EXTENDED BOUNDARY VALUE PROBLEMS

By C. LANCZOS

1. Introduction

The topic to which the following discussions are devoted is as old as the history of quantitative thinking. It was observed from the time of antiquity that continuous phenomena can be approached from the viewpoint of treating them as the limits of discontinuous happenings. This age-old problem was revitalized in our own days and became once more of topical interest on account of the sensational development of the big electronic digital computers. With the ever-increasing memory-capacity of the new machines it becomes more and more possible to tackle many of the customary types of boundary value problems in a practical way. Partial differential equations of two or even three dimensions can be coded for the big machines, and we come nearer and nearer to the state in which the physicist or the engineer may get all the relevant answers for which he is striving by putting his problems on one of the high-powered electronic machines.

In this development a very definite *approximation procedure* plays a vital role. The given partial differential equation is changed to a *difference equation* and coded as a simultaneous set of algebraic equations. If, in particular, the given differential equation is of the *linear* type—and in all the following discussions we will restrict ourselves to the domain of linear operators—the resulting set of simultaneous linear equations can be characterized by the simple matrix equation

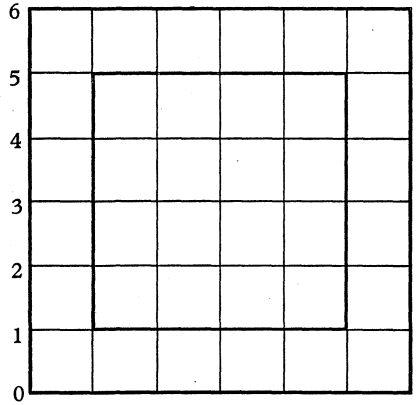
$$A\mathbf{y} = \mathbf{b}, \tag{1}$$

where the matrix A takes the place of the given linear differential operator, the vector $\mathbf{y}$ corresponds to the unknown function, while the given right side of the differential equation, including the given boundary conditions, is absorbed by the vector $\mathbf{b}$.

That this algebraization of a problem in linear differential equations is always possible is by no means self-evident. Anybody who has ever coded such a problem for the big machines will inevitably run into some questions which cannot be answered in a trivial way. Let us consider for the sake of illustration a somewhat over-simplified but characteristic example. Given the potential equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{2}$$

to be solved for a certain square-domain of the variables x and y. We set up a square grid of points and change the given differential equation into a difference equation. We observe at once that we obtain only $(n-2)^2$ equations for n^2 quantities, which demands $4n-4$ more data for a full algebraic characterization of our problem. Hence we have to add $4n-4$ boundary data. How shall we choose these data? The mathematician tells us that we will be wise if we add as boundary data the values of the



function ϕ in the grid-points along the four boundaries of our square. But we may have different ideas and tell him that we would prefer to choose our data along three sides only but omit the line 6, instead of which we want to give the functional values along the line 1. He will advise us strongly against such an idea. On the other hand, if the given differential equation happens to be

$$\frac{\partial^2\phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\phi}{\partial y^2} = 0 \tag{3}$$

he gives us exactly the opposite advice. If we try to understand these puzzling prescriptions, he refers us to the 'theory of characteristics' which, however, is not an algebraic theory, while our desire would be to understand the nature of differential equations purely from the *algebraic* point of view. This desire is not unjustified if it is true that a linear differential equation can be conceived as the limit of a set of linear algebraic equations, obtained by replacing the derivatives by difference coefficients. But then, why is it that the nature of the boundary value problem differs so completely in the elliptic and in the hyperbolic case,

although algebraically they seem to be equivalent? Before we come to the general discussion of such problems, a brief glance on the historical development of the subject will not be out of place.

2. Historical survey

The close relation between continuous and discrete operators was recognized from the very beginnings of higher mathematics. But the first example of a differential equation investigated in a consistently algebraic manner is perhaps a study of Daniel Bernoulli[1] concerning the motion of a perfectly flexible heavy string, suspended between two points. He starts with a chain, composed of two, three, four, and later an arbitrary number of links. 'Then', he says, 'by making the number of links infinite, I finally arrive at the oscillations of the completely flexible chain of either constant or variable thickness.' Later Lagrange[10], in his admirable studies of the propagation of sound, applied a similar method to the vibrations of a stretched string. He replaced the continuous manifold of points by a discrete set of points whose mutual distance could be made as small as we wish. We find the same basic idea in manifold manifestations in the works of Euler[4], who based his entire theory of differential calculus on the theory of difference equations with gradually decreasing increments. Thus he derived the fundamental differential equation of variational calculus by replacing the variational integral by a sum, and the derivatives of the unknown functions by difference coefficients. In the new form the problem became an ordinary maximum–minimum problem which could be solved by the tools of elementary calculus. He was not aware that this method involves the exchange of two limit processes which demands specific justification.

However, perhaps the greatest and most consistent exponent of the algebraic method was Lord Rayleigh[12] in his investigations of acoustic and elastic vibrations (of the years 1877–94). He gained deep insight into the nature of orthogonal function systems by his algebraic method, and it was in fact this method which led him to the discovery of the fundamental properties of orthogonal expansions. Even in his time the discretization of continuous operators was performed without any qualms of conscience, as a matter of 'physical intuition'. He starts out with the following general remarks in the introductory chapter of his great researches on vibrating systems:† 'Strictly speaking, the displacements possible to a natural system are infinitely various, and

† Cf. [12], vol. 1, chap. IV, p. 91.

cannot be represented as made up of a finite number of displacements of specified type. To the elementary parts of a solid body any arbitrary displacements may be given, subject to conditions of continuity. It is only by a process of abstraction of the kind constantly practised in Natural Philosophy, that solids are treated as rigid, fluids as incompressible, and other simplifications introduced so that the position of a system comes to depend on a finite number of co-ordinates. It is not, however, our intention to exclude the consideration of systems possessing infinitely various freedom, on the contrary, some of the most interesting applications of the results of this chapter will be in that direction. But such systems are most conveniently conceived as limits of others, whose freedom is of a more restricted kind. We shall accordingly commence with a system, whose position is specified by a finite number of independent co-ordinates $\psi_1, \psi_2, \psi_3, \ldots$, etc.'

With Lord Rayleigh we come to the turn of the century and it was exactly around that time that a new epoch of mathematical rigor takes its departure, with the classical investigations of Fredholm (1900–3) concerning the theory of a certain type of functional equations, now called 'integral equations of the Fredholm type'[6]. Fredholm tackles the problem on an algebraic basis and arrives at his results by a rigorous estimation of infinite determinants. This was the first time that the algebraization of infinitesimal processes was carried through to its final consequences with full mathematical rigor. In the dazzling light of this new approach the previous algebraic attempts were put in the balance and found wanting. While the results of Fredholm remained above all reproach, it was pointed out that similar results cannot be expected if we depart from the Fredholm type of integral equations.† The direct algebraization of differential equations was looked upon with suspicion and admitted without reservation only if the given problem could first be transformed into an integral equation of the Fredholm type.

This demand restricts, however, the type of boundary value problems admitted to an unnecessary degree. It includes only those problems which in the algebraic formulation are characterized by an $n \times n$ matrix (which means that the number of equations and unknowns must be equal), the determinant of which is different from zero. Beyond this requirement, however, the demand of the existence of a Green's function (in other words the *existence of the inverse operator*), restricts our possibilities still further.‡ Hadamard[7], in his lectures on the Cauchy

† Cf. [8], p. 1344.

‡ Cf. [8], p. 1362; [2], p. 358.

problem, called this type of boundary value problems 'well-posed', or 'correctly set' ('un problème correctement posé'). A problem is eligible to this distinction if the following two conditions are satisfied:

(1) The given data are sufficient to obtain one and only one solution.

(2) Within a certain general class of functions the given data can be prescribed freely.

Without impinging in the least on the importance of these problems, we can hardly doubt that these requirements handicap our possibilities quite severely. In the first place, how shall we decide in a given case whether a given problem is well-posed or not? This requires a very elaborate preliminary investigation of the given differential equation. Our present knowledge goes hardly beyond the realm of second-order operators. There we have the three types of elliptic, parabolic, and hyperbolic equations, and we know that a 'well-posed' problem demands in the case of elliptic differential equations the prescription of boundary conditions, in the other two cases the prescription of initial conditions. But how much or how little we should prescribe in the case of differential equations of third or fourth or higher order for the sake of a well-posed problem, it is impossible to tell.

Apart from this difficulty, we encounter well-defined and reasonable problems which do not fall in the well-posed category. Consider the case of a conservative field of force, characterized by the equation

$$\operatorname{grad} \phi = \mathbf{F}. \tag{4}$$

Here the scalar field ϕ is transformed into the vector field $\mathbf{F}$. If our aim is to obtain ϕ by observing $\mathbf{F}$, we have clearly an over-determined problem which is not solvable if $\mathbf{F}$ is freely prescribed, which is solvable, however, if $\mathbf{F}$ satisfies the compatibility condition

$$\operatorname{curl} \mathbf{F} = 0. \tag{5}$$

The algebraic picture associated with this problem involves a matrix A of n rows and m columns in which $n > m$.

On the other hand, consider the differential equation

$$\operatorname{div} \mathbf{E} = \rho, \tag{6}$$

where a vector field $\mathbf{E}$ is transformed into a scalar field ρ. The vector field $\mathbf{E}$ is by no means determined by this equation but we would like to know what conclusions can be drawn from the fact that ρ is given. Here we have an example of an under-determined system, algebraically characterized by a matrix A of n rows and m columns in which $n < m$.

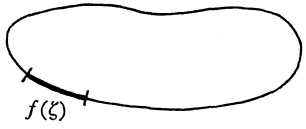
One of the most fundamental equations of the theory of analytical functions is Cauchy's integral theorem

$$f(z) = \frac{1}{2\pi i}\oint \frac{f(\zeta)\,d\zeta}{\zeta - z}, \tag{7}$$

which determines the value of $f(z)$ inside a domain if it is given on the boundary of the domain. The corresponding theorem in the theory of the Newtonian potential is the equation

$$\phi(P) = \frac{1}{4\pi}\int_S \left(\frac{\partial \phi}{\partial n}\frac{1}{r_{PS}} - \phi\frac{\partial}{\partial n}\frac{1}{r_{PS}}\right) dS,$$

by which the function ϕ can be evaluated at the inside point P, if the values of ϕ and $\partial\phi/\partial n$ are given on the boundary S of the domain. Conceiving these problems as boundary value problems, both theorems suffer from the fact of over-determination. The potential function ϕ is uniquely determined by the boundary values $\phi(S)$ *alone*, without giving the values of $(\partial\phi/\partial n)\,(S)$. In the case of the analytical function $f(z)$ it is unnecessary to give $f(\zeta)$ all along the boundary; it suffices if $f(\zeta)$ is given along an arbitrary *open* portion of the boundary and we are entitled to ask the question



(8)

how to obtain $f(z)$ in the inside of the domain in terms of these data. The problem is not of the well-posed type and has no elementary solution but we know that the solution exists and we are entitled to pose the problem.

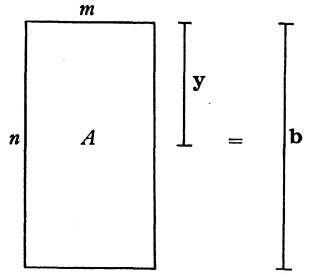
We now ask quite generally the following question. Given a linear partial differential equation or any system of such equations with added boundary conditions which are chosen in an arbitrarily judicious or injudicious fashion, thus leading to an arbitrarily over-determined or under-determined system. Can we treat such a problem successfully by an algebraic method, and if so, can we transfer the results without difficulty to the field of continuous operators?

From the historical standpoint this problem is not in line with the Fredholm type of investigations since the method of determinants loses its significance if we depart from the realm of square matrices to the realm of the more general $n \times m$ matrices. It so happens, however, that

the stakes of the algebraic method are much more widely set than the theory of determinants. If we follow up the later development of Fredholm's theory, we come to the classical investigations of Hilbert[9] who also employed the algebraic method for the development of the theory of integral equations (1904–10) but from an entirely different viewpoint. In Hilbert's theory the emphasis is laid on the geometry of second-order surfaces which are put in a Euclidean space of increasingly many dimensions. The principal axis transformation of these surfaces became the central item in Hilbert's theory. This geometrical method gave a second and independent rigorous foundation of the theory of integral equations, without recourse to the theory of determinants. The same theory was later put on a more analytical basis by E. Schmidt[13]. It is this principal axis theory of quadratic forms which provides the proper frame of reference for our much more general problem and which yields a suitable universal platform for the understanding of the behaviour of both well-posed and not well-posed boundary value problems.

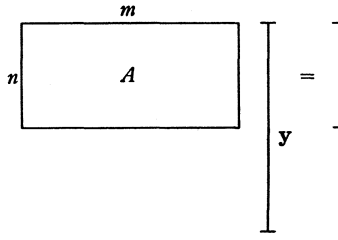
3. The fundamental eigenvalue problem

We start with the algebraic equation (1), assuming A as a general $n \times m$ matrix, with $n \gtrless m$. The case $n > m$ can be pictured as follows:



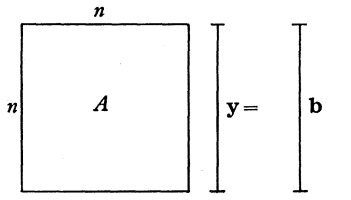
over-determined
$n > m$
$(\text{grad}\,\phi = \mathbf{F})$ (9)

The case $n < m$ can be pictured as follows:



under-determined
$n < m$
$(\text{div}\,\mathbf{E} = \phi)$ (10)

The case $n = m$ (with the added condition $\det A \neq 0$) belongs to the usual 'well-posed' case:



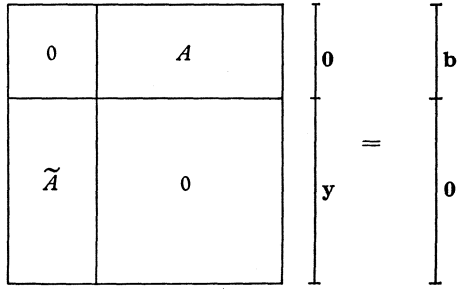
(11)

Our diagrams demonstrate by inspection an important feature of our problem. Usually we have the $n \times n$ case in hand and consider a matrix A as an operator which transforms a vector $\mathbf{y}$ into another vector $\mathbf{y}'$. But the general case $n \neq m$ shows that the case $n = m$ hides an important feature of our problem, viz. that the two vectors $\mathbf{y}$ and $\mathbf{b}$ belong to *two different spaces*. The vector $\mathbf{y}$ on which A operates belongs to an m-dimensional space, the vector $\mathbf{b}$ into which $\mathbf{y}$ is transformed, to an n-dimensional space. Hence a general matrix A takes a vector from one space and transforms it into a vector of another space. It is necessary that in all our following discussions we should keep the separateness of these two spaces clearly in mind.†

The fundamental first step, from which everything will follow with logical necessity, is an apparent triviality. We extend the basic equation (1) by a second equation and consider the resulting system of a single unit:

$$A\mathbf{y} = \mathbf{b}, \quad \tilde{A}.\mathbf{0} = \mathbf{0}. \tag{12}$$

This extension is reflected in the following matrix diagram:



(13)

† Without this distinction serious misunderstandings are prone to happen; e.g. the so-called 'integral equations of the first kind', are often declared analytically inferior to the Fredholm type of equations (see, for example, [8], p. 1453), solely because the function of the left side and the transformed function of the right side do not belong to the same class of functions.

We notice that we have extended our matrix A to a new $(n+m) \times (n+m)$ square matrix S:

$$S = \begin{array}{c|cc} & n & m \\ \hline n & 0 & A \\ m & \tilde{A} & 0 \end{array} \tag{14}$$

This matrix is not only square but even *symmetric*:

$$S = \tilde{S}. \tag{15}$$

We know that symmetric square matrices have particularly desirable properties. In particular, we know that such a matrix can always be *diagonalized* by a proper orthogonal transformation. This requires the determination of the principal axes of the matrix, on the basis of the equation

$$S\mathbf{w} = \lambda \mathbf{w}. \tag{16}$$

The vector $\mathbf{w}$ has $n+m$ components which we will record in the following form:

$$\mathbf{w} = \begin{array}{|c} \mathbf{x} \\ \\ \mathbf{y} \end{array} \tag{17}$$

In view of the specific structure of S the principal axis problem will exhibit some special features. Indeed, the basic equation (16) can be formulated in terms of the matrix A as follows:

$$A\mathbf{y} = \lambda \mathbf{x}, \quad \tilde{A}\mathbf{x} = \lambda \mathbf{y}. \tag{18}$$

We will call this system the 'shifted eigenvalue problem', because the

customary eigenvalue problem† of a square matrix is defined by the equations

$$A\mathbf{y} = \lambda\mathbf{y}, \quad \tilde{A}\mathbf{x} = \lambda\mathbf{x}, \tag{19}$$

while in our case—which remains meaningful for the general case of an arbitrary $n \times m$ matrix—the position of the vectors $\mathbf{x}$ and $\mathbf{y}$ on the right sides is *reversed.*

If we multiply the second equation (18) by A, we see at once that the $\mathbf{x}$-vectors in themselves can be defined as the solution of the eigenvalue problem

$$A\tilde{A}\mathbf{x} = \lambda^2\mathbf{x}, \tag{20}$$

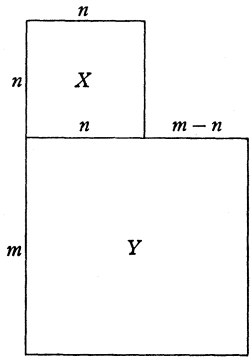
while multiplication of the first equation by $\tilde{A}$ shows that similarly the $\mathbf{y}$ vectors in themselves can be defined as the solution of the eigenvalue problem

$$\tilde{A}A\mathbf{y} = \lambda^2\mathbf{y}. \tag{21}$$

Since, however, the matrices $A\tilde{A}$ and $\tilde{A}A$ are in themselves respectively symmetric $n \times n$ and $m \times m$ matrices, we see that the $\mathbf{x}$ vectors themselves are mutually orthogonal to each other, and so are the $\mathbf{y}$ vectors:

$$\mathbf{x}_i . \mathbf{x}_k = \delta_{ik}, \quad \mathbf{y}_i . \mathbf{y}_k = \delta_{ik}. \tag{22}$$

Moreover, the vectors $\mathbf{x}_i$, if plotted as columns, fill out a full $n \times n$ space, the vectors $\mathbf{y}_j$ a full $m \times m$ space. Hence we can include the solution of the given eigenvalue problem (18) in the following matrix diagram:



(23)

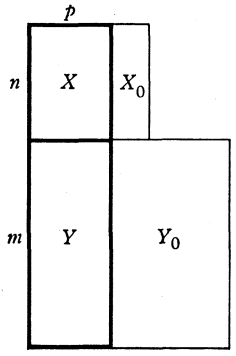
Although the two spaces X and Y are separated, there is a correlation between them on the basis of the equation

$$\mathbf{x}_i = \frac{A\mathbf{y}_i}{\lambda_i}, \tag{24}$$

† Cf. [5], p. 64.

which shows that to every $\mathbf{y}_i$ vector a corresponding $\mathbf{x}_i$ vector can be found. And yet our diagram shows that this pairing between the two kinds of vector cannot hold generally if n and m are not equal. In our illustration $m > n$ and $m-n$ $\mathbf{y}_i$ vectors remain unpaired. But the relation (24) breaks down only for $\lambda_i = 0$. Thus we see that the eigenvalue $\lambda = 0$ must be present at least $m-n$ times. In actual fact the multiplicity of the zero eigenvalue may be still larger since it is possible that even some of the $\mathbf{x}_i$ vectors belong to the eigenvalue $\lambda = 0$.

It will now be our aim to *separate the zero eigenvalue from the non-vanishing eigenvalues*. We will thus bracket out all those principal axes $\mathbf{x}_i$ and $\mathbf{y}_i$ which are truly paired on the basis of a λ_i which is not zero. The matrix of these $\mathbf{x}_i$ vectors shall be denoted by X, the matrix of the corresponding $\mathbf{y}_i$ vectors by Y, while the subspaces associated with the zero eigenvalue shall be denoted by X_0 and Y_0.



(25)

We see here the emergence of a new integer, say p, which characterizes the number of $\mathbf{x}_i$, $\mathbf{y}_i$ vectors included in the spaces X and Y. The matrix X is an $n \times p$ orthogonal matrix

$$\tilde{X}X = I, \tag{26}$$

the matrix Y an $m \times p$ orthogonal matrix

$$\tilde{Y}Y = I. \tag{27}$$

(The products $X\tilde{X}$ and $Y\tilde{Y}$, however, need not be equal to I, because generally, if $p < \frac{n}{m}$, the matrices X and Y do not fill out their spaces). The multiplicity of the zero eigenvalue is now

$$(n-p)+(m-p) = n+m-2p.$$

The total number of eigenvalues must become equal to $n+m$, since the matrix S has $n+m$ rows and columns. Consequently, we get for the number of non-zero eigenvalues:

$$n+m-(n+m-2p) = 2p.$$

Why does our diagram display only p instead of $2p$ axes? The reason is that to every solution

$$(\mathbf{x}_i, \mathbf{y}_i, \lambda_i)$$

a second solution can be constructed, namely

$$(\mathbf{x}_i, -\mathbf{y}_i, -\lambda_i).$$

Hence all non-zero eigenvalues appear with both plus and minus signs. We will agree to omit all the negative eigenvalues and keep only the positive ones, since the negative eigenvalues do not add anything new to the eigenvalue problem. We can thus characterize p as the *number of positive eigenvalues* for which the shifted eigenvalue problem (18) is solvable. It can assume any value between 1 and the smaller of the two numbers n and m:

$$1 \leqslant p \leqslant \min\{n, m\}. \tag{28}$$

This number p coincides in fact with the 'rank' of the matrix A. We see that this fundamental number, which in the usual algebraic theory of Kronecker and Frobenius is defined on a completely different basis,† enters our eigenvalue problem again as a quantity of decisive importance.

4. The fundamental decomposition theorem

Apart from the two matrices X and Y we construct a diagonal matrix Λ which contains in the diagonal all the positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, for which the shifted eigenvalue problem (18) is solvable:

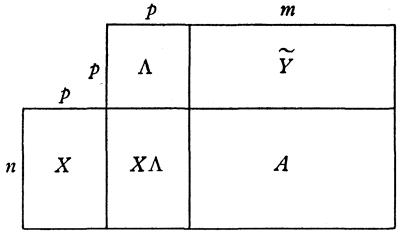
$$\Lambda = \begin{bmatrix} \lambda_1 & . & . & . & . & . \\ . & \lambda_2 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & \lambda_p \end{bmatrix}.$$

Then the principal axis transformation of the matrix S leads to the

† Cf. [11], p. 10.

following fundamental decomposition theorem which holds without any exception for any matrix A which does not vanish identically:

$$A = X\Lambda\tilde{Y}. \tag{29}$$



The most remarkable feature of this decomposition is that the operator A *completely by-passes the zero fields* X_0, Y_0. Only those principal axes of S are needed for the construction of A, which belong to positive eigenvalues. Hence we will call the columns of the matrices X and Y the 'essential axes' of our problem. It will be possible to formulate the entire theory of solving the basic equation (1) in terms of the matrices X and Y, without any reference to the fields X_0, Y_0 which are associated with the eigenvalue $\lambda = 0$.†

5. Solution of the basic equation

We return to our original problem of solving the matrix equation (1). We will, however, change our notation by denoting the right side of the equation by $\mathbf{x}$:

$$A\mathbf{y} = \mathbf{x}. \tag{30}$$

Then, substituting for A the product (29) we obtain

$$X\Lambda\tilde{Y}\mathbf{y} = \mathbf{x}. \tag{31}$$

We transform our variables according to the law

$$\mathbf{x} = X\mathbf{x}', \quad \mathbf{y} = Y\mathbf{y}', \tag{32}$$

† Professor A. S. Householder called the author's attention to a paper of E. G. Kogbetliantz, 'Diagonalization of general complex matrices as a new method for solution of linear equations', *Proceedings of the International Congress of Mathematicians*, (North-Holland, Amsterdam, 1954), II, p. 356, which describes a numerical process for the diagonalization of an arbitrary matrix, on the basis of the equation $U^*AV = D$ (which in our notation becomes $\tilde{U}AV = D$ since it is tacitly understood that 'transposition' in the presence of complex elements includes the change of i to $-i$). No reference is made to the 'shifted eigenvalue problem' (18), nor to the decomposition theorem (29), (in which the zero-field is clipped away).

and obtain for the new variables the relation

$$\Lambda \mathbf{y}' = \mathbf{x}', \tag{33}$$

which can be solved in the form

$$\mathbf{y}' = \Lambda^{-1}\mathbf{x}'. \tag{34}$$

This solution is always possible because the diagonal matrix can have no vanishing elements in the diagonal. But premultiplication of (32) by $\tilde{X}$, respectively $\tilde{Y}$, gives

$$\mathbf{x}' = \tilde{X}\mathbf{x}, \quad \mathbf{y}' = \tilde{Y}\mathbf{y}, \tag{35}$$

and thus we obtain

$$\mathbf{y} = Y\Lambda^{-1}\tilde{X}\mathbf{x}. \tag{36}$$

This solution gives the impression that *every* linear system has a solution and, in fact, a *unique* solution, which can hardly be expected of arbitrarily over-determined or under-determined systems. But actually this solution depended on the assumption (32) which is equivalent to the statement that the vector $\mathbf{x}$ is inside the X-space, the vector $\mathbf{y}$ inside the Y-space. Let us first consider the latter statement.

The operator A operates solely in the subspaces X and Y. We could describe the situation by imagining that in the matrix diagram (25) the fields X, Y are illuminated while the fields X_0, Y_0 remain entirely in the dark. Now the vector $\mathbf{y}$ can have a projection into Y_0 as well as a projection into Y. However, the given equation determines solely the projection into Y but leaves the projection into Y_0 completely undetermined. Under these circumstances it seems natural that we place $\mathbf{y}$ completely into the space Y and leave the determination of the projection into Y_0 to some additional information. The second equation of the system (32) can thus be conceived as a natural *normalization* of our solution, by putting it into the *space of the least number of dimensions* which is able to hold it.

While the second equation of (32) can be conceived as a matter of choice, this is not so with the first equation. The fact that the left side $\mathbf{x}$ is inside the space X, is not a matter of choice but a consequence of the given equation. If the given left side $\mathbf{x}$ does not satisfy this condition, then the given system is self-contradictory and thus unsolvable. We can thus conceive the condition.

$$\mathbf{x} = X\mathbf{x}', \tag{37}$$

as the *compatibility condition* of the given system which is necessary and sufficient for the existence of a solution. Hence we see that the two

conditions (32) solve the problem of over-determination and under-determination. The first condition expresses the compatibility of the given system in the case of over-determination, the second condition normalizes the solution in the case of under-determination. The uniqueness of the solution (36) is thus explained.

The compatibility conditions (37) can be expressed in various equivalent forms. We can put it, for example, in the form of an *orthogonality condition*, expressing the orthogonality of the vector $\mathbf{x}$ to the space X_0:

$$\tilde{X}_0 . \mathbf{x} = 0. \tag{38}$$

This leads to the traditional formulation of the compatibility condition of an arbitrarily given linear system: 'A given linear system of equations is solvable if and only if the right side of the system is orthogonal to every independent solution of the adjoint homogeneous system.'

It is more adequate, however, to avoid any reference to the space X_0, in which the operator A is not active. We can stay completely within the confines of the space X and express the compatibility condition (37) in the form

$$\mathbf{x} = X\tilde{X}\mathbf{x}. \tag{39}$$

But we can go still further and include all the compatibility conditions of an arbitrary linear system into one single *scalar condition*. If the vector $\mathbf{x}$ falls completely within X, then the length of the vector and the length of its projection into this space become equal:

$$\mathbf{x}^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_p^2, \tag{40}$$

where

$$\xi_i = \mathbf{x}_i . \mathbf{x}. \tag{41}$$

The converse of the theorem is equally true. The *scalar condition* (40) can thus be considered as the *necessary and sufficient condition for the solvability of the system* (30).

Transition into the realm of differential operators. It will now be our aim to translate these algebraic results into the domain of continuous operators, particularly *differential* operators. And here we encounter first of all the following deviations from the algebraic case: (1) The matrix $\tilde{A}$ is defined by a transposition of rows and columns. This operation does not allow a direct interpretation in the realm of differential equations. (2) A differential equation is usually associated with certain *boundary conditions*. We have to find a way of associating the matrix operators A and $\tilde{A}$ with the given differential operator, *plus* the proper boundary conditions. (Infinite domains are excluded from our considerations.)

We find the answer to these questions by considering the so-called 'bilinear identity'

$$\mathbf{x}.A\mathbf{y}-\mathbf{y}.\tilde{A}\mathbf{x}\equiv 0, \tag{42}$$

which holds for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$. If for a given matrix A we succeed in finding a matrix B which for arbitrary $\mathbf{x}$ and $\mathbf{y}$ satisfies the identity

$$\mathbf{x}.A\mathbf{y}-\mathbf{y}.B\mathbf{x}\equiv 0, \tag{43}$$

then we know that

$$B=\tilde{A}. \tag{44}$$

Hence we can define $\tilde{A}$ as that particular matrix B which satisfies the identity (43).

Now in the theory of linear differential equations the bilinear identity (42) appears in the form of 'Green's identity':

$$\int (vDu-u\tilde{D}v)\,d\tau = \text{boundary integral}, \tag{45}$$

which is always obtainable by the method of integrating by parts. The notation D refers to an arbitrary given linear differential operator (ordinary or partial), or systems of such operators (such as for example the left sides of the Cauchy–Riemann equations, or Maxwell's equations, etc.).

In addition to the differential equation

$$Du=\rho$$

some more or less stringent *boundary conditions* may be prescribed for u and its derivatives on the boundary of the domain. These conditions may generally be of the homogeneous or inhomogeneous type. We will agree, however, to replace any inhomogeneous boundary condition by the corresponding *homogeneous* condition. This is always possible by the device of replacing the original unknown function $\overline{u}$ by the sum

$$\overline{u}=u_0+u, \tag{46}$$

where u_0 is chosen as some function which absorbs the given inhomogeneous boundary conditions, without satisfying, however, any differential equation. Then, considering u as the new unknown, we obtain the differential equation

$$Du=\rho-Du_0, \tag{47}$$

plus boundary conditions which are now of the *homogeneous* type; (the right side being equal to *zero*, instead of some prescribed values).†

Together with the given operator Du we will consider the 'adjoint (transposed) operator' $\tilde{D}v$ which is likewise augmented by suitably

† After obtaining our result for u it is not difficult to return to the original function $\overline{u}$ and formulate the results in terms of the original inhomogeneous boundary values (with $\rho=0$ in most cases). The explicit carrying out of the substitution (46) is thus seldom demanded.

chosen boundary conditions. The definition of $\tilde{D}v$ follows from the expression on the left side of Green's identity (45). The definition of the 'adjoint boundary conditions' follows from a careful study of the boundary integral on the right side of Green's identity (45). We prescribe for v (and its derivatives) the minimum number of boundary conditions which are necessary and sufficient to make the boundary integral vanish at all points of the boundary. These conditions are once more of the homogeneous kind. The more over-determined the original problem was, the more under-determined is the adjoint problem, and vice versa.

If we now consider the 'shifted eigenvalue problem'

$$Du = \lambda v, \quad \tilde{D}v = \lambda u, \tag{48}$$

where u is subjected to the given, v to the adjoint boundary conditions, this problem has always the right degree of determination and allows an infinity of possible solutions for an infinite—but discrete—set of eigenvalues λ_i, among which we may find $\lambda = 0$ represented with a finite or infinite multiplicity. We *drop* these latter solutions and keep only the solutions which belong to the non-vanishing (and even positive) λ_i. We thus obtain an infinite set of orthogonal (and normalized) eigenfunctions $u_i(\tau)$ and a corresponding infinite set of orthogonal (and normalized) eigenfunctions $v_i(\sigma)$. These two sets of functions operate generally in two separate portions of Hilbert space, although they belong to the same domain of the independent variables. There is, however, a one-to-one correspondence between these two sets of functions, on account of the relation

$$v_i(\tau) = \frac{Du_i(\tau)}{\lambda_i}. \tag{49}$$

6. Solution of the differential equation $Du = v$

We will now consider the solution of the differential equation

$$Du = v, \tag{50}$$

where u is subjected to the given homogeneous boundary conditions. Although we have dropped the eigenfunctions associated with the eigenvalue zero, the remaining functions $u_i(\tau)$ and $v_i(\tau)$ are still sufficiently complete to serve as base vectors for the representation of the functions $u(\tau)$, respectively $v(\tau)$. Hence we can put

$$\left.\begin{aligned} u(\tau) &= \sum_{i=1}^{\infty} c_i u_i(\tau), \\ v(\tau) &= \sum_{i=1}^{\infty} \gamma_i v_i(\tau). \end{aligned}\right\} \tag{51}$$

The differential equation (50) establishes the following relation between the two sets of coefficients:

$$c_i = \frac{\gamma_i}{\lambda_i}. \tag{52}$$

Our equation (50) is thus solvable by the infinite expansion

$$u(\tau) = \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_i} u_i(\tau), \tag{53}$$

where

$$\gamma_i = \int v(\sigma)\, v_i(\sigma)\, d\sigma. \tag{54}$$

However, in view of the fact that the functions $v_i(\tau)$ are generally *not* complete (since we have dropped the $v_j^{(0)}$ which belong to $\lambda = 0$), it is necessary to test the given function $v(\tau)$ concerning compatibility. The solvability of the equation (50) demands that $v(\tau)$ is completely within the subspace of the $v_i(\tau)$. This means

$$\int v^2(\sigma)\, d\sigma = \sum_{i=1}^{\infty} c_i^2. \tag{55}$$

This 'completeness relation'—which corresponds to the previous algebraic relation (40)—represents the compatibility condition demanded of the right side of the differential equation (50) in the case of an over-determined system.

7. The two kinds of boundary value problems

If we approach our problem from the previously pursued algebraic angle by considering the given differential equation (50) as the limit of an algebraic set of equations, our general expectations will be as follows. In view of the arbitrarily judicious or injudicious choice of boundary conditions we are confronted with a system which may be arbitrarily over-determined or under-determined. As far as under-determination goes, the uncertainty of the solution is eliminated by a natural normalization of our solution, viz. by putting the solution into that u-space in which the operator D is activated. As far as over-determination goes, we have to test the right side of the equation whether or not it satisfies the condition that it is completely contained in that v-space in which the operator D is activated. If this condition is not satisfied, no solution is possible. If this condition is satisfied, a unique solution of the given boundary value problem is obtained. We do not see any further complications which may arise.

However, limit processes have their own intricacies and we know

that unexpected things can happen because in the limit something may occur which did not occur any time during the limit process. For example the limit of a continuous set of functions may be a discontinuous function. A closer analysis reveals that something similar is at work in our problem.

Let us plot the entire λ_i-spectrum on the positive half-line, from zero to infinity. We have omitted the zero eigenvalue as not included by the operator. And yet, the zero eigenvalue may come into evidence in a more subtle manner. It is possible that our eigenvalue spectrum does not start with a certain finite $\lambda_1 = \epsilon$ but that $\lambda = 0$ is a *limit point*, which means that there are infinitely many eigenvalues which, although discrete, come to zero as close as we wish. These eigenvalues cannot be omitted as not belonging to the operator. They do belong to the operator and their presence has a strong influence on the solution of our problem.

Under these circumstances we can put the entire class of possible boundary value problems into two categories, according to the presence or absence of the limit point $\lambda = 0$.

7.1. Boundary value problems of the first kind. This class of problems is characterized by the condition that $\lambda = 0$ *is not a limit point of the eigenvalue spectrum.* The traditional type of boundary value problems fall into this category. Within this class of problems we distinguish two subgroups:

(*a*) $\lambda = 0$ *is not included among the eigenvalues of the adjoint operator* $\tilde{D}$. This means that under the adjoint boundary conditions the equation

$$\tilde{D}v = 0 \tag{56}$$

has no non-vanishing solutions.

(*b*) $\lambda = 0$ *is included among the eigenvalues of the adjoint operator* $\tilde{D}$. This means that the equation (56) (under the adjoint boundary conditions) has a finite or infinite number of non-vanishing solutions.

The case (*a*). This is distinguished by the property that the given problem is solvable for arbitrarily prescribed $v(\tau)$,† without the demand of specific compatibility conditions. The solution can be given in the form of the infinite expansion (53). But the same solution may also be put in operational form, corresponding to the matrix solution (36) of the algebraic problem:

$$u(\tau) = \int G(\tau, \sigma)\, v(\sigma)\, d\sigma, \tag{57}$$

† The expression 'arbitrary' refers to an aribtrary function from the general class of 'functions of bounded variation'.

where $G(\tau, \sigma)$, the 'Green's function', is defined by the following infinite expansion:

$$G(\tau, \sigma) = \sum_{i=1}^{\infty} \frac{u_i(\tau)\, v_i(\sigma)}{\lambda_i}. \tag{58}$$

This expansion is a natural generalization of the well-known bilinear expansion† of a symmetric kernel function $K(\tau, \sigma) = K(\sigma, \tau)$:

$$K(\tau, \sigma) = \sum_{i=1}^{\infty} \frac{u_i(\tau)\, u_i(\sigma)}{\lambda_i}. \tag{59}$$

As we know, this expansion is not always convergent, although convergence can be obtained by averaging over an arbitrarily small neighbourhood of the point σ. The same holds of the expansion (58). Moreover, while the expansion (58) itself may diverge, we get a convergent result if it is applied in (57) under the integral sign, integrating term by term.

The function $G(\tau, \sigma)$ may also be characterized by the solution of the differential equation

$$DG(\underline{\tau}, \sigma) = \delta(\underline{\tau}, \sigma), \tag{60}$$

where $\delta(\tau, \sigma)$ is Dirac's delta function (the underlining of τ indicates that we consider τ as the variable while σ is a mere parameter).

As a special case within the subgroup (a) we can go one step still further and demand that not only the adjoint homogeneous equation (56) but also the given homogeneous equation

$$Du = 0 \tag{61}$$

(under the given boundary conditions) shall have no non-vanishing solutions. Then the solution given in the form (53), or in the alternate form (57), (58), is unique not merely by normalization but unconditionally. In the algebraic sense we now have the limiting case $n = m = p$. This most restricted group of boundary value problems corresponds to Hadamard's 'well-posed' problem.

The case (b). If the equation (56) possesses non-vanishing solutions

$$v_1^{(0)}(\tau), \quad v_2^{(0)}(\tau), \quad \ldots, \quad v_r^{(0)}(\tau) \tag{62}$$

we come to the next group of boundary value problems. We *exclude* the solutions $v_i^{(0)}(\tau)$ from our system of eigenfunctions $v_i(\tau)$, and thus lose the completeness of our function system (we do the same with the $u_j^{(0)}(\tau)$, if they exist, with respect to the system $u_i(\tau)$).

The solution (53) is once more valid, and it is also possible to give the solution in terms of the Green's function (57), (58). This function is

† Cf. [2], p. 134.

occasionally called the 'generalized Green's function'† because, although the bilinear expansion (58) retains its form without any change, the defining differential equation (60) is now to be modified to

$$\left.\begin{aligned} DG(\underline{\tau}, \sigma) &= \delta(\underline{\tau}, \sigma) - \sum_{i=1}^{r} v_i^{(0)}(\underline{\tau})\, v_i^{(0)}(\sigma), \\ \tilde{D}G(\tau, \underline{\sigma}) &= \delta(\tau, \underline{\sigma}) - \sum_{i=1}^{s} u_i^{(0)}(\tau)\, u_i^{(0)}(\underline{\sigma}). \end{aligned}\right\} \tag{63}$$

(The numbers r and s—which are completely independent of each other —may be finite or infinite. If $s = 0$, the solution is unique not only by normalization, but unconditionally.)

The only important difference between the cases (a) and (b) is that the function $v(\tau)$ can no longer be chosen freely but has to satisfy the compatibility conditions

$$\gamma_i^{(0)} = \int v(\sigma)\, v_i^{(0)}(\sigma)\, d\sigma = 0 \quad (i = 1, 2, \ldots, r). \tag{64}$$

All these conditions are replaceable by one single scalar condition, viz. the 'completeness relation' (55), which expresses the fact that $v(\tau)$ is inside that subspace of the function space which is spanned by the $v_i(\tau)$ alone, without the $v_j^{(0)}(\tau)$.

7.2. Boundary value problems of the second kind. This class of boundary value problems is characterized by the property that $\lambda = 0$ *is a limit point of the eigenvalue spectrum.* These problems fall outside the realm of the traditional type of boundary value problems. We encounter them if an elliptic differential equation is characterized by hyperbolic type of boundary conditions, or vice versa, or if, for example, the heat conduction equation is characterized by end-values instead of initial values. The problem indicated in figure (8) falls likewise into this category.

These problems are solvable but require a more circumspect approach than that employed in the previous class of problems. The characteristic feature of the new eigenvalue problem is the *unusual distribution of the eigenvalues* λ_i. This feature is deeply interwoven with a fundamental question that concerns the general structure of the function space. The concept of the 'function space' envisages a Euclidean space of infinitely many dimensions. The various orthogonal function systems associated with self-adjoint differential operators can be conceived as various

† Cf. [2], p. 356.

systems of orthogonal base vectors which in principle are all equivalent to each other. And yet the complete homogeneity of a Euclidean space in every direction does not correspond to the actual structure of the function space. In a Euclidean space the sequence in which we arrange our co-ordinate axes is entirely immaterial. In function space a definite *ordering* of the axes is demanded, in view of the fact that the infinite expansion into eigenfunctions has to *converge* to a definite limit $f(\tau)$. 'Convergence' by its very definition means that the terms of high order contribute negligibly small amounts to the expansion. This, however, assumes that we have arranged our functions properly, namely in order of decreasing significance. From where do we obtain such an ordering principle in a Euclidean space which is homogeneous in every direction?

We are used to the Sturm–Liouville type of eigenfunction problems† in which the ordering of the eigenfunctions is quite systematic and determined by the *magnitude of the eigenvalues* λ_i. By putting the λ_i in increasing order: $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \ldots$, we obtain a natural ordering principle for the associated eigenfunctions $\phi_1(x), \phi_2(x), \phi_3(x), \ldots$. This principle carries over into the realm of partial differential equations if we deal with boundary value problems of the conventional type. In our present problem, however, we encounter a situation which is not of the conventional type. First of all, if $\lambda = 0$ is a limit point of the eigenvalue spectrum, the ordering of the λ_i in increasing order is no longer possible. But this ordering would not even be justified since it is no longer true that the orthogonal functions of primary importance belong to the smallest eigenvalues. We experience a peculiar 'inversion of eigenvalues', due to which certain eigenfunctions which should appear quite *late* in order of importance, appear in fact *very early* in the λ_i spectrum. For this reason we will speak of a 'residual spectrum', to which we will relegate all the eigenvalues (and associated eigenfunctions) which cluster around the value zero and which in fact represent eigenfunctions of high order.

In view of this situation we will establish the following procedure for a definite ordering of the λ_i. We prescribe a certain arbitrarily small ϵ and put the eigenvalues into two categories:

Group 1: all the eigenvalues $\lambda_i \geqslant \epsilon$, arranged in increasing order: $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \ldots$ (together with the corresponding $u_i(\tau)$, $v_i(\tau)$).

Group 2: all the eigenvalues (now denoted by λ_i') for which $\lambda_i' < \epsilon$, arranged in decreasing order: $\lambda_1' \geqslant \lambda_2' \geqslant \lambda_3' \geqslant \ldots$ (together with the corresponding $u_i'(\tau)$, $v_i'(\tau)$).

† Cf. [2], p. 291.

The solution of the equation (50) under the present circumstances occurs once more by the infinite expansions (51) if we pay attention to the proper ordering of the eigenfunctions:

$$v(\tau) = \sum_{i=1}^{\infty} \gamma_i v_1(\tau) + \sum_{j=1}^{\infty} \gamma'_j v'_j(\tau), \tag{65}$$

where

$$\gamma_i = \int v(\sigma)\, v_i(\sigma)\, d\sigma, \quad \gamma'_j = \int v(\sigma)\, v'_j(\sigma)\, d\sigma, \tag{66}$$

and

$$u(\tau) = \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_i} u_i(\tau) + \sum_{j=1}^{\infty} \frac{\gamma'_j}{\lambda'_j} u'_j(\tau). \tag{67}$$

In contradistinction to the previous type of problems, a solution in terms of the Green's function is no longer possible. We can define the kernel function

$$G(\tau, \sigma) = \sum_{i=1}^{\infty} \frac{u_i(\tau)\, v_i(\sigma)}{\lambda_i}, \tag{68}$$

but the corresponding function for the primed eigenfunctions does not exist. Hence we can give the solution only *partially* in terms of the Green's function; the addition of an infinite sum, extended over the residual spectrum, cannot be avoided:

$$u(\tau) = \int G(\tau, \sigma)\, v(\sigma)\, d\sigma + \sum_{i=1}^{\infty} \frac{\gamma'_i}{\lambda'_i} u'_i(\tau). \tag{69}$$

It is exactly this sum which represents the difference between the conventional and the unconventional type of boundary value problems. We observe that the division by very small λ'_i has the consequence that the sum on the right side (69) will generally *not converge*. Hence the function $v(\tau)$ cannot be prescribed arbitrarily. In order that $u(\tau)$ shall be quadratically integrable, it is necessary and sufficient that the following condition shall be satisfied:

$$\sum_{i=1}^{\infty} \left(\frac{\gamma'_i}{\lambda'_i}\right)^2 < \infty. \tag{70}$$

Beyond this condition, however, we will be generally obliged to demand the absolute convergence of the expansion coefficients:

$$\sum_{i=1}^{\infty} \left|\frac{\gamma'_i}{\lambda'_i}\right| < \infty. \tag{71}$$

The convergence conditions (70) and (71) *restrict* the class of functions $v(\tau)$ for which the given boundary value problem is solvable, although this restriction is *less stringent* than that encountered in the class I, case (*b*) type of problems. There the conditions (64) demanded that all the expansion coefficients $\gamma_i^{(0)}$ which belonged to the eigenfunctions associated

with the eigenvalue zero must *vanish*. Now the demand is that the expansion coefficients γ_i' which belong to the eigenfunctions associated with *almost vanishing* eigenvalues, need not be zero but are *sufficiently small*.

The two subgroups (a) and (b) of the previous class of boundary value problems can once more be distinguished, with quite analogous results:

Case (a). *The value zero is not included among the eigenvalues of the adjoint operator* $\tilde{D}$. This case is covered by our foregoing discussions. The function $v(\tau)$ need not satisfy additional compatibility conditions, beyond the convergence conditions (70) and (71).

Case (b). *The value zero is included among the eigenvalues of the adjoint operator* $\tilde{D}$. Here the solvability of the given problem demands that the function $v(\tau)$ shall satisfy the additional completeness relation:

$$\int v^2(\sigma)\, d\sigma = \sum_{i=1}^{\infty} \gamma_i^2 + \sum_{j=1}^{\infty} \gamma_j'^2. \tag{72}$$

8. An alternative treatment

The same results can be obtained by a somewhat different approach in which the residual spectrum is brought in in the form of a *limit process*. We start again with a given ϵ and all the λ_i which are greater or equal to ϵ. They are once more ordered in increasing magnitude and once more we define the Green's function by the expansion (68). We indicate, however, that this function depends on ϵ:

$$G_\epsilon(\tau, \sigma) = \sum_{i=1}^{\infty} \frac{u_i(\tau)\, v_i(\sigma)}{\lambda_i}. \tag{73}$$

We now decrease ϵ to smaller and smaller values which means that $G_\epsilon(\tau, \sigma)$ goes more and more out of bound. However, while $G_\epsilon(\tau, \sigma)$ does not approach any limit, it is possible that the sequence of functions

$$u_\epsilon(\tau) = \int G_\epsilon(\tau, \sigma)\, v(\sigma)\, d\sigma \tag{74}$$

formed with the help of these $G_\epsilon(\tau, \sigma)$, approaches a definite limit:

$$u(\tau) = \lim_{\epsilon \to 0} u_\epsilon(\tau). \tag{75}$$

If this limit does not exist, the function $v(\tau)$ was not chosen from the class of permissible functions and the given problem has no solution. If the limit (75) does exist, the function $v(\tau)$ was given properly and the limit $u(\tau)$ represents the solution of the given boundary value problem.

9. Elastic vibrations

An interesting example of such a 'boundary value problem of the second kind', which at the same time has physical significance, is provided by the problem indicated in figure (8). Here the function $f(z)$ of the complex variable $z = x+iy$ is given along an open boundary only and our aim is to obtain $f(z)$ inside the given domain with the help of the boundary data. By putting

$$f(z) = \frac{\partial \phi}{\partial y} + i\frac{\partial \phi}{\partial x} \tag{76}$$

we can formulate our problem in the following alternative fashion. 'Given the potential function $\phi(x, y)$ and its normal derivative $\partial\phi/\partial n$ along an arbitrarily small open portion s of the boundary curve S. Find the value of ϕ inside the given domain.' The associated eigenvalue problem is now

$$\Delta u = \lambda v, \quad \Delta v = \lambda u \tag{77}$$

with the boundary conditions

$$\left.\begin{aligned} u = 0, \quad \frac{\partial u}{\partial n} = 0, \quad &\text{along } s, \\ v = 0, \quad \frac{\partial v}{\partial n} = 0, \quad &\text{along } S - s. \end{aligned}\right\} \tag{78}$$

Now it is exactly this eigenvalue problem which characterizes the *vibrations of an elastic sheet, clamped along s and free along* $S-s$. This problem is well investigated for simple (particularly rectangular and circular) boundaries but only under the assumption that the *full* boundary is clamped. The interesting fact that for any partially open boundary the eigenvalue $\lambda = 0$ is a limit point of the eigenvalue spectrum—i.e. that there exist infinitely many vibrational modes which belong to arbitrarily small frequencies—has (to the author's knowledge) escaped the attention of the research workers in this field. And yet, we can demonstrate the existence of this limit point without any detailed calculations, solely on the basis of a logical argument.

First of all, the examination of the boundary conditions (78) shows that —in view of the complete symmetry of the differential equations (77) with respect to u and v—the following two problems possess exactly the same eigenvalue spectrum: (a) the sheet clamped along s and free otherwise; (b) the sheet clamped along $S-s$ and free otherwise. Let us now assume that $\lambda = 0$ is *not* a limit point of the eigenvalue spectrum. Then there must exist a definite *smallest* vibrational frequency ν_1. If

we now enlarge the length of s by a small amount, we have from the standpoint of (*a*) *tightened* the boundary conditions and thus ν_1 must *increase*, while from the standpoint of the complementary problem (*b*) we have *weakened* the boundary conditions and thus ν_1 must *decrease*.† These two statements are self-contradictory, which disproves the existence of ν_1 and thus proves the existence of the limit point $\lambda = 0$.

How far can we tighten the boundary conditions before this limit point disappears? We have to clamp in fact the *entire boundary* for this purpose. Even an arbitrarily small unclamped part of the boundary will cause arbitrarily small frequencies. The limit point $\lambda = 0$ disappears suddenly and jumps up to a finite ν_1 in the moment that even the last portion of the free boundary becomes clamped. But can we believe that this will really happen? Can we assume that any clamping mechanism is so perfect that not even an arbitrarily small portion of the sheet might maintain its free mobility? If this question is answered by 'no', then we automatically admit a distribution of vibrational frequencies which is very different from the traditional one. We should then find that an elastic sheet, if struck with a hammer, contains overtones which are exceedingly *low*-pitched. This can be tested in an indirect way by putting the sheet under lateral pressure. This pressure decreases the vibrational frequencies, until a 'critical load' is reached at which the smallest frequency becomes reduced to zero. At that moment the elastic sheet collapses and we experience the phenomenon of 'buckling'.‡

Now it is a well-known fact that the actual critical load at which buckling occurs is by far *lower* than that evaluated theoretically.§ This surprising result is entirely understandable, if we take into account the effect of the 'residual spectrum' which must accompany even the slightest imperfections of the imposed boundary conditions. 'But'—we may be inclined to say—'this implies another absurdity since now buckling could occur under the slightest lateral pressure, which is certainly not the case.' This possibility is prevented, however, by that peculiar 'reversal of eigenvalues' that we have discussed before. A very small eigenvalue does not belong to a vibrational mode of *low* but of *high* order. This means that a very small eigenvalue would demand a rippling of the sheet of such fineness which is physically unrealizable. And thus the 'residual spectrum' in its *physical* manifestation does not start with the frequency zero but with a definite finite frequency which, however, is still much lower than the lowest frequency calculated for the case of perfect clamping.

† Cf. [2], p. 407. ‡ Cf. [14], chap. IX, pp. 439–497. § Cf. [14], p. 462.

The explanation suggested here of the reduced elastic stability of thin sheets is open to experimental verification, since imperfect clamping conditions can be experimentally generated and the corresponding critical loads (at least in rough approximation) calculated. The currently accepted explanation is based on very different considerations. The theory proposed by von Karman[15] operates with vibrations of finite amplitudes, while the theory of Donnell[3] is based on large initial deflections, caused by imperfections of the cylindrical shape. In both cases the discrepancy between theory and experiment is conceived as a non-linear effect which leads to exceedingly complex calculations. The present considerations do not go beyond the realm of the classical eigenvalue theory, although attention is called to the fact that the 'reversal of eigenvalues' which must come in operation under these conditions, confronts us with a situation which is not of the conventional type and which cannot be treated by the customary methods.

10. Summary

The present investigation endeavours to establish a common platform for the theory of linear partial differential equations, subjected to boundary conditions which may be chosen in an arbitrarily judicious or injudicious fashion and given in an arbitrarily over-abundant or under-abundant number. The problem is approached from the domain of algebra, exploiting the isomorphism which exists between linear differential equations and systems of linear algebraic equations. First the general problem of $n \times m$ linear systems is solved on the basis of an eigenvalue problem which yields a unique solution of the problem if the right side satisfies the proper compatibility conditions. This method is then translated into the domain of differential equations. It is found that a general boundary value problem belongs to one of two categories, depending on the question whether the eigenvalue spectrum excludes or includes the value $\lambda = 0$ as a limit point. The conventional boundary value problems belong to the first category. The eigenvalue method is applicable, however, to both categories and yields the solution on the basis of an expansion into eigenfunctions. The solvability of the problem demands that the data satisfy compatibility conditions—and in the second category certain convergence conditions—which can be explicitly stated. The explicit construction of the solution and the testing of the data concerning solvability presupposes, however, the preliminary solution of the eigenvalue problem, excluding the eigensolutions

associated with the eigenvalue zero which are not needed for the construction of the solution and which are dispensable also from the standpoint of the compatibility conditions.

REFERENCES

[1] Bernoulli, D. *Commentarii Academiae Petropolitanae*, 6, 108 (1732).

[2] Courant, R. and Hilbert, D. *Methods of Mathematical Physics*, I. Interscience Publishers, New York, London, 1953.

[3] Donnell, L. H. and Wan, C. C. Effect of imperfections on buckling of thin cylinders and columns under axial compression. *J. appl. Mech.* 17, 73–83 (1950).

[4] Euler, L. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes* (1744).

[5] Frazer, R. A., Duncan, W. J. and Collar, A. R. *Elementary Matrices.* Cambridge University Press, 1950.

[6] Fredholm, I. Sur une classe d'équations fonctionelles. *Acta Math.* 27, 365–390 (1903).

[7] Hadamard, J. *Lectures on Cauchy's Problem.* Dover Publications, New York, 1952.

[8] Hellinger, E. and Toeplitz, O. *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten.* Chelsea Publishing Co., New York, 1953.

[9] Hilbert, D. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen.* Chelsea Publishing Co., New York, 1953.

[10] Lagrange, J. *Miscellanea Taurinensia*, 1, pt. 3, 1759.

[11] MacDuffee, C. C. *The Theory of Matrices.* Chelsea Publishing Co., New York, 1946.

[12] Rayleigh, Lord. *The Theory of Sound*, I. Macmillan and Co., London, 1937.

[13] Schmidt, E. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.* 64, 161–174 (1907).

[14] Timoshenko, S. *Theory of Elastic Stability.* McGraw-Hill, New York and London, 1936.

[15] Kármán, Th. von and Tsien, H. The buckling of thin cylindrical shells under axial compression. *J. aero. Sci.* 8, 303–312 (1941).