and λ_2, there always exists a column, C say, such that $(K-\lambda_1 I)(K-\lambda_2 I)C = 0$. For this purpose let us return to the Cayley-Hamilton theorem in the form (4) of §1, because it is general for any sort of latent roots. From an arbitrary column A let the associates of all the other latent roots be removed by forming C defined thus:

$$C = (K-\lambda_3 I)(K-\lambda_4 I)\dots(K-\lambda_n I)A, \tag{147}$$

to $(n-2)$ brackets.

Then from (4) of § 1 it follows that

$$(K-\lambda_1 I)\,(K-\lambda_2 I)\,C = \mathbf{0}. \tag{148}$$

We may now identify C with Q and $(K-\lambda_1 I)\,(K-\lambda_2 I)$ with $K^2+gK+hI$. It follows by solving a quadratic equation that

$$\lambda_1,\ \lambda_2 = \tfrac{1}{2}\{-g \pm \sqrt{(g^2-4h)}\}. \tag{149}$$

Then if $g^2<4h$ the latent roots λ_1, λ_2 are complex. If $g^2 = 4h$, the latent roots are both equal to $-\frac{1}{2}g$; and if, moreover, $(K+\frac{1}{2}gI)\,C$ does not vanish then these roots are associated with a quadratic 'elementary divisor' in Weierstrass's nomenclature.

Example 10. $K = \begin{bmatrix} 4 & 1 \\ -1 & 2 \end{bmatrix}$ has latent roots $\lambda_1 = \lambda_2 = 3$, as may be shown by solving the quadratic for λ. Now take an arbitrary column $\{a, b\}$ and premultiply it by $K-3I$ thus:

$$(K-3I)\,\{a,b\} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ -a & -b \end{bmatrix} \quad \text{not} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

So $K-3I$ is not an annihilating multiplier for an arbitrary column. However, on repeating the operation,

$$(K-3I)^2\,\{a,b\} = (K-3I)\begin{bmatrix} a+b \\ -a & -b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

So there is a quadratic elementary divisor.

Example 11.
$$K = \begin{bmatrix} 6 & -3 & 4 & 1 \\ 4 & 2 & 4 & 0 \\ 4 & -2 & 3 & 1 \\ 4 & 2 & 3 & 1 \end{bmatrix}.$$

This is taken from Aitken (1937), who stated that 'The latent roots are actually

$$\lambda_1 = \lambda_2 = 5{\cdot}236068, \quad \lambda_3 = \lambda_4 = 0{\cdot}763932,$$

and these equal pairs are each associated with a quadratic elementary divisor'. It follows that $(K-\lambda_1 I)\,(K-\lambda_3 I) = K^2-6K+4I$. This includes both roots. Yet

$$(K^2-6K+4I)\,\{1,1,1,1\} = \{8,20,8,24\}, \quad \text{not} \quad \{0,0,0,0\}.$$

On repeating the multiplication, however,

$$(K^2-6K+4I)\,\{8,20,8,24\} = \{0,0,0,0\}.$$

Let us now return to the quadratic annihilator in general.

If $g^2>4h$, the latent roots λ_1 and λ_2 are real and distinct; but in that case the simpler division test would have sufficed, and the computer would be left wondering how he had missed it. It has been shown in § 3 that the division test also suffices for the equal latent roots of symmetrical matrices.

Tests by simultaneous equations in more than two unknowns corresponding to annihilators of degree higher than the second

If λ_1, λ_2, λ_3 are any three latent roots, it can be proved in a similar manner that there always exists a column D such that

$$(K-\lambda_1 I)\,(K-\lambda_2 I)\,(K-\lambda_3 I)\,D = \mathbf{0}. \tag{150}$$

If the computer cannot find a column P such that $KP-\lambda P=\mathbf{0}$, nor a column Q such that $K^2Q+gKQ+hQ=\mathbf{0}$, he may perhaps find a column R such that

$$K^3R+fK^2R+gKR+hR=\mathbf{0}. \tag{151}$$

If so, it is possible from f, g, h to find $\lambda_1, \lambda_2, \lambda_3$. And so on to higher degrees, but naturally with increasing difficulties.

5. An example on non-linear elementary divisors

This example is no. 11 reconsidered. Aitken (1937, p. 273) premultiplied $\{1,1,1,1\}$ by K six times and remarked on the slowness of convergence. He mentioned (p. 280) that $\lambda_1=5{\cdot}236068$ was found by evaluating the product of the k greatest roots. Knowing λ_1 he proceeded, by his method of λ-differences, to show that this root is double. The same process yielded one of the latent columns P_1. Then λ_3 and P_3 were found in a similar manner (pp. 300–301). Aitken's treatment illustrates several of his ingenious devices of wider application. His arithmetic is published only in brief abstract; so I do not know whether it is less or more laborious than that of which a summary is given below. The given matrix would, I expect, be troublesome by any iterative method. One might be tempted to dismiss this example by saying that two pairs of equal roots are not likely to occur in statistical practice. Nevertheless, I think that this K is worthy of attention because its troublesomeness is specially instructive.

Let us now see how a computer, who thinks in terms of purification, may reasonably be supposed to tackle this matrix. He notices that K is unsymmetrical. He is aware that in this case the division test is only the test which one tries first, and he is ready, if it proves indecisive, to proceed to a test by simultaneous equations. He does not know any of the latent roots, nor their associated columns, nor that the elementary divisors are non-linear. He prefers, whenever possible, to work with whole numbers, because of their simplicity and accuracy, having noticed Aitken's precept (1937, p. 271) about the meaning which resides in apparently idle digits. As the initial column he chooses $\{1,1,1,1\}$. He observes that $\lambda_1+\lambda_2+\lambda_3+\lambda_4=6+2+3+1=12$; and accordingly he premultiplies by $K-3I$ in the hope of clearing away the associates of any roots near to their mean. Actually there are no roots near the mean, but the computer does not yet know that. The arithmetic begins in summary thus:

$X^0=\{1,1,1,1\};$	$KX^0\div X^0=\{8,10,6,10\};$	range 4.
$X^1=(K-3I)\,X^0=\{5,7,3,7\};$	$KX^1\div X^1=\{5{\cdot}6,6{\cdot}6,7{\cdot}3,7{\cdot}1\};$	range 1·7.
$X^2=(K-3I)\,X^1=\{13,25,13,29\};$	$KX^2\div X^2=\{6{\cdot}46,6{\cdot}16,5{\cdot}38,5{\cdot}86\};$	range 1·08.

The range of the estimates of λ has diminished. A latent root in the neighbourhood of $\lambda=6$ seems to be emerging. The computer guesses that there may be a latent root about as far away from the mean on the other side, that is to say, near $\lambda=0$. Accordingly he purifies X^2 by premultiplying it thrice by $K-0I$, thus:

$X^3=KX^2=\{84,154,70,170\};$	$KX^3\div X^3=\{5{\cdot}86,5{\cdot}80,5{\cdot}83,6{\cdot}02\};$	range 0·22.
$X^4=KX^3=\{492,924,408,1024\};$	$KX^4\div X^4=\{5{\cdot}76,5{\cdot}90,5{\cdot}80,5{\cdot}92\};$	range 0·16.
$X^5=KX^4=\{2836,5448,2368,6064\};$	$KX^5\div X^5=\{5{\cdot}71,5{\cdot}82,5{\cdot}75,5{\cdot}84\};$	range 0·13.

The mean estimate of λ from X^5 is 5·8; but the range of the four estimates, though less than it was, has decreased slowly and is still too large. This makes the computer suspect that something unusual is happening. He divides X^4 and X^5 by their first elements and finds

$$X^4 = 492\{1,\ 1{\cdot}8780,\ 0{\cdot}8292,\ 2{\cdot}0814\}, \quad X^5 = 2836\{1,\ 1{\cdot}9210,\ 0{\cdot}8350,\ 2{\cdot}1382\}.$$

Thus X^4 is considerably out of proportion to X^5. This strengthens the computer's suspicions. First he searches for a mistake in his arithmetic, but, on finding none, he decides to try a test as to whether scalars g and h can be found such that

$$(K^2+gK+hI)\,X^5 = \mathbf{0}.$$

On working out KX^5 and K^2X^5 the suggested relation is found to be

$$\begin{bmatrix} 91984 \\ 182720 \\ 77664 \\ 204512 \end{bmatrix} + g \begin{bmatrix} 16208 \\ 31712 \\ 13616 \\ 35408 \end{bmatrix} + h \begin{bmatrix} 2836 \\ 5448 \\ 2368 \\ 6064 \end{bmatrix} = \mathbf{0}.$$

Although he has returned here to simultaneous equations, it should be noticed that they are not the n simultaneous equations in n unknowns which iterative methods are designed to avoid. For here are only two unknowns and the equations are soluble in pairs quite easily.

From the first and second	$g = -10{\cdot}443,$	$h = 27{\cdot}254.$
From the second and third	$g = -10{\cdot}467,$	$h = 27{\cdot}389.$
From the third and fourth	$g = -10{\cdot}424,$	$h = 27{\cdot}143.$
From the fourth and first	$g = -10{\cdot}416,$	$h = 27{\cdot}095.$

The computer decides to accept the means of g and h as the best information about the latent roots so far available. They are $\bar{g} = -10{\cdot}438$, $\bar{h} = 27{\cdot}220$. It follows that the corresponding latent roots, say λ_1 and λ_2, are approximately

$$\lambda_1,\ \lambda_2 = \tfrac{1}{2}\{-\bar{g} \pm \sqrt{(\bar{g}\bar{g} - 4\bar{h})}\} = 5{\cdot}35,\ 5{\cdot}08.$$

Their near-equality appears to the computer as an explanation of his former suspicions. Intending to reconsider λ_1 and λ_2 at the stage of fine purification, and mindful of the precept that it is desirable to locate all the latent roots roughly before seeking any of them accurately, the computer now attends to λ_3 and λ_4. Their sum is $12-\lambda_1-\lambda_2 = 12+\bar{g} = 1{\cdot}56$, or thereabouts. To find them separately he first prepares a column free from the associates of λ_1 and λ_2. As raw material the computer chooses, just from habit, $\{1, 1, 1, 1\} = Y^0$, say. As a matter of fact this is a very bad choice, but of course the computer could not know that. The method will rescue him. The appropriate purifier is $(K-\lambda_1 I)\,(K-\lambda_2 I)$; and this, to the best of his information, equals $K^2+\bar{g}K+\bar{h}I$. Accordingly he works out

$$(K^2 - 10{\cdot}44K + 27{\cdot}22I)\,Y^0 = \{-4{\cdot}30,\ -1{\cdot}18,\ 4{\cdot}58,\ 2{\cdot}82\} = Y^1, \quad \text{say.}$$

Because $\bar{g}$ and $\bar{h}$ are only approximations to $-\lambda_1-\lambda_2$ and $\lambda_1\lambda_2$ respectively, a second purification is desirable, using the same premultiplier, and retaining all digits. It is found that

$$(K^2 - 10{\cdot}44K + 27{\cdot}22I)\,Y^1 = Y^2, \text{ say } = \{-104{\cdot}4732,\ -19{\cdot}2540,\ 106{\cdot}8708,\ 103{\cdot}2804\}.$$

Evidently Y^2 is not a scalar multiple Y^1. That is to say, there is no evidence that the sequence Y^0, Y^1, Y^2 has converged. This might make the computer feel despondent, if he relied on convergent sequences of columns. He tries the division test and finds

$$KY^2 \div Y^2 = \{3{\cdot}7,\ 1{\cdot}5,\ 0{\cdot}4,\ -0{\cdot}3\}.$$

The discordance of those four numbers might deepen his gloom, if he relied exclusively on the division test. However, he believes that Y^2 is almost free from the associates of λ_1 and λ_2; and, as there can be only two more latent roots, he expects that

$$(K^2+g_3K+h_3I)\,Y^2 = \mathbf{0},$$

where g_3 and h_3 are scalars, of which g_3 is probably about $-1{\cdot}56$. He already has KY^2 and now works K^2Y^2 so as to state the last equation in the form

$$\begin{bmatrix} 2{\cdot}3952 \\ -33{\cdot}0576 \\ 5{\cdot}5968 \\ -110{\cdot}0736 \end{bmatrix} + g_3 \begin{bmatrix} -38{\cdot}3136 \\ -28{\cdot}9176 \\ 44{\cdot}5080 \\ -32{\cdot}5080 \end{bmatrix} + h_3 \begin{bmatrix} -104{\cdot}4732 \\ -19{\cdot}2540 \\ 106{\cdot}8708 \\ 103{\cdot}2804 \end{bmatrix} = \mathbf{0}.$$

From the first and second elements $g_3 = -1{\cdot}5327$, $h_3 = 0{\cdot}5850$.
From the second and third elements $g_3 = -1{\cdot}5335$, $h_3 = 0{\cdot}5863$.
From the third and fourth elements $g_3 = -1{\cdot}5291$, $h_3 = 0{\cdot}5845$.
From the fourth and first elements $g_3 = -1{\cdot}5303$, $g_3 = 0{\cdot}5841$.

The fair agreement between these estimates of g_3 and h_3 is in cheerful contrast with the wild discordance of $KY^2 \div Y^2$. The computer accepts the means $\bar{g}_3 = -1{\cdot}5314$, $\bar{h}_3 = 0{\cdot}5850$, and from them he deduces, as a rough preliminary estimate, that

$$\lambda_3,\ \lambda_4 = \tfrac{1}{2}\{-\bar{g}_3 \pm \sqrt{(\bar{g}_3\bar{g}_3 - 4\bar{h}_3)}\} = 0{\cdot}802,\ 0{\cdot}730.$$

The near equality of these roots explains the lack of convergence of Y^0, Y^1, Y^2.

All four roots are now roughly known, so that the groping stage is ended.

Fine purification is by comparison a straightforward routine. To improve the estimates of λ_1 and λ_2 the purest column for a starting point is

$$X^7 = K^2X^5 = \{91984,\ 182720,\ 77664,\ 204512\}.$$

This should be further purified by premultiplication by $K^2+\bar{g}_3K+\bar{h}_3I$. The result is

$$(K^2-1{\cdot}531K+0{\cdot}585I)\,X^7 = X^8, \text{ say}$$
$$= \{2{,}171{,}612{\cdot}37;\ 4{,}432{,}190{\cdot}21;\ 1{,}850{,}257{\cdot}44;\ 4{,}981{,}614{\cdot}24\}.$$

The computer expects X^8 to be almost free from the associates of λ_3 and λ_4, but to involve those of both λ_1 and λ_2. He is therefore prepared to distrust a division test. Merely from curiosity he tries it and finds

$$KX^8 \div X^8 = \{5{\cdot}579,\ 5{\cdot}630,\ 5{\cdot}596,\ 5{\cdot}637\}.$$

In view of the previous lengthy purification the rough agreement between these four numbers is not satisfactory. The computer rejects them and proceeds to inquire whether there are scalars g_1 and h_1 which satisfy $(K^2+g_1K+h_1I)\,X^8 = \mathbf{0}$.

By taking the first member with each of the others it is found that

$$g_1 = -10{\cdot}472{,}136{,}0, \quad h_1 = 27{\cdot}416{,}408{,}1,$$
$$g_1 = -10{\cdot}472{,}136{,}2, \quad h_1 = 27{\cdot}416{,}409{,}2,$$
$$g_1 = -10{\cdot}472{,}135{,}9, \quad h_1 = 27{\cdot}416{,}407{,}6.$$

The agreement here is 100,000 times better than in $KX^8 \div X^8$. The mean

$$\tfrac{1}{2}(\lambda_1+\lambda_2) = -\tfrac{1}{2}g_1 = 5{\cdot}236{,}068_{,0}$$

in agreement with Aitken. Also $g_1^2 - 4h_1 = 0{\cdot}000{,}000$; but this unfortunately only guarantees that λ_1 and λ_2 are equal to within $\pm 0{\cdot}000{,}5$.

If improved estimates of λ_3 and λ_4 were also required, the first step would be to purify Y^2 by forming

$$Y^3 = (K^2 + g_1 K + h_1 I)\, Y^2.$$

Then h_3 would be redetermined from

$$(K^2 + g_3 K + h_3 I)\, Y^3 = \mathbf{0},$$

in which $-g_3 = 12 + g_1 = 1{\cdot}527{,}864$. Finally, λ_3 and λ_4 would be obtained as

$$\tfrac{1}{2}\{-g_3 \pm \sqrt{(g_3^2 - 4h_3)}\}.$$

By this time the reader has probably felt that it would have been easier to have evaluated the determinant and solved the quartic equation. The author agrees, but points out that the relative merits of the two methods are reversed when the given matrix has 10 rows and 10 columns.

6. General theory in the background needed for unsymmetrical matrices

This theory is said to be in the background because a computer might, with luck, arrive at the desired latent root, and prove its correctness, without knowing anything about the theory of this section. Nevertheless, the theory is certainly helpful, especially in awkward cases. After the computer has found out all about the given matrix, he might look back over his arithmetic and explain what had happened in terms of the fundamental classification of matrices by their 'elementary divisors' (Bôcher 1935, chapter xx), but he may need, as will be shown in § 8, also another characteristic. When, however, the computer is beginning his investigation, the concept of elementary divisors must remain in the background of his thought, while he attends immediately to the symmetry or asymmetry of the given matrix, to appearances of convergence, and to tests of achievement by way of annihilating multipliers.

Aitken (1937), who brought a profound knowledge of matrix theory to bear on the various possibilities that may arise, classified the given matrices thus: Case (i), roots real, single or multiple, elementary divisors linear. Case (ii), conjugate complex roots, possibly multiple, linear elementary divisors. Case (iii), multiple real roots, non-linear elementary divisors. Case (iv), multiple complex pairs of roots, non-linear elementary divisors. Aitken added the remark that combinations of these cases may occur. The numerical examples by which Aitken illustrated his classification show that it corresponds, more or less, to the progression from ease in case (i) to awkwardness in case (iv). Complex roots and equal roots cause little trouble provided the associated elementary divisors are linear.

To explain the fine purification which follows after all the latent roots have been approximately located, the Cayley-Hamilton theorem provides, or leads easily to, all the general theory that is necessary. On the contrary, to explain the computer's initial gropings, a different theory is necessary, namely, that in which an arbitrary column A is expressed as a sum of latent columns, or the like, each multiplied by a scalar coefficient. The computer does not need to find the coefficients in the expansion, but needs only to know that it exists, and how its terms behave when multiplied by $K-\lambda I$. For the symmetrical matrix the n latent columns $P_1, P_2, \ldots, P_n$ are mutually orthogonal; so that the coefficient of P_ν can be found by premultiplying A by $\tilde{P}_\nu$. That is like the pleasant and familiar process of Fourier analysis. On the contrary, there are at least some unsymmetrical matrices for which the latent columns are not all mutually orthogonal; for example, that of § 5. The general method of expansion of A depends instead on the linear independence of n standard columns, and on repeated multiplications by $K-\lambda_j I$ for suitably chosen j. Courant & Hilbert (1924, pp. 30–31) state some of the chief relations, but without any proof. Their notation, for example $f_{e_h}^{(h)}$, is elaborate; perhaps unavoidably so for a quite general statement. A pleasanter notation can be used if instead of stating generalities one may illustrate them by types, the further generalization of which is almost obvious. Thus it may suffice to consider a six-rowed matrix having elementary divisors $(\lambda-\lambda_1)^1$, $(\lambda-\lambda_1)^4$, $(\lambda-\lambda_6)^1$. Here, however, for consistency with the rest of this work, the theory will be expressed in terms of purification.

A latent column and its 'ancestors'

Suppose that there happened to be five equal latent roots $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$. Take an arbitrary column A and operate on it with all the other roots thus:

$$(K-\lambda_6 I)(K-\lambda_7 I)\ldots(K-\lambda_n I)A = T, \quad \text{say.} \tag{152}$$

Then the Cayley-Hamilton theorem leads to the simple result that

$$(K-\lambda_1 I)^5 T = \mathbf{0}. \tag{153}$$

In order to expand an arbitrary column B, say, consisting of n elements, we shall need n standard columns. Besides T four more are needed to make up the quota for the five equal roots. So define S, R, Q, P thus:

$$(K-\lambda_1 I)T = S, \quad (K-\lambda_1 I)S = R, \quad (K-\lambda_1 I)R = Q, \quad (K-\lambda_1 I)Q = P. \tag{154}$$

It follows from (153) and (154) that

$$(K-\lambda_1 I)P = \mathbf{0}. \tag{155}$$

Some of the arbitrariness of A may have passed over into S, R, Q, P. The purpose of this section is to provide a theory capable of extension to all the unusual cases. To illustrate one such case, let us suppose that a particular A, say A_1, leads to T_1, and S_1 such that $S_1 = 0$, and therefore

$$(K-\lambda_1 I)T_1 = \mathbf{0}. \tag{156}$$

We should then naturally call T_1 a latent column and should rename it P_1. Next let A_2 be another raw material, unlike A_1. From A_2 let T_2, S_2, R_2, Q_2, P_2 be prepared by the operations (152), (154). Again it may happen that some of S_2, R_2, Q_2, P_2 vanish. However, for the sake of variety in illustration, let us suppose that none of them do. Then P_2 is a second latent

column belonging to the same latent root as P_1. For subsequent operations P_2 will need to be linearly independent of P_1, so that

$$c_1 P_1 + c_2 P_2 \neq 0, \quad \text{unless} \quad c_1 = 0 \quad \text{and} \quad c_2 = 0. \tag{157}$$

This is likely to happen, because two columns chosen at random are usually linearly independent. If it does not happen, then A_2 must be rejected and a new start made from say A_2' or A_2''... until linear independence occurs. The columns P_1 and P_2 are both annihilated by $(K-\lambda_1 I)$, and so are latent columns in the meaning of that phrase as applied to symmetrical K. But S_2, R_2, Q_2 are annihilated by different multipliers, thus:

$$(K-\lambda_2 I)^4 S_2 = \mathbf{0}, \quad (K-\lambda_2 I)^3 R_2 = \mathbf{0}, \quad (K-\lambda_2 I)^2 Q_2 = \mathbf{0}. \tag{158}$$

It does not seem suitable to call Q_2, R_2, S_2 simply latent columns. Turnbull & Aitken (1932, p. 47) call them 'vectors of grades 2, 3, 4' respectively. Would it be suitable to call Q the 'parent' of P because $(K-\lambda I)\,Q$ gives rise to P? Similarly, R may be called the 'grandparent' of P; and in general Q, R, S, T may be called the 'ancestors' of P. The usually unwritten factor of unity in $(K-\lambda I)\,Q = 1 \times P$ may then be called a 'fertility', equal to unity. I shall use this nomenclature tentatively.

The equations (158) containing squares, cubes, ... of $(K-\lambda_2 I)$ do not lead back to the definitions (154), because $K-\lambda_2 I$, being a singular matrix, has no reciprocal.

The linear independence of the n standard columns

The possible varieties may be sufficiently illustrated by a matrix K of six rows, having five equal roots $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$ and one different root λ_6. As in the previous illustration, let the columns associated with the equal roots split into two sets distinguished by suffixes 1 and 2. The complete set of columns is accordingly

$$T_1 = P_1; \quad P_2, Q_2, R_2, S_2; \quad P_6.$$

Of these P_6 satisfies

$$(K-\lambda_6 I)\,P_6 = \mathbf{0}, \tag{159}$$

and the others satisfy the equations (156) and (158).

Multiply the six columns by scalars c_1, c_2, c_3, c_4, c_5, c_6 and add so as to form

$$Z = c_1 T_1 + c_2 P_2 + c_3 Q_2 + c_4 R_2 + c_5 S_2 + c_6 P_6. \tag{160}$$

The six columns will be linearly independent if Z can be zero only when $c_1, \ldots, c_6$ are all zero (Bôcher 1935, chapter III). At first, however, let Z be as general as it can, and be operated on by various annihilating multipliers. The associates of the equal roots can be simply removed thus:

$$(K-\lambda_1 I)^5 Z = (K-\lambda_1 I)^5 c_6 P_6 = (\lambda_6-\lambda_1)^5 c_6 P_6. \tag{161}$$

The removal of the column P_6 belonging to the single root λ_6 is not so simple, but goes thus:

$$\begin{aligned}(K-\lambda_6 I)\,Z &= (K-\lambda_6 I)\,(Z-c_6 P_6) \\ &= (\lambda_1-\lambda_6)\,(c_1 T_1 + c_2 P_2 + c_3 Q_2 + c_4 R_2 + c_5 S_2) + c_3 P_2 + c_4 Q_2 + c_5 R_2.\end{aligned} \tag{162}$$

This expression can next be broken down by multiplication by powers of $(K-\lambda_1 I)$. Thus $(K-\lambda_1 I)^3$ annihilates the existing T_1, P_2, Q_2, R_2, and changes S_2 into P_2 so that

$$(K-\lambda_1 I)^3\,(K-\lambda_6 I)\,Z = (\lambda_1-\lambda_6)\,c_5 P_2. \tag{163}$$

Similarly,
$$(K-\lambda_1 I)^2 (K-\lambda_6 I) Z = (\lambda_1-\lambda_6)(c_4 P_2 + c_5 Q_2) + c_5 P_2, \tag{164}$$
$$(K-\lambda_1 I)(K-\lambda_6 I) Z = (\lambda_1-\lambda_6)(c_3 P_2 + c_4 Q_2 + c_5 R_2) + c_4 P_2 + c_5 Q_2. \tag{165}$$

Now suppose that Z is zero, and take the equations in the order (161), (163), (164), (165), (162). It follows in turn that
$$0 = c_6 = c_5 = c_4 = c_3, \tag{166}$$
and that
$$c_1 T_1 + c_2 P_2 = 0. \tag{167}$$

But in the manufacture of P_2 it was arranged that this last equation could be satisfied only when c_1 and c_2 were both zero. Therefore the six columns form a linearly independent set.

The possibility of expansion (and the determination of the coefficients)

There is a theorem that 'm sets of n constants each are always linearly dependent if $m>n$' (Bôcher 1935, p. 37). So if A is an arbitrary column there is always an equation
$$c_0 A = c_1 T_1 + c_2 P_2 + c_3 Q_2 + c_4 R_2 + c_5 S_2 + c_6 P_6, \tag{168}$$
in which the $n+1$ scalars $c_0, c_1, \ldots, c_6$ are not all zero. That alone does not assure us of the possibility of expanding A; for c_0 might be zero. However, if c_0 were zero, then not all of $c_1, \ldots, c_6$ could be zero, and this would contradict the previous theorem that the columns on the right are linearly independent. Therefore the expansion of an arbitrary column is possible.

We are now at liberty to regard Z in equation (160), not as a specially defined sum, but as an arbitrary column. When all the latent roots, the latent columns, and their ancestors are known, then the coefficients c in the expansion can be determined after the manner of equations (161) to (167). The first members of those equations can be computed from Z and the equations then can be solved for the c. This process replaces the Fourier type of determination by orthogonality. It should be emphasized that the computer does not ordinarily need to know the coefficients. Nevertheless to explain the example of § 5 they have by exception been worked out for it on p. 478.

Standardizing any latent column and its ancestors

Suppose that numerical columns P, Q, R, S have somehow been computed so as to satisfy
$$(K-\lambda I) S = R, \quad (K-\lambda I) R = Q, \quad (K-\lambda I) Q = P, \quad (K-\lambda I) P = \mathbf{0}. \tag{169}$$

These relations, in which the suffices are implied, have so far been regarded as the definitions of P, Q, R, S. They leave, however, a considerable arbitrariness. For if $\epsilon, \zeta, \eta, \theta$ are any scalars, and if the letters in heavy type represent compound columns made up thus:
$$\mathbf{P} = \epsilon P, \quad \mathbf{Q} = \epsilon Q + \zeta \mathbf{P}, \quad \mathbf{R} = \epsilon R + \zeta \mathbf{Q} + \eta \mathbf{P}, \quad \mathbf{S} = \epsilon S + \zeta \mathbf{R} + \eta \mathbf{Q} + \theta \mathbf{P}, \tag{170}$$
then the $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, when substituted respectively for P, Q, R, S, are found to satisfy the same definitions (169). As much arbitrariness as this may be inconvenient for some purposes. If we begin at P there is one new scalar for each new column. There are not enough disposable constants to provide for mutual orthogonality between four or more columns. But there are just enough to permit normalization, and, for present purposes, normalization is desirable. So let
$$\tilde{\mathbf{P}}\mathbf{P} = 1, \quad \tilde{\mathbf{Q}}\mathbf{Q} = 1, \quad \tilde{\mathbf{R}}\mathbf{R} = 1, \quad \tilde{\mathbf{S}}\mathbf{S} = 1. \tag{171}$$

The equations can then be solved in turn for $\epsilon, \zeta, \eta, \theta$. Because the equations are quadratic, some choices still remain open; at the outset ϵ may have either sign.

Example 11 (*continued* from pp. 470 to 476). The matrix

$$K = \begin{bmatrix} 6 & -3 & 4 & 1 \\ 4 & 2 & 4 & 0 \\ 4 & -2 & 3 & 1 \\ 4 & 2 & 3 & 1 \end{bmatrix}$$

was studied by Aitken (1937), who stated that its latent roots are

$$\lambda_1 = \lambda_2 = 5{\cdot}236068, \quad \lambda_3 = \lambda_4 = 0{\cdot}763932,$$

and that these equal pairs are each associated with a quadratic elementary divisor. From Aitken I took P_1 and P_3. Then I manufactured Q_1 and Q_3 thus:

$$Q_1 = (K-\lambda_3 I)^2\{1,1,1,1\}, \quad Q_3 = (K-\lambda_1 I)^2\{1,1,1,1\}.$$

After normalization the columns are

$$\mathbf{P}_1 = \begin{bmatrix} 0{\cdot}2628 \\ 0{\cdot}6153 \\ 0{\cdot}2350 \\ 0{\cdot}7051 \end{bmatrix}, \quad \mathbf{Q}_1 = \begin{bmatrix} 0{\cdot}3554 \\ 0{\cdot}6020 \\ 0{\cdot}2854 \\ 0{\cdot}6556 \end{bmatrix}, \quad \mathbf{P}_3 = \begin{bmatrix} 0{\cdot}3312 \\ -0{\cdot}1131 \\ -0{\cdot}2962 \\ -0{\cdot}8887 \end{bmatrix}, \quad \mathbf{Q}_3 = \begin{bmatrix} -0{\cdot}2066 \\ -0{\cdot}3678 \\ 0{\cdot}2919 \\ -0{\cdot}8584 \end{bmatrix},$$

where $(K-\lambda_1 I)\,\mathbf{Q}_1 = \mathbf{P}_1$ and $(K-\lambda_3 I)\,\mathbf{Q}_3 = \mathbf{P}_3$. No two of these four columns form an orthogonal pair. In particular, $\tilde{\mathbf{P}}_1\mathbf{Q}_1 = 0{\cdot}9931$. By Cauchy's inequality the inner product $\tilde{A}B$ of any two non-proportional normalized columns A and B must lie between -1 and $+1$. So $\tilde{\mathbf{P}}_1\mathbf{Q}_1$ is nearly up to one of the inevitable bounds. We may say that $\mathbf{P}_1$ is nearly proportional to $\mathbf{Q}_1$. This peculiarity of the matrix K was a cause of embarrassment to the computer in § 5, although he did not know about it at the time. It is not a necessary associate of a quadratic elementary divisor, for $\tilde{\mathbf{P}}_3\mathbf{Q}_3$ is not close to unity, being 0·6496.

The expansion of $\{1,1,1,1\}$ is of interest, because so many authors have taken $\{1,1,\ldots,1\}$ as their raw material. So let

$$\{1,1,1,1\} = a_1\mathbf{P}_1 + b_1\mathbf{Q}_1 + a_3\mathbf{P}_3 + b_3\mathbf{Q}_3,$$

where $\mathbf{P}_1, \mathbf{Q}_1, \mathbf{P}_3, \mathbf{Q}_3$ are the standardized columns given above. It is required, as an exercise, to determine the scalars a_1, b_1, a_3, b_3. Of these a_3 and b_3 can be eliminated by premultiplying both sides of the last equation by $(K-\lambda_3 I)^2$. The result of this operation is

$$\begin{aligned}(K-\lambda_3 I)^2\{1,1,1,1\} &= \{40{\cdot}36068,\ 61{\cdot}30495,\ 31{\cdot}41641,\ 65{\cdot}30495\} \\ &= a_1(\lambda_1-\lambda_3)^2\mathbf{P}_1 + b_1(\lambda_1-\lambda_3)^2\mathbf{Q}_1 + 2b_1(\lambda_1-\lambda_3)\mathbf{P}_1.\end{aligned}$$

The coefficient 2 is noteworthy. This expression involves both a_1 and b_1. However, premultiplication by $K-\lambda_1 I$ cuts out the existing terms in $\mathbf{P}_1$ and changes $\mathbf{Q}_1$ into $\mathbf{P}_1$ with the result that

$$(K-\lambda_1 I)\,(K-\lambda_3 I)^2\{1,1,1,1\} = \{37{\cdot}888_6,\ 88{\cdot}721_3,\ 33{\cdot}888_8,\ 101{\cdot}665_7\} = b_1(\lambda_1-\lambda_3)^2\mathbf{P}_1,$$

in which $(\lambda_1-\lambda_3)^2 = 20$, exactly. This equation gives four values of b_1 which agree, apart from errors due to the rounding off, at $b_1 = 7{\cdot}209_4$. When this value of b_1 is substituted in

the previous equation it gives four values of a_1 which agree as far as $a_1 = -5{\cdot}296$. The process of finding these coefficients involves rather more computation than it would if orthogonality were available, but on the other hand an automatic $(n-1)$-fold check on the arithmetic is included.

The coefficients a_3 and b_3 were found in a similar manner. The final result was

$$\{1, 1, 1, 1\} = -5{\cdot}296\mathbf{P}_1 + 7{\cdot}209\mathbf{Q}_1 - 0{\cdot}317\mathbf{P}_3 + 0{\cdot}319\mathbf{Q}_3,$$

and this has been verified directly.

In § 5 the computer, who was unaware of this expansion, followed custom by taking $\{1, 1, 1, 1\}$ as his raw material. It is rich in the associates of λ_1, poor in those of λ_3. The poverty of a raw material would be of slight importance if the methods of purification were perfect, as they would be in this example if λ_1 were accurately known. Actually the computer in § 5 had at first only a rough estimate of λ_1, and so some of $\mathbf{P}_1$ and $\mathbf{Q}_1$ got through as a contamination to $\mathbf{P}_3$ and $\mathbf{Q}_3$.

Unreliability of Rayleigh's mean when there are non-linear elementary divisors

To take the simplest case let a column X be a mixture of a latent column $\mathbf{P}_1$ with its parent $\mathbf{Q}_1$ thus:

$$X = a\mathbf{P}_1 + b\mathbf{Q}_1. \tag{172}$$

By the fundamental relations

$$(K-\lambda_1 I)\,\mathbf{P}_1 = \mathbf{0}, \quad (K-\lambda_1 I)\,\mathbf{Q}_1 = \mathbf{P}_1.$$

Therefore

$$KX = \lambda_1 X + b\mathbf{P}_1.$$

So Rayleigh's mean μ works out thus:

$$\mu = \frac{\tilde{X}KX}{\tilde{X}X} = \frac{\lambda_1 \tilde{X}X + b\tilde{X}\mathbf{P}_1}{\tilde{X}X}$$

$$= \lambda_1 + b\,\frac{(a\tilde{\mathbf{P}}_1 + b\tilde{\mathbf{Q}}_1)\,\mathbf{P}_1}{(a\tilde{\mathbf{P}}_1 + b\tilde{\mathbf{Q}}_1)\,(a\mathbf{P}_1 + b\mathbf{Q}_1)}. \tag{173}$$

We have

$$\tilde{\mathbf{P}}_1\mathbf{P}_1 = 1, \quad \tilde{\mathbf{Q}}_1\mathbf{Q}_1 = 1.$$

Let

$$\tilde{\mathbf{Q}}_1\mathbf{P}_1 = \tilde{\mathbf{P}}_1\mathbf{Q}_1 = z, \quad \text{say.} \tag{174}$$

Then

$$\mu = \lambda_1 + \frac{b(a+bz)}{a^2+2abz+b^2}. \tag{175}$$

Thus $\mu = \lambda_1$ when $b = 0$, that is, when the parent is absent. On the contrary, when $a = 0$ the parent alone gives $\mu = \lambda_1 + z$. The essence of Rayleigh's principle, in other cases, is that a special mean is almost stationary for slight variations of the impurities about zero. Here the impurity is measured by b. When b alone varies,

$$\left(\frac{\partial\mu}{\partial b}\right)_{\text{at } b=0} = \frac{1}{a}. \tag{176}$$

This is not zero unless $a \to \infty$. So Rayleigh's mean loses its otherwise remarkable detergent property when there is a non-linear elementary divisor.

[*Note added* 21 *January* 1950. A *generalization* designed to avoid this failure of Rayleigh's mean is

$$\tilde{X}K^2X - 2\lambda\tilde{X}KX + \lambda^2\tilde{X}X = 0. \tag{177}$$

This quadratic in λ has roots λ_1 and $2\mu-\lambda_1$ when X is of the form (21). But unfortunately if X involves standard columns not orthogonal with $\mathbf{P}_1$ and $\mathbf{Q}_1$, then (26) also fails.]

The known connexion of the Weierstrassian elementary divisors with the families of characteristic columns has been vaguely mentioned in the foregoing pages, but carefully avoided in any proof.

The basic ideas of these two systems seem at first sight to be detached from one another; for the doctrine of elementary divisors begins with the greatest common divisor of the i-rowed determinants of $|K-\lambda I|$ (Bôcher 1935, p. 263), and this seems far away from multiplication by columns. The connexion between the two systems was exhibited by Weierstrass in a paper about differential equations (1875, reprinted 1895). The result is very simple. To each elementary divisor $(\lambda-\lambda_j)^{e_i}$ there corresponds a family of e_i characteristic columns P, Q, R, ..., one of which, P, satisfies $(K-\lambda_j I)\,P=\mathbf{0}$, while any others are connected in successive pairs thus:

$$(K-\lambda_j I)\,Q=P, \quad (K-\lambda_j I)\,R=Q,$$

and so on. Weierstrass's original proof was very complicated; nor have I found in modern texts any complete proof which is easy (for me) to understand; although the materials for building a proof are all ready (Bôcher 1935, chapters xx and xxi; Turnbull & Aitken 1932). So I offer the following arrangement of known theorems.

It is proved (Bôcher 1935, pp. 262–277 that the elementary divisors of $K-\lambda I$ are the same as those of $H(K-\lambda I)\,H^{-1}$, where H is any non-singular square matrix independent of λ and having the same number of rows as K. It is proved (Bôcher 1935, pp. 279–283) that if the matrices $K-\lambda I$ and $k-\lambda I$ have the same elementary divisors, then an H always exists such that $H(K-\lambda I)\,H^{-1}=k-\lambda I$. It is shown (Bôcher 1935, pp. 288–289) that for a given set of elementary divisors a delightfully simple, alias canonical, form of $H(K-\lambda I)\,H^{-1}$ is that in which the elements are zero everywhere except on the principal diagonal, where $\lambda_j-\lambda$ occurs, and on the line next above the diagonal where units may, or may not, occur. According to the pattern of any units in the superdiagonal, the matrix $H(K-\lambda I)\,H^{-1}$ can be partitioned into principal minors which do not overlap. Each such minor corresponds to one elementary divisor, and conversely. These general statements will be illustrated by supposing that the elementary divisors of $H(K-\lambda I)\,H^{-1}$ happen to be $(\lambda_1-\lambda)^1$, $(\lambda_1-\lambda)^4$, $(\lambda_6-\lambda)^1$. For in that case

$$H(K-\lambda I)\,H^{-1}=HKH^{-1}-\lambda I=\begin{bmatrix} (\lambda_1-\lambda) & 0 & 0 & 0 & 0 & 0 \\ 0 & (\lambda_1-\lambda) & 1 & 0 & 0 & 0 \\ 0 & 0 & (\lambda_1-\lambda) & 1 & 0 & 0 \\ 0 & 0 & 0 & (\lambda_1-\lambda) & 1 & 0 \\ 0 & 0 & 0 & 0 & (\lambda_1-\lambda) & 0 \\ 0 & 0 & 0 & 0 & 0 & (\lambda_6-\lambda) \end{bmatrix}. \tag{178}$$

Each non-overlapping principal minor is enclosed in a dotted rectangle. On postmultiplying this canonical matrix by a column $\{a,b,c,d,e,f\}$ consisting of arbitrary elements, it is evident that a interacts only with the minor belonging to the elementary divisor $(\lambda_1-\lambda)^1$; that b, c, d, e interact only with the minor belonging to $(\lambda_1-\lambda)^4$ and that f interacts only with

the minor belonging to $(\lambda_6-\lambda)$. This separation is evidently not confined to the present illustration. So we can treat each elementary divisor separately by preparing a column having zeroes in all the rows which belong to the other elementary divisors. Moreover, as common factors are of scant interest, the letters may then be replaced by units. In particular,

$$(HKH^{-1}-\lambda_1 I)\{1,0,0,0,0,0\} = \{0,0,0,0,0,0\}. \tag{179}$$

So denote $\{1,0,0,0,0,0\}$ by p_1 because it is a latent column of the simplified matrix and corresponds to the latent root λ_1. For a similar reason let $\{0,0,0,0,0,1\} = p_6$.

The relations of the quartic elementary divisor are found by multiplying by the four-rowed minor, after setting $\lambda = \lambda_1$. For brevity the inactive rows and columns may be omitted thus

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} c \\ d \\ e \\ 0 \end{bmatrix}. \tag{180}$$

The effect of the multiplication is to push the elements b, c, d, e up one place, b being pushed out at the top, and a zero entering at the bottom (Turnbull & Aitken 1932, pp. 62–63). Thus $\{b,c,d,e\}$ is *not* a latent column. But if we make all but one of b, c, d, e, zero, we obtain a family of columns connected by the 'parental' relationship. They may be lettered p_2, q_2, r_2, s_2 to correspond with the P, Q, R, S used in computing operations. The zero elements opposite to the minors of the other elementary divisors will now be restored, and p_2, q_2, r_2, s_2 defined thus:

$$\left.\begin{aligned} p_2 &= \{0,1,0,0,0,0\}, \quad q_2 = \{0,0,1,0,0,0\}, \\ r_2 &= \{0,0,0,1,0,0\}, \quad s_2 = \{0,0,0,0,1,0\}. \end{aligned}\right\} \tag{181}$$

For then

$$\left.\begin{aligned} (HKH^{-1}-\lambda_1 I)\, p_2 &= \mathbf{0}, \quad (HKH^{-1}-\lambda_1 I)\, q_2 = p_2, \\ (HKH^{-1}-\lambda_1 I)\, r_2 &= q_2, \quad (HKH^{-1}-\lambda_1 I)\, s_2 = r_2. \end{aligned}\right\} \tag{182}$$

The six columns p_1; p_2, q_2, r_2, s_2; p_6 are mutually orthogonal. They relate to the simplified, alias canonical, matrix HKH^{-1}. Our concern, however, is with the given matrix K. Is there a way back? Take a typical relationship such as

$$(HKH^{-1}-\lambda_1 I)\, q_2 = p_2, \tag{183}$$

and premultiply it by H^{-1} obtaining

$$(K-\lambda_1 I)\, H^{-1} q_2 = H^{-1} p_2. \tag{184}$$

So if Q_2 and P_2 are defined by $\quad Q_2 = H^{-1}q_2, \; P_2 = H^{-1}p_2, \tag{185}$

it follows that $\quad (K-\lambda_1 I)\, Q_2 = P_2. \tag{186}$

Similarly for all the other columns p_1, r_2, s_2, p_6. Thus the connexion between the elementary divisors and the families of characteristic columns is established.

But what of orthogonality? It has been proved that $\tilde{p}_2 q_2 = 0$. Does $\tilde{P}_2 Q_2$ therefore vanish? From the definitions

$$\tilde{P}_2 Q_2 = (H^{-1}p_2)^{\sim} H^{-1} q_2 = \tilde{p}_2 (H^{-1})^{\sim} H^{-1} q_2. \tag{187}$$

In general $(H^{-1})^{\sim} H^{-1}$ is a square matrix which cannot be extricated from its position between $\tilde{p}_2$ and q_2. There is thus no reason to expect that P_2 should be orthogonal with Q_2.

The relation between elementary divisors and annihilating multipliers

Each of the columns associated with the elementary divisor $(\lambda-\lambda_\nu)^r$ is annihilated on premultiplication by $(K-\lambda_\nu I)^r$, whereas $(K-\lambda_\nu)^{r-1}$ would leave one of those columns extant. Thus a classification of matrices by their elementary divisors is closely related to a classification of the columns belonging to those matrices by their simplest annihilating multipliers. From the computer's point of view an annihilating multiplier is what he is immediately seeking, whereas an elementary divisor is part of a far-away theory.

The correspondence between the above two classifications is close, provided that complex multipliers are admitted. The computer may, however, prefer to work only with real multipliers, and if so a quadratic multiplier $\{K^2-2\alpha K+(\alpha^2+\beta^2)\, I\}$ is required to annihilate a conjugate pair of complex roots $\lambda = \alpha \pm i\beta$, each of which corresponds to a linear elementary divisor.

Cleaning operations when the roots may be complex or equal, but the elementary divisors are linear

In the first example (p. 442) the columns P_2, P_3 associated with a pair of complex latent roots λ_2, $\lambda_3 = -3{\cdot}986 \pm 0{\cdot}918i$ were gradually cleaned away by the real premultiplier $K+4I$, leaving the column associated with $\lambda_1 = 1{\cdot}973$. The explanation is, briefly, that on the Argand diagram a circle centred at the point -4 and having the point $1{\cdot}973$ on its circumference, contains both the points $-3{\cdot}986 \pm 0{\cdot}918i$.

The general explanation of this case is as follows. An arbitrary initial column can be expanded in the form

$$X^0 = \sum_{\nu=1}^{\nu=n} a_\nu P_\nu, \tag{188}$$

where the a_ν and P_ν may be complex, but ancestors such as Q, R, ... do not appear. The a_ν are single complex numbers, the P_ν are columns of numbers annihilated linearly, thus

$$(K-\lambda_\nu I)\, P_\nu = \mathbf{0}, \tag{189}$$

where the latent root λ_ν may be complex. Let the premultiplier for making the iterations be $(K-\theta I)$, so that

$$X^1 = (K-\theta I)\, X^0, \tag{190}$$

where θ may be complex. It follows that

$$X^1 = \sum_{\nu=1}^{\nu=n} (\lambda_\nu-\theta)\, a_\nu P_\nu. \tag{191}$$

Let λ_s be the desired latent root. To keep the coefficient of P_s unchanged, so that it may serve as a standard, let X^0 be compared, not with X^1 but with

$$\frac{X^1}{\lambda_s-\theta} = \sum_{\nu=1}^{\nu=n} \frac{\lambda_\nu-\theta}{\lambda_s-\theta}\, a_\nu P_\nu. \tag{192}$$

The desired cleaning will proceed if $|\lambda_\nu-\theta| < |\lambda_s-\theta|$ for every ν except s. These moduli are distances on the Argand diagram. So the requirement can be expressed by saying that

for a linear purifier the circle with centre θ and having the desired root on its circumference must contain all the unwanted roots. That is necessary for convergence, fast or slow. To make convergence rapid it is desirable, if possible, to choose θ so that the unwanted roots lie near the centre of the circle.

There may be no such circle; for example, if the desired root λ_s is surrounded by unwanted roots. The computer must then either alter his desires, or else choose a non-linear purifier, say $(K-\theta_1 I)(K-\theta_2 I)$, where both θ_1 and θ_2 may be complex. For comparison with X^0 it is proper now to take

$$\frac{(K-\theta_1 I)(K-\theta_2 I)X^0}{(\lambda_s-\theta_1)(\lambda_s-\theta_2)} = \sum_{\nu=1}^{\nu=n} \frac{(\lambda_\nu-\theta_1)(\lambda_\nu-\theta_2)}{(\lambda_s-\theta_1)(\lambda_s-\theta_2)} a_\nu P_\nu. \tag{193}$$

As the location of the roots λ_ν is often very vaguely known, it is suitable to omit the suffix ν and to regard λ as a movable point on the Argand diagram. Then

$$|(\lambda-\theta_1)(\lambda-\theta_2)| = |(\lambda_s-\theta_1)(\lambda_s-\theta_2)| \tag{194}$$

is a curve, and λ_s lies on it. It may be called the 'null curve', because, if any other root lies on it, the multiplication by $(K-\theta_1 I)(K-\theta_2 I)$ has no relative effect on its latent column. On one side of the null curve lie those latent roots whose latent columns are gradually cleaned away. Any latent roots on the other side of the null curve have latent columns which would gradually become more prominent than that of the desired root; therefore the computer should if possible choose θ_1 and θ_2 so that all the unwanted roots lie on the cleanable side of the null curve. The factors $|\lambda-\theta_1|$ and $|\lambda-\theta_2|$ are the lengths of the radii, r_1 and r_2, say, from the poles θ_1 and θ_2 respectively to the movable point λ. Thus on the null curve $r_1 r_2$ is constant. Lamb (1924, p. 321) in connexion with bipolar co-ordinates, discusses the family of curves $r_1 r_2 =$ constant. They are known as the ovals of Cassini. Lamb gives a diagram. When the desired root λ_s is near enough to either θ_1 or θ_2 the cleanable region consists of two separate areas, one around each pole. As λ_s moves away these coalesce into a single cleanable area. The form at which coalescence just occurs is a figure-of-eight known as the lemniscate of Bernoulli.

Cleaning operations when the elementary divisors are non-linear

This case may be troublesome, as Aitken pointed out (1937, p. 275). The expansion of an arbitrary column X^0 may now contain not only P but the 'ancestors' of P which have been denoted by Q, R, S, ... On multiplication by K each ancestor throws off descendants which complicate the theory. I refrain from attempting any formal general statement. For the general formulae would be choked with exponents of elementary divisors, which the computer could ascertain only after considerable exploration. It is more practical for him to remember that he may on rare occasions encounter awkward special cases. One of them is illustrated in §§ 5 and 8.

7. Tests of achievement versus proofs that infinitely many operations would converge

H. F. Baker (1902, 1916) made remarkable applications of infinite series of matrices, reducing them with the aid of the Cayley-Hamilton theorem to polynomials. Duncan & Collar (1934, p. 882), Aitken (1937), and Hotelling (1943*a*) all define infinite sequences of

computing operations, and all prove that they will converge. In the present paper, by contrast, no infinite sequences are contemplated. The process, like a game, is not completely defined in advance. The computer is at liberty either to follow a carefully planned and verified succession of operations, or else to muddle through anyhow, provided only that his result passes a test of achievement. There are some theorems, e.g. in § 3, which would prove the convergence of infinite sequences, if such were contemplated; but actually the theorems are only used to show that one or two or a few operations will make an improvement. Since the time of Cauchy, proofs of convergence have been greatly respected in mathematics. Any treatment of a limiting operation which omits a proof of convergence is nowadays habitually condemned as lax. Evidently I must defend myself against a prospective charge of laxity. The question at issue is mainly about the tense of the verb; 'has converged' versus 'will converge'. It is of interest wider than that of matrices; and it brings in considerations from psychology as well as from mathematics. A 'convergence theorem' in the ordinary meaning of the phrase is, from the computer's point of view, a proof that infinitely many operations would converge.

A. The correctness of the latent roots and latent columns is certified in the present method by the equality of the n, or fewer, estimates of a latent root obtained either by the division test or by a redundant set of simultaneous equations. The agreement is usually to a finite number of decimal places. There may be a question as to what residual uncertainty in a latent column belongs to the observed residual disagreement between the n estimates of its latent root (see § 8). But anyway, a proof that the process must converge, if continued for ever, could not make the results at any stage more certain.

B. In the absence of a test of achievement, a proof that the process must converge may mislead the computer into thinking that it already has converged. An illustration may be drawn from infinite series. The computer, let us suppose, has added terms until each separately has become negligibly small. A knowledge that the series is convergent tempts the computer to think that he has done enough, but does not prove that he has. If he were provided with bounds to the sum of the remainder that would be a test of achievement.

C. Whenever an easy test of achievement is available to the computer, he is more likely to feel tired than encouraged by a proof that an infinite number of his operations would lead to the desired result.

D. The computer does, however, need to be led on by a reasonable hope that his process will probably converge sufficiently after a moderate number of operations to pass the test of achievement.

E. A proof of convergence may prevent the exercise of skill and judgement by laying down in advance a rigid programme of operations. If the computer relies on a proof that a defined process will converge, he must for honesty's sake not make a mistake anywhere in that process. If, on the contrary, he relies on a test of achievement there is no dishonesty in his making mistakes or intentional alterations at any stage except that of the test itself.

F. In view of the possibility of mistakes, it is very advantageous if a few operations starting anywhere tend to improvement. This is the 'pleasing characteristic of iterative processes' mentioned by Whittaker & Robinson (1924, p. 81) and by various other writers (Richardson 1910, p. 325; Duncan & Collar 1934, p. 866; Hotelling 1943 *a*). In the present work the proofs relate to the conditions under which a few operations will make an improvement.

They may be called 'proofs of progress' in contradistinction to the orthodox 'proofs of convergence' which relate to infinitely many operations.

G. The motives and convictions of computers may usefully be compared with those of people in other occupations. A ship's captain, wishing to take his vessel from New York to Liverpool, does not require in advance a formal proof that he will certainly converge on Liverpool; he is content with a reasonable expectation of so doing, and is guided by a test of achievement, namely, the sextant observations, which show how he is getting on. In like manner the chemical distiller often proceeds, not knowing quite what will happen, but guided by the temperature of the vapour, which is one of his many tests of achievement. Similarly the lens polisher knows that his machine tends usually to make surfaces nearly spherical, but he is guided by interference fringes, which show him how nearly sphericity has been attained. Instead of an initial proof that one will arrive (proof of convergence) there is in these, and in many other practical affairs, a final proof that one has arrived (test of achievement). Why should not similarly skilful operations be considered respectable in arithmetic?

H. Iterative computations are not quite limiting operations in the strictest sense; for, although the computer can always attain greater accuracy by doing more arithmetic, yet in fact he is usually content with less than twenty significant digits; so that theorems about infinite accuracy are not required. Here is another resemblance to lens-polishing; for if a lens is correct to within a tenth of the wave-length of light, that is near enough.

8. A caution relating to almost proportionality

The division test gives n estimates of a latent root. In many examples the true value has been found to lie between the greatest and the least of the estimates. The comfortable feeling of security induced by such experiences received a rude shock in the example of § 5, where

$$KX^5 \div X^5 = \{5{\cdot}71,\ 5{\cdot}82,\ 5{\cdot}75,\ 5{\cdot}84\},$$

but the true value of the latent root was *outside* the range of these estimates at $\lambda_1 = 5{\cdot}236$. The computer abandoned the division test because the column X^5 was out of proportion to X^4. If he had not noticed that, the division test might have misled him.

The peculiar behaviour persisted after the associates of the other equal roots λ_3, λ_4 had been cleaned away, leaving a column X^8; for it was found that

$$KX^8 \div X^8 = \{5{\cdot}579,\ 5{\cdot}630,\ 5{\cdot}596,\ 5{\cdot}637\},$$

which again shows a spurious appearance of agreement.

In search for an explanation of these peculiarities, let a column X be expressible in terms of the latent column P_1 and its parent Q_1, thus:

$$X = aP_1 + bQ_1, \tag{195}$$

where a and b are scalars. The fundamental relations are

$$KQ_1 = \lambda_1 Q_1 + P_1 \quad \text{and} \quad KP_1 = \lambda_1 P_1;$$

so that

$$KX = \lambda_1 X + bP_1.$$

Accordingly

$$KX \div X = \lambda_1 + b(P_1 \div X). \tag{196}$$

In general $P_1 \div X$ is a column of unequal numbers, so that the n estimates $KX \div X$ are discordant unless $b = 0$; and therefore concordance is a proof that X is a pure multiple of P_1. In the present example, however, Q_1 is nearly proportional to P_1 and so also to X. It follows that $P_1 \div X$ is a column of nearly equal numbers; and therefore the approximate concordance of the n estimates $KX \div X$ can occur with quite a large admixture of Q_1. The misleading appearance is thus not due to a non-linear elementary divisor alone, but to that in combination with the almost proportionality of the latent column to its parent. These circumstances are fortunately rare, and in particular they cannot occur when the matrix K is symmetrical. Moreover, the theorem (p. 476) that the n standard columns characteristic of K are linearly independent, shows that the difficulty of almost proportionality can always be overcome by working to a sufficient number of digits.

9. Summary and conclusions as to matrices

The given matrix K has n rows and n columns. A column X of n numbers is chosen as a raw material for purification; X may be arbitrary, but is preferably a good guess at the desired latent column. The routine is to premultiply X by several factors of the form $K - \rho I$, where I is the unit matrix and ρ is, in the simplest idea, an estimate of an unwanted latent root. The best choice of $\rho_1, \rho_2, \ldots$ is considered from the point of view of least squares. Duncan & Collar (1934), Aitken (1937) and Hotelling have published methods like this, except that they took $\rho = 0$. The advantage of an adjustable ρ is that it allows the computer quickly to clean away suspected impurities. Tests are given for deciding whether a column, Z say, has been sufficiently purified. The test which is simplest and best known, though not always appropriate, is to divide each element of KZ by the corresponding element of Z. If the n ratios are equal, then Z is a latent column. More elaborate tests are given. There is no need to consider infinite sequences of computing operations.

10. Application to differential equations of purifiers prepared for matrices

The beginning of the present work on matrices was an idea borrowed from a paper on partial differential equations (Richardson 1910); now at the end there are some ideas to pay back. The constants α and β, summarized above on pp. 467–469, are likely to be useful to anyone who solves partial differential equations by such methods (e.g. Newing 1941). But one needs to see just how the constants fit in.

Since 1938 Southwell and his collaborators have solved very many and important problems involving partial differential equations by an arithmetical method which he has named 'relaxation'. A recent summary by him is in book form (1946). I suppose—though this may be controversial—that 'relaxation' is sufficiently alike to 'purification' for the choice of purifiers to have some relevance to relaxation.

In connexion with the historical development, Sir Richard Southwell has drawn my attention to a remarkably extensive, though still incomplete, bibliography prepared by T. J. Higgins (1949). Higgins found that Runge (1908–9) was the first to apply finite differences to the partial differential equation of Saint-Venant's torsion problem. Incidentally, I must protest against Higgins's suggestion that there is a contrast between iteration and what he calls Richardson's method of 'normal functions'. For, in fact, I cleaned away unwanted normal functions by iteration.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The present comparison of techniques will, however, relate to my paper of 1910. The phraseology of cleaning was not then mentioned, but in retrospect it seems appropriate. Differential equations were replaced by difference equations, and the approach to the limit was deferred (Richardson & Gaunt 1927), so that all the operations were on finite sets of arithmetical numbers. [Bogoliouboff & Kryloff (1928), who called this 'Rayleigh's Principle', evaluated the error committed at the nth approximation. There are numerous related papers by Kryloff.] Other broad resemblances between the treatment of the matrices and that of the differential equations include the following: in both there is an initial guess; in both the errors of that guess are expandable in a finite set of orthogonal functions; in both the computer does not need to know the details of the expansion; in both unwanted orthogonal functions are gradually cleaned away by iteration. For the partial differential equations the greatest and the least of the eigenvalues were first roughly estimated. For the matrices the greatest and least of the latent roots can conveniently be estimated before the others. In Richardson's paper of 1910 the figures 1, 2, 3 represent cleaning operations, and have co-ordinates comparable with those of the present figures 1 and 2; the ordinates are the same; the abscissae differ only by changes of origin and scale. To see the resemblance one has to bear in mind that the notations are different, λ^2 there corresponding to λ here, and λ_n^2 to λ_1. The present choice of purifiers for symmetrical matrices, being based on Legendre's principle of least squares, is much more systematic than those I published in 1910.

In order to see more exactly how the purifiers chosen for matrices can be applied to differential equations it is desirable to consider two of the simplest jury problems to which purification is applicable. Each problem can be regarded as a set of n simultaneous linear algebraic equations; the contrast between the problems is that the first set of equations is homogeneous, the second not. Here δ is a finite difference operator; it is centered and partial.

Problem 1. Given that

$$\left.\begin{aligned} \frac{\delta^2 P}{\delta x^2}+\frac{\delta^2 P}{\delta y^2} &= \lambda P \quad \text{at every one of } n \text{ body-points,} \\ \text{and that} \qquad P &= 0 \quad \text{at every boundary-point,} \end{aligned}\right\} \tag{197}$$

required to find λ and P. It is well known that a solution exists only for special values of λ, called the eigenvalues. Let them be denoted by $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$. For each eigenvalue λ_ν there is a corresponding $P = P_\nu$, which, however, is arbitrary as to a multiplicative constant. That is exactly like finding a latent column for a given symmetrical matrix. Indeed, by arranging the n body values in a linear order, the difference equations can be expressed as a matrix equation $(K-\lambda I)\,P = \mathbf{0}$. Such a linear arrangement is, however, artificial in a two-dimensional problem. For the purpose of choosing purifiers, a different analogy is preferable. Let $\mathscr{K}$ be a contraction for the difference operator thus:

$$\mathscr{K} = \frac{\delta^2}{\delta x^2}+\frac{\delta^2}{\delta y^2}. \tag{198}$$

Then the given equation can be written

$$\mathscr{K} P = \lambda P. \tag{199}$$

Let ϕ_1 be an initial guess, consisting of zero at each boundary-point and any real number at each body-point. Then ϕ_1 can be expanded thus:

$$\phi_1 = \sum_{\nu=1}^{\nu=n} A_\nu P_\nu, \tag{200}$$

where the A_ν are constants. To make the A_ν definite, let us suppose that the P_ν have been normalized. Let a second approximation ϕ_2 be obtained from ϕ_1 by the operation

$$\phi_2 = \left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2}\right)\phi_1 - \rho\phi_1. \tag{201}$$

This, being $\phi_2 = \mathscr{K}\phi_1 - \rho\phi_1$, resembles the matrix operation

$$X^2 = KX^1 - \rho X^1 = (K - \rho I)\, X^1,$$

which has been thoroughly discussed in the previous pages. If $\rho = \lambda_\nu$, the corresponding P_ν does not appear in the expansion of ϕ_2: that normal mode has been cleaned away. In order to make such purifications, estimates of the greatest and the least of the eigenvalues are needed. When $\delta x = \delta y = 1$, then λ_1 is close to 8. A sufficient estimate of λ_n can be obtained by considering some simple but slightly larger boundary, perhaps a rectangle, for which formulae for P_n and λ_n are known (Richardson 1910, pp. 320–321). Suppose now that it is desired to find more accurately one eigenvalue λ_s and its normal mode P_s. Then all the other $n-1$ normal modes must be cleaned away. This is exactly like finding a latent root and latent column of a matrix. The constants α and β, summarized on pp. 467–469 above, apply to this problem, as fixing the best choice of ρ in terms of the rough estimates of λ_1, λ_n, λ_s and the number m of operations in prospect.

Problem 2. For the same geometrical pattern of points as in problem 1 let it be given that

$$\left.\begin{aligned} &\frac{\delta^2\phi}{\delta x^2} + \frac{\delta^2\phi}{\delta y^2} = 0 \quad \text{at every body-point,} \\ &\text{and let } \phi \text{ be given at every boundary-point.} \end{aligned}\right\} \tag{202}$$

It is well known that there is only one solution ϕ, which, unlike P_k, does not involve any arbitrary constant. The initial guess ϕ_1 must now have the given boundary values instead of zeroes. Not ϕ_1 but its error $\phi_1 - \phi$ is now expandable in the form

$$\phi_1 - \phi = \sum_{\nu=1}^{\nu=n} A_\nu P_\nu. \tag{203}$$

The purifying operation $\phi_{n+1} = \mathscr{K}\phi_n - \rho\phi_n$, which suited problem 1, will not do for problem 2, because if $\mathscr{K}\phi_n = 0$, then must $\phi_{n+1} = \phi_n$, whereas the multiplier $-\rho$ gets in the way. The operation is therefore altered so as to become

$$\phi_{t+1} = \phi_t - \frac{1}{\rho_t}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2}\right)\phi_t. \tag{204}$$

Then if $\rho_t = \lambda_\nu$, the error of ϕ_{t+1} does not involve the normal mode P_ν specified in problem 1. Purifications can therefore proceed much as for the matrices; but not quite so. For a clean sweep must now be made of all, not all but one, of the P_ν. Equation (204) can be rearranged in terms of the errors thus:

$$\phi_{t+1} - \phi = \left\{1 - \frac{1}{\rho_t}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2}\right)\right\}(\phi_t - \phi). \tag{205}$$

But by (203) and (197) $$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)(\phi_1-\phi)=\sum_{\nu=1}^{\nu=n}A_\nu\lambda_\nu P_\nu. \tag{206}$$

So $$\phi_2-\phi=\sum_{\nu=1}^{\nu=n}A_\nu\left(1-\frac{\lambda_\nu}{\rho_1}\right)P_\nu. \tag{207}$$

On repeating the operation (205) for $t = 2, 3, 4, \ldots, m$ in turn we arrive at

$$\phi_{m+1}-\phi=\sum_{\nu=1}^{\nu=n}A_\nu P_\nu\left(1-\frac{\lambda_\nu}{\rho_1}\right)\left(1-\frac{\lambda_\nu}{\rho_2}\right)\ldots\left(1-\frac{\lambda_\nu}{\rho_m}\right), \tag{208}$$

a result given in a different notation in 1910. As the λ_ν are unknown, except for estimates of λ_1 and λ_n, the suffix ν may suitably be omitted and ω_m may be defined thus:

$$\omega_m=\left(1-\frac{\lambda}{\rho_1}\right)\left(1-\frac{\lambda}{\rho_2}\right)\ldots\left(1-\frac{\lambda}{\rho_m}\right). \tag{209}$$

Here ω may be called a ratio of purification. It should be made near to zero for all points in the closed interval $\lambda_n\leqslant\lambda\leqslant\lambda_1$. The distinction from problem 1 is that there is here no desired P_s which can serve as a standard of comparison. The ratio of purification f_m, which was used in connexion with the matrices on p. 457, became unity for some finite λ; on the contrary ω_m is equal to unity at $\lambda = 0$. That is why the abscissae of the diagrams have different origins. Otherwise similar notions apply. Thus let a 'root-mean-square ratio of purification', J, be defined by

$$J^2=\frac{1}{\lambda_1-\lambda_n}\int_{\lambda_n}^{\lambda_1}\omega_m^2\,d\lambda. \tag{210}$$

From the point of view of least squares, it is desirable to make

$$J^2 \text{ a minimum for variations of } \rho_1, \rho_2, \ldots, \rho_m. \tag{211}$$

To facilitate comparison with the integral G^2 which was applied to the matrices in (65) of § 3, transform the variables by division by the greatest of the eigenvalues, defining x, γ, ϵ thus:

$$\lambda/\lambda_1=x, \tag{212}$$

$$\rho_j/\lambda_1=\gamma_j, \tag{213}$$

$$\lambda_n/\lambda_1=\epsilon; \tag{214}$$

for then $$J^2=\frac{1}{1-\epsilon}\int_\epsilon^1\left(1-\frac{x}{\gamma_1}\right)^2\left(1-\frac{x}{\gamma_2}\right)^2\ldots\left(1-\frac{x}{\gamma_m}\right)^2dx. \tag{215}$$

The integral in (215) differs from that whereby the α were determined in having a lower terminus ϵ instead of zero. Many of the theorems in § 3 are true for any fixed termini, and so apply here. In particular, (98) and (102) prove that it is advantageous to scatter the γ, as A. E. H. Love suggested. Also the conditions for finding the best γ simplify, as in (106), to

$$\int_\epsilon^1 x\omega_m\,dx=0,\quad \int_\epsilon^1 x^2\omega_m\,dx=0,\quad \ldots,\quad \int_\epsilon^1 x^m\omega_m\,dx=0. \tag{216}$$

But at the next stage the terminus ϵ affects the result. Thus for $m = 1$ it follows from (216) that

$$\gamma=\frac{2}{3}\left(\frac{1-\epsilon^3}{1-\epsilon^2}\right)=\frac{2}{3}\left(\frac{1+\epsilon+\epsilon^2}{1+\epsilon}\right); \tag{217}$$

whereas α was simply $\frac{2}{3}$. For $m = 2, 3, 4, \ldots$ the best purifiers γ might be laboriously obtained from (216) as rather complicated functions of ϵ; but there is a short cut. For to make a good finite-difference approximation to a differential equation, δx and δy need to be small fractions of the diameter of the figure; and therefore λ_n is in practice usually a small fraction of λ_1; and so ϵ is much nearer to 0 than to 1. Also the α were chosen so as to make $\int_0^1 f_m^2 dx$ a minimum. When this has been done, f_m^2 is greater near $x = 0$ than anywhere else in the range $0 \leqslant x \leqslant 1$. So for ϵ near zero

$$\frac{1}{1-\epsilon}\int_\epsilon^1 f_m^2 dx < \int_0^1 f_m^2 dx. \tag{218}$$

Now the α make the second member as small as it can be. So the α make the first member usefully small. The conclusion is that: *the α specified on pp.* 467–468 *for finding the least latent root of a matrix are also usefully good, though probably not quite the best, for solving problem* 2 *relating to Laplace's equation in a plane*. That is to say take, instead of (213),

$$\rho_t = \rho_j = \alpha_j \lambda_1, \tag{219}$$

where λ_1 is approximately

$$4/(\delta x)^2 + 4/(\delta y)^2, \tag{220}$$

for insertion in (204).

References

Aitken, A. C. 1937 The evaluation of the latent roots and latent vectors of a matrix. *Proc. Roy. Soc. Edinb.* **62**, 269–304.

Baker, H. F. 1902 On the integration of linear differential equations. *Proc. Lond. Math. Soc.* **35**, 333–378.

Baker, H. F. 1916 On certain linear differential equations of astronomical interest. *Phil. Trans.* A, **216**, 129–186.

Bôcher, M. 1935 *Introduction to higher algebra*. New York: Macmillan.

Bogoliouboff, N. & Kryloff, N. 1928 *Ann. Math. Princeton, etc.*, 2nd ser., **29**, 255–275.

Bromwich, T. J. I'A. 1906 *Quadratic forms and their classification by means of invariant factors*. Cambridge University Press.

Courant, R. & Hilbert, D. 1924 *Methoden der mathematischen physik*. Berlin: Springer.

Duncan, W. J. & Collar, A. R. 1934 A method for the solution of oscillation problems by matrices. *Phil. Mag.* May 1934.

Duncan, W. J. & Collar, A. R. 1935 Matrices applied to the motions of damped systems. *Phil. Mag.* (7), **19**, 197–219.

Frazer, R. A., Duncan, W. J. & Collar, A. R. 1938 *Elementary matrices, with some applications to dynamics and differential equations*. Cambridge University Press.

Goursat, E. 1924 *Cours d'analyse mathematique*. Paris: Gauthier-Villars.

Higgins, T. J. 1949 Typescript for Grinter's *Numerical methods of analysis in engineering*. New York: Macmillan.

Hobson, E. W. 1921 *Plane trigonometry*. 5 ed. Cambridge University Press.

Holzinger, K. J. & Harman, H. H. 1948 *Factor analysis*. University of Chicago Press.

Hotelling, H. 1933 Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441 and 498–520.

Hotelling, H. 1943*a* Some new methods in matrix calculation. *Ann. Math. Statist.* **14**, 1–34.

Hotelling, H. 1943*b* Further points on matrix calculations and simultaneous equations. *Ann. Math. Statist.* **14**, 440–441.

Jeffreys, H. 1939 *Theory of probability*. Oxford: Clarendon Press.
Jeffreys, H. & Jeffreys, B. S. 1946 *Methods of mathematical physics*. Cambridge University Press.
Kryloff, N. 1926 *Ann. Fac. Sci., Toulouse*, pp. 1–34.
Kryloff, N. 1929 *Bull. Acad. Sci. U.R.S.S.* pp. 457–470 (in French).
Lamb, H. 1924 *Infinitesimal calculus*. Cambridge University Press.
MacDuffee, C. C. 1933 *The theory of matrices*. Berlin: Springer.
Newing, S. T. 1941 Shearing stresses in axially symmetrical shafts. *Phil. Mag.* (7), **32**, 33–49.
Rayleigh, Baron 1894 *The theory of sound*. London: Macmillan.
Richardson, L. F. 1910 The approximate arithmetical solution by finite differences of physical problems. *Phil. Trans.* A, **210**, 307–357.
Richardson, L. F. & Gaunt, J. A. 1927 The deferred approach to the limit. *Phil. Trans.* A, **226**, 299–361.
Richardson, L. F. 1939 Generalized foreign politics. *Brit. J. Psychol. Monogr. Suppl.* no. 23.
Richardson, L. F. 1947 *Arms and insecurity*, on 35 mm. microfilm. From the author. New edition 1949.
Richardson, L. F. 1950 A method for computing principal axes. *Brit. J. Psychol.* **3**, 16–20. (*Statistical Section.*)
Runge, C. 1908–9 Über eine Methode die partielle Differential-gleichung $\Delta u =$ Konstante numerisch zu integrieren. *Z. Math. Phys.* **56**, 225–232.
Southwell, Sir Richard 1946 *Relaxation methods in theoretical physics*. Oxford: Clarendon Press.
Temple, G. & Bickley, W. G. 1933 *Rayleigh's principle*. Oxford University Press.
Temple, G. 1939 The general theory of relaxation methods applied to linear systems. *Proc. Roy. Soc.* A, **169**, 476–500.
Thurstone, L. L. 1935 *The vectors of mind*. Chicago University Press.
Titchmarsh, E. C. 1947 *Theory of functions*. Oxford University Press.
Turnbull, H. W. & Aitken, A. C. 1932 *Theory of canonical matrices*. Glasgow: Blackie.
Weierstrass, K. 1895 *Werke*, **2**. Berlin: Mayer and Müller.
Whittaker, E. T. & Robinson, G. 1924 *The calculus of observations*. Glasgow: Blackie.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES