1. Kernel Polynomials in Linear Algebra and Their Numerical Applications[1]

Eduard L. Stiefel[2]

CONTENTS

1. Introduction

In recent years many computational procedures have been developed in order to handle problems of linear algebra, in particular for solving linear equations and computing eigenvalues. The principal need for this research in a relatively elementary field of mathematics has arisen in the use of high-speed computational equipment, where numerical stability and the self-correcting features of a numerical method are very important. As a result, some of the classical techniques have had to be improved and refined. Much progress was made, especially in the domain of iteration processes. There are various reasons for preferring iterations in high-speed computing. First of all, each step of an iteration routine gives a new and better estimate of the solution. Furthermore, iteration is easy to code. Finally, in linear algebra an iteration uses the given matrix over and over again and does not modify the matrix during the computation. This is most important if the matrix contains many zeros, as often happens in solving boundary-value problems in mathematical physics by the method of finite differences.

The author feels that there is a common theoretical background in many apparently distinct modern iteration techniques. This is the theory of orthogonal polynomials. Let us take it as a basis for better understanding and relating known methods, and perhaps for discovering new ones. It is fairly obvious that polynomials play an important part in the theory of linear problems. From the point of view of abstract algebraic structure, the computation with a given matrix A and its powers is isomorphic to the computation in the ring of polynomials in a variable λ. Orthogonality of polynomials is important because orthogonal polynomials are distinguished by extremum properties. This fact is familiar to computers in the special case of Chebyshev polynomials. Last but not least, polynomials (and rational functions) are the only functions that an electronic computer is able to handle. Hence in working with polynomials we introduce into pure mathematics the practical properties of our computational equipment.

2. Review of the Elementary Properties of Orthogonal Polynomials

The contents of this section are classic and should be taken as a review preparatory to further investigations. An excellent textbook on the theory of orthogonal polynomials has been written by Szegö [1].[3]

[1] This paper was presented in April 1955 in the form of four lectures that were prepared under contract between the National Bureau of Standards and American University with the sponsorship of the Office of Naval Research.

[2] Eidgenössische Technische Hochschule, Zürich.

[3] Figures in brackets indicate the literature references at the end of this paper.

Let $a \leq \lambda \leq b$ be a given finite interval on the axis of the real variable λ. A real function, $\rho(\lambda)$, defined in this interval will be called a *density function* if it is either a continuous function positive in the open interval

$$\rho(\lambda) > 0, \qquad a < \lambda < b, \tag{1}$$

or if it is of the form

$$\rho(\lambda) = \sum_{j=1}^{m} k_j^2 \, \delta(\lambda - \lambda_j), \qquad a < \lambda_j < b, \tag{2}$$

where the k_j are given real numbers not equal to 0. In this formula δ stands for the Dirac δ-function. $\delta(\lambda)$ is an improper function vanishing in each point $\lambda \neq 0$, but having at $\lambda = 0$ so high and steep a peak that

$$\int_{-\epsilon}^{+\epsilon} \delta(\lambda) d\lambda = 1. \tag{3}$$

Therefore, the density function (2) has peaks at the points $\lambda_1, \lambda_2, \ldots, \lambda_m$, but vanishes everywhere else in the interval (a, b). If $f(\lambda)$ is any function defined in (a, b), we have from (3),

$$\int_a^b f(\lambda) \delta(\lambda - \lambda_j) d\lambda = f(\lambda_j). \tag{3a}$$

Hence, using the density (2),

$$\int_a^b f(\lambda) \rho(\lambda) d\lambda = \sum_{j=1}^{m} k_j^2 f(\lambda_j). \tag{4}$$

Therefore, the purpose of introducing the δ-function is only to write finite sums as integrals. Of course this can be done also by the concept of Stieltjes integrals. The density (2) will be called *of Dirac's type.*

Corresponding to a given density function $\rho(\lambda)$, there is a sequence of polynomials

$$P_0(\lambda) = 1, \; P_1(\lambda), \; P_2(\lambda), \ldots, P_n(\lambda), \ldots, \tag{5}$$

such that P_n has the degree n (λ^n having a nonvanishing coefficient) and

$$\int_a^b P_i(\lambda) P_k(\lambda) \rho(\lambda) d\lambda = 0, \quad \text{if} \quad i \neq k. \tag{6}$$

This is to say, two polynomials of the sequence are orthogonal with respect to density ρ. In order to prove this statement, let us construct the sequence by recursion in the following way. Assume that the sequence is already established up to P_n. Then λP_n has the degree $(n+1)$. In order to make it orthogonal to the two preceding polynomials P_n, P_{n-1}, we add a linear combination of P_n, P_{n-1}, thus putting

$$P_{n+1}(\lambda) = \lambda P_n(\lambda) - \alpha_{n+1} P_n(\lambda) - \beta_n P_{n-1}(\lambda). \tag{7}$$

The orthogonality of P_{n+1}, P_n and P_{n+1}, P_{n-1} yields

$$\alpha_{n+1} = \frac{\int \lambda P_n(\lambda)^2 \rho(\lambda) d\lambda}{\int P_n(\lambda)^2 \rho(\lambda) d\lambda}, \qquad \beta_n = \frac{\int \lambda P_n(\lambda) P_{n-1}(\lambda) \rho(\lambda) d\lambda}{\int P_{n-1}(\lambda)^2 \rho(\lambda) d\lambda}, \tag{8}$$

the limits of the integrals being always a, b. The polynomial (7) is automatically orthogonal to $P_i(\lambda)$ for $i < (n-1)$. This is established by induction, admitting that the three-term recursion formula (7) is true for indices lower than n. From definition (7):

$$\int P_{n+1} P_i \rho d\lambda = \int \lambda P_n P_i \rho d\lambda.$$

From the hypothesis of induction,

$$\lambda P_i = P_{i+1} + \alpha_{i+1} P_i + \beta_i P_{i-1}.$$

Hence

$$\int P_{n+1} P_i \rho d\lambda = \int P_n P_{i+1} \rho d\lambda + \alpha_{i+1} \int P_n P_i \rho d\lambda + \beta_i \int P_n P_{i-1} \rho d\lambda.$$

But each of the integrals on the right-hand side vanishes.

The fact that there is a recursion formula (7) linking together three consecutive polynomials is most important for this theory of solving linear problems.

In the case of a continuous density, the denominators in (8) are always nonzero, and thus our construction never breaks down. In the case of a Dirac density, however, this is not true. The orthogonality (6) becomes then, by (4),

$$\sum_{j=1}^{m} k_j^2 P_i(\lambda_j) P_k(\lambda_j) = 0,$$

which is to say that the vectors

$$[k_1 P_i(\lambda_1), k_2 P_i(\lambda_2), \ldots, k_m P_i(\lambda_m)], \qquad i = 0,1,2, \ldots, \tag{9}$$

build an orthogonal set in the m-dimensional vector space. Because there are at most m linear independent vectors in this space, for some $i \leq m$, the vector (9) must have the length 0:

$$\sum k_j^2 P_i(\lambda_j)^2 = \int P_i(\lambda)^2 \rho(\lambda) d\lambda = 0. \tag{10}$$

Thus

$$P_i(\lambda_j) = 0, \qquad j = 1,2, \ldots, m. \tag{11}$$

But a polynomial of degree i can only have m roots if $i \geq m$, hence $i = m$. Therefore, the following statements are correct:

$$\int P_i(\lambda)^2 \rho(\lambda) d\lambda \neq 0, \quad \text{for} \quad i < m, \quad P_m(\lambda_j) = 0. \tag{12}$$

This proves

THEOREM 1. *In the case of a density of Dirac's type with m peaks, the sequence of orthogonal polynomials finishes exactly with the polynomial of degree m. This last polynomial has its roots at the abscissas of the peaks.*

EXAMPLE. *Chebyshev polynomials.* Let us take as interval $(-1,1)$ and as a continuous density

$$\rho(\lambda) = \frac{1}{\sqrt{1-\lambda^2}}. \tag{13}$$

Introducing ϕ by $\cos \phi = \lambda$, we observe that

$$P_n(\lambda) = \cos n\phi \tag{14}$$

is a polynomial in λ because $\cos n\phi$ is a linear combination of the powers of $\cos \phi$. These polynomials are orthogonal with respect to the given density; indeed,

$$\int_{-1}^{+1} P_i P_k \rho d\lambda = \int_0^{\pi} \cos i\phi \cos k\phi \frac{\sin \phi}{\sqrt{1-\cos^2 \phi}} d\phi = 0.$$

From $\cos(n+1)\phi = 2 \cos \phi \cos n\phi - \cos(n-1)\phi$ we obtain the recursion formula,

$$P_{n+1} = 2\lambda P_n - P_{n-1}.$$

Expansions. Any function $f(\lambda)$ defined in (a,b) may be formally expanded into a series of orthogonal polynomials:

$$f(\lambda) \backsimeq \sum_{(i)} c_i P_i(\lambda). \tag{15}$$

Multiplying this by $P_k(\lambda)$ and integrating, we obtain as a result for the Fourier coefficient c_k

$$c_k=\frac{\int f(\lambda)P_k(\lambda)\rho(\lambda)d\lambda}{N_k}. \tag{16}$$

Here N_k is an abbreviation for the norm,

$$N_k=\int P_k(\lambda)^2\rho(\lambda)d\lambda. \tag{17}$$

Take as a special case $f(\lambda)=\lambda^n$. (In the case of a Dirac density, n should be less than or equal to the number of peaks.) It is obvious by induction that λ^n is a linear combination of $P_0,P_1, \ldots, P_n$. This is to say that the expansion (15) is not formal but exact and finite. Furthermore, for $k>n$ we have $c_k=0$, or

$$\int \lambda^n P_k(\lambda)\rho(\lambda)d\lambda=0, \quad \text{for} \quad k>n. \tag{18}$$

By simple superposition these results carry over from λ^n to any polynomial $\Pi_n(\lambda)$ of degree n. There is a finite expansion

$$\Pi_n(\lambda)=\sum_{i=0}^{n} c_i P_i(\lambda), \tag{19}$$

and

$$\int \Pi_n(\lambda)P_k(\lambda)\rho(\lambda)d\lambda=0, \quad \text{for} \quad k>n. \tag{20}$$

Expansion of the δ-function, kernel polynomials. Let λ_0 be a point in the open interval (a, b). We want to construct a polynomial that has a very high and steep peak at λ_0 but takes on small values elsewhere. This can be done by expansion of the δ-function,

$$\delta(\lambda-\lambda_0)\cong\sum_{(i)} c_i P_i(\lambda),$$

truncating the series after a finite number of terms. From (16) and (3a),

$$c_k=\frac{P_k(\lambda_0)\rho(\lambda_0)}{N_k}.$$

Thus

$$\delta(\lambda-\lambda_0)=\rho(\lambda_0)\sum_{(k)}\frac{P_k(\lambda_0)P_k(\lambda)}{N_k}. \tag{21}$$

The constant factor $\rho(\lambda_0)$ is not essential for our purposes. The partial sum,

$$K_n(\lambda_0,\lambda)=\sum_{k=0}^{n}\frac{P_k(\lambda_0)P_k(\lambda)}{N_k}, \tag{22}$$

is called the nth *kernel polynomial* with respect to the density ρ, and some formal properties of those polynomials are listed, *dropping the assumption that* λ_0 *is inside the interval.* Take any polynomial Π_n of maximal degree n and its expansion (19):

$$\Pi_n(\lambda)=\sum_{i=0}^{n} P_i(\lambda)\frac{1}{N_i}\int \Pi_n(\mu)P_i(\mu)\rho(\mu)d\mu$$

$$=\int \Pi_n(\mu)\rho(\mu)\sum_{i=0}^{n}\frac{P_i(\lambda)P_i(\mu)}{N_i}d\mu$$

$$\Pi_n(\lambda)=\int K_n(\lambda,\mu)\Pi_n(\mu)\rho(\mu)d\mu. \tag{23}$$

The polynomial Π_n is reproduced by integration with the kernel K_n. Thus

THEOREM 2. *The function $K_n(\lambda,\mu)$ is a reproducing kernel in the field of polynomials of maximal degree n.* Take, in particular, $\Pi_n(\lambda)=K_n(\lambda_0,\lambda)$. Then

$$K_n(\lambda_0,\lambda)=\int K_n(\lambda,\mu)K_n(\lambda_0,\mu)\rho(\mu)d\mu. \tag{24}$$

This important identity gives for $\lambda=\lambda_0$ the result

$$\int K_n(\lambda,\mu)^2\rho(\mu)d\mu=K_n(\lambda,\lambda), \tag{25}$$

which is an evaluation of the *norm of the kernel polynomial.* Now take $\Pi_n(\lambda)=(\lambda-\lambda_0)\Pi_i(\lambda)$, with $i<n$. From (23),

$$\int K_n(\lambda,\mu)\Pi_i(\mu)(\mu-\lambda_0)\rho(\mu)d\mu=(\lambda-\lambda_0)\Pi_i(\lambda),$$

and for $\lambda=\lambda_0$,

$$\int K_n(\lambda_0,\mu)\Pi_i(\mu)(\mu-\lambda_0)\rho(\mu)d\mu=0.$$

Thus $K_n(\lambda_0,\lambda)$ is orthogonal to any polynomial of degree $i<n$ with respect to the density $(\lambda-\lambda_0)\cdot\rho(\lambda)$. Or

THEOREM 3. *The kernel polynomials $K_n(\lambda_0,\lambda)$ build for a fixed λ_0 an orthogonal set with respect to the density* $(\lambda-\lambda_0)\cdot\rho(\lambda)$. It is important, however, to observe that $(\lambda-\lambda_0)\cdot\rho(\lambda)$ is only a density function in the strict sense of definitions (1) and (2) if λ_0 is a point *outside* the interval (a, b), because otherwise the new density changes its sign in an interior point of the interval.

The original idea of constructing a polynomial with a dominant peak is reflected in the following exact statement.

THEOREM 4. *Among all polynomials of maximal degree n and having a given value a at the given point λ_0 the polynomial $cK_n(\lambda_0,\lambda)$ yields the least norm.* Of course c is determined by $cK_n(\lambda_0,\lambda_0)=a$.

PROOF. From (19) it follows for any Π_n by Cauchy's inequality that

$$\Pi_n(\lambda_0)^2=\left\{\sum_{i=0}^{n}(\sqrt{N_i}c_i)\frac{P_i(\lambda_0)}{\sqrt{N_i}}\right\}^2\leq\left\{\sum_i N_i c_i^2\right\}\left\{\sum_i\frac{P_i(\lambda_0)^2}{N_i}\right\}.$$

If Π_n is in the family under consideration, the left side is equal to a^2. The first factor on the right side is the norm N of Π_n by (19), and the second factor is $K_n(\lambda_0,\lambda_0)=a/c$. Hence $ac\leq N$. From (25),

$$\int K_n(\lambda_0,\mu)^2\rho(\mu)d\mu=K_n(\lambda_0,\lambda_0)=a/c.$$

Thus

$$\int\{cK_n(\lambda_0,\mu)\}^2\rho(\mu)d\mu=ac\leq N.$$

Theorem 4 establishes a kind of *filtering property* of the kernel polynomials. If the λ-axis is a frequency axis and $f(\lambda)$ a frequency function, the integral

$$\int K_n(\lambda_0,\lambda)f(\lambda)\rho(\lambda)d\lambda$$

filters out the frequency λ_0, suppressing all the other frequencies better and better with increasing This property is the key to the utility of the kernel polynomials in numerical analysis.

3. On Solving Linear Equations

Given a nonsingular real matrix, A, we want to solve the linear system,

$$Ax=k, \tag{26}$$

where x is the vector of the unknowns, and k is the given vector of constants. We propose to solve (26) by iteration. Using geometrical terms, we start with a first trial point x_0. Without any loss of generality, x_0 *may be assumed to be the origin* $x_0=0$. Then we move out of x_0 in some chosen direction to the next approximation point x_1, adding a correction, Δx_0. This construction is repeated up to a final approximation point x_n that is hoped to be near the desired solution x. The accuracy of x_i may be checked by putting this point into (26) and computing the *residual*,

$$r_i=k-Ax_i. \tag{27}$$

Such an iteration procedure is characterized by the choice of the directions of the segments of the path and by the choice of the lengths of the segments. In the following investigation we limit ourselves to the iterations where the direction of motion of x_i is indicated by the residual r_i. Thus we record the rules,

$$x_{i+1}=x_i+\Delta x_i, \qquad \Delta x_i=\left(\frac{1}{q_i}\right)r_i, \qquad x_0=0. \tag{28}$$

The scalars q_i are still arbitrary, they are called *relaxation factors*, and they determine the length of the segments of the path.[4] In order to secure the convergence of the iteration some *tactic* is needed. Normally, it is recommended that q_i be chosen in such a way that an appropriate error measure is minimized going from x_i to x_{i+1}. This tactical rule concerns only a single step. However, like a good chess player, one should not be guided by the greatest advantage to be gained in a single move, but should consider the wisest over-all strategy for winning the game and choose the set of moves that best promotes that strategy. Therefore, the following *strategy* problem must be solved. Among all paths having a given number of segments n, it is necessary to find the path yielding an endpoint x_n with least error measure. Therefore we must determine the relaxation factors $q_0, q_1, \ldots, q_{n-1}$ simultaneously. Before this can be done, some information is needed on how residuals build up.

Residual polynomials. From (28) it follows that

$$r_{i+1}=k-Ax_{i+1}=k-Ax_i-A\Delta x_i=r_i-\frac{1}{q_i}Ar_i=\left(1-\frac{A}{q_i}\right)r_i. \tag{29}$$

(Here 1 stands for the unit matrix.) Furthermore, $r_0=k-Ax_0=k$, and by repeated application of the law (29),

$$r_n=\left(1-\frac{A}{q_0}\right)\left(1-\frac{A}{q_1}\right)\ldots\left(1-\frac{A}{q_{n-1}}\right)k. \tag{30}$$

Introducing the polynomial in a real variable

$$R_n(\lambda)=\left(1-\frac{\lambda}{q_0}\right)\left(1-\frac{\lambda}{q_1}\right)\ldots\left(1-\frac{\lambda}{q_{n-1}}\right), \tag{31}$$

we may write

$$r_n=R_n(A)k. \tag{32}$$

The *residual polynomial* $R_n(\lambda)$ has the important property to take the value 1 at the origin,

$$R_n(0)=1. \tag{33}$$

[4] If A is a symmetric and definite matrix, the iteration (28) is called method of *steepest descent* or *gradient method*. This is so named for the following reason: Consider the quadratic function $F(x)=(x, Ax)-2(k,x)$, where the comma stands for the scalar product. Then grad $F=2(Ax-k)=-2r$. The direction of r is orthogonal to the level surfaces of F; in this direction F diminishes as fast as possible. The solution point x is characterized by grad $F=0$ and is thus the point where F is minimal.

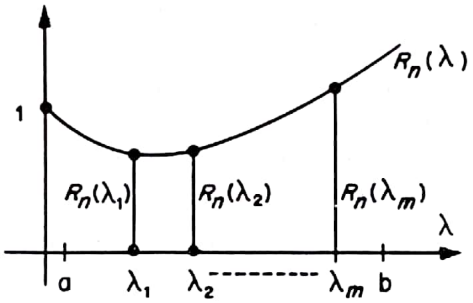
Furthermore, the relaxation factors are the roots of the residual polynomial. Inversely, any polynomial R_n having its roots real and satisfying (33) determines uniquely an iteration process yielding as last residual (32). In order to prove this, it is sufficient to take the roots of the chosen polynomial as relaxation factors.

THEOREM 5. *There is a one-to-one correspondence between iteration processes and polynomials with real roots satisfying* (33).

Let us assume from now on that the matrix A has *real* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ and can be transformed by a real coordinate transformation into the diagonal form,

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \lambda_m \end{pmatrix}.$$

(The latter assumption could be easily removed by dealing with Jordan's canonical form instead of diagonal matrices.) Later on we will make some remarks about handling more general matrices. Under these assumptions the graph of the residual polynomial $R_n(\lambda)$ gives much information about the behavior of the iteration.



Indeed,

$$R_n(A)=\begin{pmatrix} R_n(\lambda_1) & & & \\ & R_n(\lambda_2) & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & R_n(\lambda_m) \end{pmatrix}.$$

Therefore, eq (32) splits into the following relations for the components of the vectors in the new coordinate system:

$$r_{ni}=R_n(\lambda_i)k_i, \qquad i=1, 2, \ldots, m. \tag{34}$$

This is to say: *The ordinate $R_n(\lambda_i)$ of our graph is the percentage by which the ith component of the residual is reduced after the nth step of the iteration.*

It seems therefore reasonable to introduce as an error measure,

$$\int_a^b R_n(\lambda)^2 \rho(\lambda) d\lambda, \tag{35}$$

where (a, b) is an interval containing the eigenvalues and $\rho(\lambda)$ an arbitrary density function in this interval. From theorem 5 it follows that this strategy problem is equivalent to the construction of a

polynomial $R_n(\lambda)$, of given degree n, satisfying $R_n(0)=1$ and minimizing (35). The answer is given by theorem 4. We have to choose

$$R_n(\lambda)=\frac{K_n(0,\lambda)}{K_n(0,0)}. \tag{36}$$

The roots of this polynomial are indeed real.[5] Hence we may record

THEOREM 6. *The iteration of best strategy with respect to the error measure (35) is given by the following rule. Let $a\leq\lambda\leq b$ be an interval containing the eigenvalues of the given matrix. Choose any density function in (a,b). Take as relaxation factors the roots of the nth kernel polynomial with respect to this density.*

Computational technique. It is not impossible to use theorem 6 as it stands for numerical computations. However, the density $\rho(\lambda)$ must be so simple a function that the zeros of the kernel polynomials are available. D. Young [2] has worked out such a routine in the case of a Chebyshev density. But even then one may run into trouble, because the arrangement of the roots remains undetermined. Therefore, there are $n!$ different paths satisfying theorem 6 and ending in the *same* point of best strategy. Some of them might be highly unstable.

Therefore, we modify slightly the basic iteration process in the case where the eigenvalues of the given matrix A are not only real but positive, and therefore the lower bound a can be assumed to be greater than or equal to zero. Consider not only the single polynomial $R_n(\lambda)$, but the whole set of polynomials

$$R_i(\lambda)=\frac{K_i(0,\lambda)}{K_i(0,0)}, \qquad i=0,1,\ldots,n. \tag{37}$$

From theorem 3 it follows that these polynomials are *orthogonal* with respect to the density function $\lambda\rho(\lambda)$, which is now in (a,b) a density in the strict sense of definitions (1) and (2), and therefore three consecutive polynomials are linked by a recursion formula of type (7). We prefer to write this formula in the form of a finite expansion of $\lambda R_i(\lambda)$:

$$\lambda R_i(\lambda)=-q_iR_{i+1}(\lambda)+t_iR_i(\lambda)-p_iR_{i-1}(\lambda). \tag{38}$$

(There is no reason to confound the coefficients q_i with the relaxation factors denoted previously by the same letter.) From (37) it follows that $R_i(0)=1$; thus $t_i=p_i+q_i$ and

$$\lambda R_i(\lambda)=-q_iR_{i+1}(\lambda)+(p_i+q_i)R_i(\lambda)-p_iR_{i-1}(\lambda). \tag{39}$$

Now we construct a new set of vectors, $r_0=k$, $r_1,r_2,\ldots,r_n$, defined by

$$r_i=R_i(A)k. \tag{40}$$

We have the recursion formula,

$$Ar_i=-q_ir_{i+1}+(p_i+q_i)r_i-p_ir_{i-1}.$$

Introducing $\Delta r_i=r_{i+1}-r_i$, this may be written

$$\Delta r_i=\frac{1}{q_i}(p_i\Delta r_{i-1}-Ar_i). \tag{41}$$

Finally, we want to construct the iteration path yielding the vectors r_i as residual vectors. Let x_i be the approximation points, $\Delta x_i=x_{i+1}-x_i$. The objective is

$$r_i=k-Ax_i. \tag{42}$$

[5] For a proof see Szegö, Über orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehören, Math. Z. 9, 241–244 (1921).

Subtracting this from $r_{i+1}=k-Ax_{i+1}$, we get $\Delta r_i=-A\Delta x_i$. Hence, from (41)

$$\Delta x_i=\frac{1}{q_i}(r_i+p_i\Delta x_{i-1}).$$

This is the rule for the construction of the new path. It must be emphasized that *each point* of this path, and not only the last one, is a *point of best strategy*. Indeed, from (40) it follows by (37) that the residual polynomial of the ith point is the kernel polynomial. Therefore the path is perfectly stable.

THEOREM 7. *A computational routine of best strategy is the following. Let $0\leq a<b$ be an interval containing the eigenvalues. Choose any density function $\rho(\lambda)$ in (a,b). Construct the set of orthogonal polynomials $R_i(\lambda)$ with respect to the density $\lambda\rho(\lambda)$ satisfying $R_i(0)=1$. Their recurrence formula is*

$$\lambda R_i=-q_iR_{i+1}+(p_i+q_i)R_i-p_iR_{i-1}. \tag{43}$$

Carry out the iteration process:

$$x_0=0,\ x_{i+1}=x_i+\Delta x_i,\quad \text{where}\quad \Delta x_i=\frac{1}{q_i}(r_i+p_i\Delta x_{i-1}). \tag{44}$$

and r_i is the residual $r_i=k-Ax_i$.

Observe that the roots of polynomials are no longer needed. The new routine is only slightly more complicated than the original routine (28). The additional term $p_i\Delta x_{i-1}$ takes into account the history of the process and takes care of the numerical stability. At each step the accuracy is checked because the residual is computed.

Choice of the density function. This choice must still be made. Theoretically any information about the eigenvalues of the given matrix can be helpful to determine $\rho(\lambda)$. Assume, for instance, that some eigenvalues cluster around a point λ_0. Then put a heavy weight on this region, choosing a density function having a maximum at λ_0. By this technique, residuals in the region of λ_0 will be quickly liquidated. However, if $\rho(\lambda)$ is too complicated a function, the computation of the q_i, p_i might be difficult. Discussion of a few typical cases follows.

1. It is easy to find an upper bound b for the eigenvalues by Gerschgorin's circle theorem [3]. It might be much more difficult to find a fairly good lower bound $a>0$. But let us assume that such a bound is known. As suggested by Shortley and Flanders, we may take for the residual polynomials $R_i(\lambda)$ the Chebyshev polynomials adopted to the interval (a, b). This is a good choice because the Chebyshev polynomials have not only the property of yielding the least norm but also the property of yielding the *least maximal value* inside (a, b). In order to find the recurrence relation we put

$$\cos\phi=\frac{-2\lambda+(a+b)}{b-a},$$

ϕ running from 0 to π as λ goes from a to b. The residual polynomials are given by

$$R_n(\lambda)=\frac{\cos n\phi}{\cosh n\omega}, \tag{45}$$

where ω is determined by

$$\cosh\omega=\frac{b+a}{b-a}. \tag{46}$$

It is not too hard to establish the recurrence relation. The final result is the following rule for iteration:

$$\Delta x_n=\frac{1}{\cosh(n+1)\omega}\left\{\frac{4}{b-a}\cosh n\omega\cdot r_n+\cosh(n-1)\omega\cdot\Delta x_{n-1}\right\}. \tag{47}$$

For $n=0$, this formula must be replaced by $\Delta x_0=2r_0/(b+a)$. From (45) it follows that after the nth step the residuals are reduced by a factor of $1/\cosh n\omega$.

2. Suppose now that no lower bound of the eigenvalues is available. Then we have to put $a=0$, and to choose a density function $\rho(\lambda)$ in the interval $(0, b)$. Technique (47) does not work any more

because $\omega=0$ and the residuals are not reduced at all. Without loss of generality we assume for the following discussion $b=1$. In [4] the choice,

$$\rho(\lambda)=\lambda^{\alpha-1}(1-\lambda)^{\beta}, \qquad \alpha>0, \qquad \beta\geq-\tfrac{1}{2}, \tag{48}$$

has been investigated. The corresponding residual polynomials are orthogonal with respect to the density $\lambda^{\alpha}(1-\lambda)^{\beta}$ as follows from theorem 7. They are therefore the general *hypergeometric* or *Jacobi polynomials*. The final rule for iteration is given by the values of the relaxation coefficients,

$$\begin{aligned} q_n&=\frac{1}{4}\left\{1+\frac{(\alpha-\beta+1)(\alpha+\beta+1)}{2n+\alpha+\beta+1}-\frac{(\alpha-\beta)(\alpha+\beta)}{2n+\alpha+\beta+2}\right\}\\ p_n&=\frac{1}{4}\left\{1-\frac{(\alpha-\beta+1)(\alpha+\beta+1)}{2n+\alpha+\beta+1}+\frac{(\alpha-\beta)(\alpha+\beta)}{2n+\alpha+\beta}\right\}. \end{aligned} \tag{49}$$

These relaxation coefficients are rational numbers, which is an advantage for automatic computation. (In (47) they are given by a transcendental function.) One should be careful in applying this hypergeometric relaxation, because for some values of α and β it has the character of overrelaxation, as discussed in [4]. Indeed, for high values of β, formula (48) shows that there is almost no weight on the higher parts of the spectrum. Therefore, in the beginning of the relaxation the higher residuals are not attacked. This may be advantageous and may speed up convergence, but has to be investigated. The special case $\alpha=+\frac{1}{2}$, $\beta=-\frac{1}{2}$ was discussed earlier by Lanczos and can be recommended as always safe. It yields the simple procedure,

$$\Delta x_n=\frac{1}{2n+3}\{4(2n+1)r_n+(2n-1)\Delta x_{n-1}\}. \tag{50}$$

In order to check hypergeometric relaxation, Dirichlet's boundary-value problem was solved at Zürich in a domain containing 81 grid points. With $\alpha=\beta=\frac{3}{2}$, the maximal error of the function was reduced after 14 steps to 0.3 percent of its original value.

3. Let us now discuss densities of Dirac's type. In order to do this, we must assume that the *matrix* A of the given linear system (26) is *symmetric* and *positive definite*. Denote by m the number of unknowns, or—geometrically speaking—the dimension of the space that the operator A is working in. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be again the eigenvalues of A, and $k_1, k_2, \ldots, k_m$ the components of the given vector k, in the reference system of the principal axis of A. This coordinate system is now a Cartesian system. Take as density function

$$\rho(\lambda)=\sum_{j=1}^{m} k_j^2\delta(\lambda-\lambda_j), \tag{51}$$

having its peaks at the eigenvalues. In order to apply theorem 7 and to establish the recurrence relation

$$\lambda R_t=-q_tR_{t+1}+(p_t+q_t)R_t-p_tR_{t-1}, \tag{52}$$

we use the fact that the R_t build an orthogonal set with respect to $\lambda\rho(\lambda)$. Multiplying (52) successively by R_{t+1}, R_t, R_{t-1} and integrating, we find that

$$q_t=\frac{-\int\lambda^2R_tR_{t+1}\rho d\lambda}{C_{t+1}}, \qquad p_t+q_t=\frac{\int\lambda^2R_t^2\rho d\lambda}{C_t}, \qquad p_t=\frac{-\int\lambda^2R_tR_{t-1}\rho d\lambda}{C_{t-1}},$$

where

$$C_t=\int\lambda R_t^2\rho d\lambda. \tag{53}$$

It is a little more convenient to write

$$q_t=\frac{1}{C_t}\int\lambda^2R_t^2\rho d\lambda-p_t, \qquad p_t=\frac{C_t}{C_{t-1}}q_{t-1}. \tag{54}$$

Now, always using formula (4):

$$C_i = \int \lambda R_i^2 \rho d\lambda = \sum_{(j)} \lambda_j R_i(\lambda_j)^2 k_j^2.$$

Taking into account the diagonal form of the matrix in the reference system of principal axis, together with (34), we have

$$C_i = \sum_{(j)} \lambda_j r_{ij}^2,$$

where r_{ij} is the jth component of the residual vector r_i. This sum is nothing but the scalar product of the vectors r_i and Ar_i. Therefore,

$$C_i = (r_i, Ar_i).$$

By straightforward computation of the same type, we obtain

$$\int \lambda^2 R_i^2 \rho d\lambda = (Ar_i, Ar_i).$$

Using (41), the final routine is now

$$r_0 = k, \qquad r_{i+1} = r_i + \Delta r_i, \qquad \Delta r_i = \frac{1}{q_i}(p_i \Delta r_{i-1} - Ar_i),$$

where

$$q_i = \frac{1}{C_i}(Ar_i, Ar_i) - p_i, \ (p_0 = 0), \qquad p_i = \frac{C_i}{C_{i-1}} q_{i-1}, \tag{55}$$

and

$$C_i = (r_i, Ar_i).$$

This yields the successive residuals. Afterwards, compute

$$x_0 = 0, \qquad x_{i+1} = x_i + \Delta x_i, \qquad \Delta x_i = \frac{1}{q_i}(r_i + p_i \Delta x_{i-1}). \tag{56}$$

This iteration has some remarkable properties. First of all, the error measure (35) turns out to be

$$\int R_n(\lambda)^2 \rho(\lambda) d\lambda = \sum_{(j)} R_n(\lambda_j)^2 k_j^2 = (r_n, r_n).$$

Because of the best strategy, the iteration (55) yields therefore the least-square residual after n steps. Furthermore, it follows from theorem 1 that $R_m(\lambda_j) = 0$; thus, $r_{mj} = R_m(\lambda_j) k_j = 0$ and $r_m = 0$. This is to say that the routine gives the exact solution point after m steps, m being the number of unknowns of the given linear system.

THEOREM 8. *The iteration (55), (56) yields after a given number of steps n a smaller length of the residual vector than any iteration of type (28). Moreover, it reaches the solution point after a finite number of steps. This number is the dimension of the vector space mapped onto itself by matrix A.*

This iteration is a modification of the *method of conjugate gradients* described in the monograph [5].

Let us close this section with some recommendations on how to solve a given linear system having eigenvalues that are real and positive. If the system is well conditioned, a fairly good lower bound, $a > 0$, of the eigenvalues will be known and the method of case 1 is most successful. But we have to deal also with *ill-conditioned* systems. Such systems occur, for instance, in solving a boundary-value problem using difference methods in a fine grid. Then the following combination of 1 and 3 seems to be the best. Compute an upper bound b of the eigenvalues. Chose a number $a > 0$ that is not necessarily a lower bound of the eigenvalues but leaves only a small number μ of eigenvalues on the left side. Start with routine (47). This will liquidate the components of the residual vector corresponding to the eigenvalues to the right of a. The remaining residual is therefore a vector in the space R^μ spanned by the eigenvectors of A corresponding to the eigenvalues to the left of a. Operator A leaves this space invariant. Now switch to the routine (55), (56). During this routine all the residuals remain in R. Therefore it follows from theorem 8 that after μ steps the residual is practically 0.

As an example, Laplace's equation, $\Delta u=1$, was solved in a square with vanishing boundary values, using a net of 10×10 points. With $a=2$, $b=8$, eleven steps of the Chebyshev routine (47) were carried out. This reduced the maximal error of the desired function u to 70 percent of its initial value. Then two steps of (55), (56) reduced this further to 0.014 percent. This amazing result is of course due to the special symmetries of the problem.

The described combination is very successful if there are only a few eigenvalues near the origin $\lambda=0$. This happens in particular in solving *integral equations*. In most cases the eigenvalues then cluster around a point $\lambda\neq 0$.

More general matrices. If the eigenvalues of the given matrix A are still real but located on both sides of the point $\lambda=0$, the basic method of theorem 7 breaks down because $\lambda\rho(\lambda)$ is no longer a positive density function in the interval (a,b) of the eigenvalues. However, the technique of theorem 6 is still available.

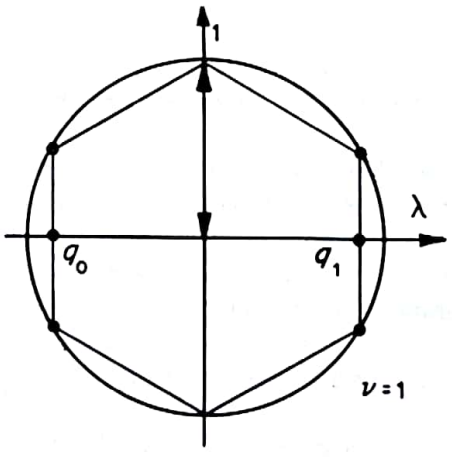
EXAMPLE. Let the eigenvalues be in the interval $(-1,+1)$ and take Chebyshev's density function,

$$\rho(\lambda)=\frac{1}{\sqrt{1-\lambda^2}}. \tag{57}$$

With $\lambda=\cos\phi$, the orthogonal polynomials are $P_n(\lambda)=\cos n\phi$. It is not too hard to establish the kernel polynomial. Up to a constant factor,

$$K_n(0,\lambda)=\begin{cases}\dfrac{\cos n\phi}{\cos\phi}, & n \text{ odd}\\[2ex] \dfrac{\cos(n+1)\phi}{\cos\phi}, & n \text{ even.}\end{cases} \tag{58}$$

Observe that two consecutive polynomials coincide! From theorem 6 it follows that the relaxation factors q_i in the old routine (28) are determined by the roots of $\cos(2\nu+1)\phi$, where ν is a given integer. They may be constructed in the following way. Construct a regular polygon of $2(2\nu+1)$ edges in the unit-circle, one corner lying above the origin $\lambda=0$. Then the relaxation factors are the abscissas of the corners except the origin itself. In order to avoid numerical instability, the relaxation factors should be arranged in *decreasing* distance from the origin $\lambda=0$ during the iteration (28).



There is an almost trivial trick to replace any given linear system $Ax=k$ by a system with a symmetric and positive definite matrix. Introduce a new unknown y by letting $x=A^*y$, where A^* is the transposed matrix of A. Then $(AA^*)y=k$, and AA^* is symmetric and definite. Of course in a practical numerical problem the matrix AA^* is not computed. But in order to apply it on a vector u, first compute $v=A^*u$ and then Av. It is obvious that the transformation almost doubles the computational labor. Applications to problems of the calculus of observations are described in [6].

4. On Computing Eigenvalues

In this section we assume that the given matrix A has *real eigenvalues*. Without loss of generality we may restrict ourselves to matrices with *positive eigenvalues*. Indeed, if a is a lower bound of the spectrum of A, we may replace A by $A+aE$, where E is the unit matrix.

Iteration by kernel polynomials. Again let $0 \leq a < b$ be an interval containing the eigenvalues of A, and $\rho(\lambda)$ a chosen density function in this interval. Furthermore, let k be an arbitrary vector. We want to investigate the iterated vectors,

$$r_n = \frac{K_n(\lambda_0, A)}{K_n(\lambda_0, \lambda_0)} k, \qquad n=0, 1, 2, \ldots, \tag{59}$$

where $K_n(\lambda_0,\lambda)$ is the kernel polynomial of degree n with respect to the density $\rho(\lambda)$ and the parameter λ_0. Remember that it has a peak at the abscissa λ_0. It must be emphasized that now the vectors r_n are not necessarily the residuals of a system of linear equations.

There are two computational routines available to find the iterates r_n:

1. If $\lambda_0=0$, it follows from (36) that $K_n(0,\lambda)/K_n(0,0)$ is the residual polynomial $R_n(\lambda)$ investigated in section 3. Therefore, we have to establish the recurrence relation (39),

$$\lambda R_n(\lambda) = -q_n R_{n+1}(\lambda) + (p_n+q_n) R(\lambda) - p_n R_{n-1}(\lambda), \tag{60}$$

and we may use (41):

$$\Delta r_n = \frac{1}{q_n}(p_n \Delta r_{n-1} - A r_n), \qquad r_{n+1} = r_n + \Delta r_n. \tag{61}$$

In the case of a density of Dirac's type the coefficients are given by (55).

2. If $\lambda_0 \neq 0$, the original definition (22) of kernel polynomials is applicable. Hence

$$r_n = \frac{1}{K_n(\lambda_0,\lambda_0)} \sum_{i=0}^{n} \frac{P_i(\lambda_0)}{N_i} k_i, \qquad k_i = P_i(A)k. \tag{62}$$

The $P_i(\lambda)$ are the orthogonal polynomials with respect to the density $\rho(\lambda)$, and N_i stands for the norm of $P_i(\lambda)$. It remains to compute the sequence of vectors k_i. For this purpose we use the recursion formula (7):

$$P_{i+1}(\lambda) = \lambda P_i(\lambda) - \alpha_{i+1} P_i(\lambda) - \beta_i P_{i-1}(\lambda). \tag{63}$$

Thus

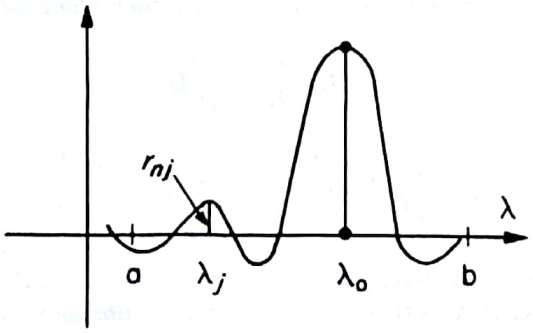
$$k_{i+1} = A k_i - \alpha_{i+1} k_i - \beta_i k_{i-1}, \qquad k_0 = k. \tag{64}$$

Observe that the k_i are independent of λ_0 and therefore yield by (62) the vectors r_n for *any value of* λ_0.

There are different applications of this vector iteration by kernel polynomials.

1. In the reference system of principal axes of A, eq. (59) splits into the following equations for the components of the vectors:

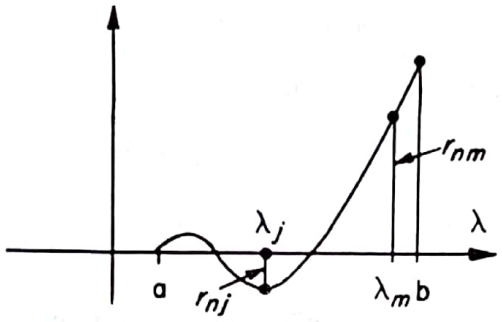
$$r_{nj} = \frac{K_n(\lambda_0,\lambda_j)}{K_n(\lambda_0,\lambda_0)} \cdot k_j. \tag{65}$$



Suppose now that λ_0 is an eigenvalue of A. From the filtering property of kernel polynomials it follows that the component of r_n with respect to the eigenvector corresponding to λ_0 is large and that the components with respect to the other eigenvectors are very small. Therefore formula (62) yields the eigenvector corresponding to λ_0 if $n \to \infty$.

2. In a paper in [11], Lanczos suggests the use of this property for the practical computation of eigenvalues by scanning the interval (a,b), moving λ_0 from a to b. He plots one component of r_n (with respect to the original coordinate system) as a function of λ_0 in n equidistant points. For large n, the abscissas of the maxima of this graph are approximations to the eigenvalues. Lanczos takes the Chebyshev polynomials for the $P_i(\lambda)$. Because those polynomials are given by cosine functions, formula (62) then has the character of a Fourier transformation.

3. Purification of a vector k. The filtering by kernel polynomials may facilitate the computation of the dominant eigenvalue λ_m by the well-known *power method*. (By dominant eigenvalue is understood the eigenvalue next to b.)



Iterating with the kernel polynomial $K_n(b,A)$, one gets a vector r_n having a large component with respect to the eigenvector corresponding to λ_m, the other components, r_{nj}, being small again (for large n and not too high a j). Therefore, the vectors r_n, Ar_n, A^2r_n, A^3r_n, . . . will tend more rapidly to the dominant eigenvector than they do without preceding purification. From the computational point of view it is agreeable to replace matrix A by $A-bE$. Then the iteration by the kernel polynomial $K_n(b,\lambda)$ is replaced by the iteration by $K_n(0,\lambda)$, and therefore the simpler routine (61) can be used.

If A is symmetric, it is possible to proceed from the computation of the dominant eigenvector e_m to the computation of e_{m-1} by the following well-known and stable routine. Orthogonalize r_n with respect to e_m, carry out A, orthogonalize again with respect to e_m, and so on.

Spectral transformations. A very powerful tool to compute eigenvalues is provided by the following fact. Let λ_j $(j=1, 2, \ldots, m)$ be the eigenvalues of A, and $f(\lambda)$ a rational function of the variable λ. Then $f(A)$ has the eigenvalues $f(\lambda_j)$. This is to say, the eigenvalues of $f(A)$ are the ordinates of the graph of $f(\lambda)$ corresponding to the abscissas λ_j. In many problems of applied mathematics, we are not interested in all the eigenvalues, but only in the eigenvalues contained in a region of the λ-axis around a given point λ_0. For instance, in solving the problem of critical angular velocity of a rotating shaft, λ_0 is the actual working velocity. Or, in a problem of vibrating membranes and plates where it is necessary to find only some low frequencies, $\lambda_0=0$.

In order to attack such a restricted problem, we introduce the matrices

$$B_n=\frac{K_n(\lambda_0,A)}{K_n(\lambda_0,\lambda_0)}, \tag{66}$$

with the eigenvalues

$$\Lambda_j=\frac{K_n(\lambda_0,\lambda_j)}{K_n(\lambda_0,\lambda_0)}, \qquad j=1, 2, \ldots, m. \tag{67}$$

The graph of the polynomial $K_n(\lambda_0,\lambda)$ with its peak at λ_0 shows immediately that by this spectral transformation, the eigenvalue λ_j next to λ_0 is transformed into the dominant eigenvalue Λ_j of B_n. (To perform

this, n need not be very large.) Therefore Λ_j may be computed by the power method, constructing with an arbitrary vector k the sequence

$$k,\ B_n k,\ B_n^2 k,\ B_n^3 k,\ \ldots . \tag{68}$$

The computational procedure is as follows. From (66),

$$B_n k = \frac{K_n(\lambda_0, A)}{K_n(\lambda_0, \lambda_0)} \cdot k = r_n$$

in the notation of formula (59). Hence every multiplication by B in (68) is equivalent to n steps of the iteration (61) or (64). It is obvious that the same technique is perfectly useful in order to *separate close eigenvalues* or to *refine* a rough *estimate* λ_0 of an eigenvalue λ_j.

For the latter purpose Wielandt suggested his *fractional iteration*. He takes $f(\lambda)=1/(\lambda-\lambda_0)$. The graph of this function is a hyperbola with vertical asymptote at λ_0, and therefore the eigenvalue λ_j next to λ_0 is transformed by the spectral transformation into a highly dominant value. The analog to B_n here is $B=(A-\lambda_0 E)^{-1}$, and in order to compute $x=Bk$ we have to solve the linear system $(A-\lambda_0 E)x=k$. The method suffers a little from the fact that this system is highly ill-conditioned if the estimate λ_0 is already close to the wanted eigenvalue, λ_j.

The characteristic polynomial. Let us assume that the given m-row matrix A is symmetric, and let k_0 be an arbitrary vector with components k_{0j} in the *orthogonal* coordinate system of the principal axis of A. Again we choose the density of Dirac's type,

$$\rho(\lambda)=\sum_{j=1}^{m} k_{0j}^2 \delta(\lambda-\lambda_j),$$

where the λ_j are the eigenvalues of A. If k_0 is not a very particular vector, we may assume that $k_{0j}\neq 0$ for $j=1, 2, \ldots, m$. From theorem 1 it follows at once that the last polynomial, $P_m(\lambda)$, of the orthogonal set corresponding to this density has its roots at the eigenvalues and is therefore the characteristic polynomial. In order to compute it we construct sequence (62) of vectors $k_i=P_i(A)k_0$. From the orthogonality of the polynomials it follows, for $i\neq l$,

$$\begin{aligned} 0&=\int P_i(\lambda)P_l(\lambda)\rho(\lambda)d\lambda \\ &=\int P_i(\lambda)P_l(\lambda)\sum_{j=0}^{m} k_{0j}^2\delta(\lambda-\lambda_j)d\lambda \\ &=\sum_{j=0}^{m} k_{0j}^2 \int P_i(\lambda)P_l(\lambda)\delta(\lambda-\lambda_j)d\lambda \\ &=\sum_{j=0}^{m} k_{0j}^2 P_i(\lambda_j)P_j(\lambda_j) \\ &=\sum_{j=0}^{m} k_{ij}k_{lj}=(k_i, k_l), \end{aligned} \tag{69}$$

where k_{ij} is the jth component of k_i. Therefore, the *vectors* k_i build an *orthogonal* set. (This may be used in order to correct rounding-off errors.) From this fact and the recursion formula (64) there immediately follows Lanczos' rule,

$$k_{i+1}=Ak_i-\alpha_{i+1}k_i-\beta_i k_{i-1},$$

where

$$\alpha_{i+1}=\frac{(Ak_i,k_i)}{|k_i|^2}, \qquad \beta_i=\frac{(Ak_i,k_{i-1})}{|k_{i-1}|^2}, \qquad (\beta_0=0), \tag{70}$$

$$P_{i+1}(\lambda)=\lambda P_i(\lambda)-\alpha_{i+1}P_i(\lambda)-\beta_i P_{i-1}(\lambda),$$

and $P_m(\lambda)$ is the characteristic polynomial.

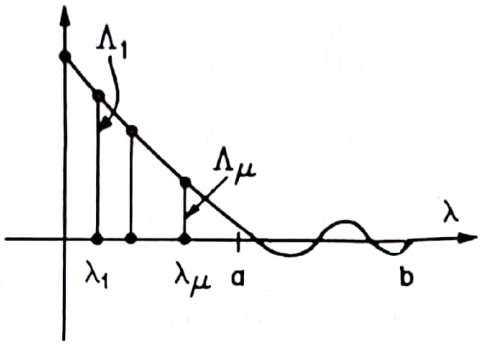
After computation of the eigenvalues as roots of $P_m(\lambda)$, we may construct the *eigenvectors* e_j by (62), observing that

$$N_i=\int P_i(\lambda)^2\rho(\lambda)d\lambda=(k_i,k_i)=|k_i|^2,$$

which follows from (69). Thus

$$e_j=\sum_{i=0}^{m-1}\frac{P_i(\lambda_j)}{|k_i|^2}k_i. \tag{71}$$

For nonsymmetric matrices, the process of orthogonalization of the vectors k_i has to be replaced by a process of biorthogonalization as developed in Lanczos' original paper [7]. This covers even the case of complex eigenvalues. As it stands, Lanczos' algorithm can only be successful for low-order matrices with nicely separated eigenvalues. For larger matrices the rounding-off errors destroy quickly the orthogonality of the vectors. As in solving linear equations, it is necessary to find for such matrices a suitable *combination of the methods* available. Let us be explicit in the following example. Assume that the problem is to compute the lower frequencies of a vibrating plate. We use the difference technique, constructing a net in the region of the plate. If we are careful, the resulting matrix A is symmetric and definite. It is easy to find an upper bound b for the eigenvalues, but the desired low eigenvalues cluster near the origin $\lambda=0$, and should be separated before an attempt is made to compute them.



In order to perform this, we choose a bound $a>0$, which leaves approximately the desired number μ of low eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_\mu$ to the left of it. Then we make a spectral transformation (66) with $\lambda_0=0$:

$$B_n=\frac{K_n(0,A)}{K_n(0,0)}, \tag{72}$$

taking as polynomial $K_n(0,\lambda)$, a high Chebyshev polynomial adapted to the interval (a,b). The vector $r_n=B_nk$ is then given by the recursion

$$\Delta r_n=\frac{1}{\cosh(n+1)\omega}\left\{\cosh(n-1)\omega\cdot\Delta r_{n-1}-\frac{4}{b-a}\cosh n\omega\cdot Ar_n\right\}, \qquad (r_0=k), \tag{73}$$

with

$$\cosh\omega=\frac{b+a}{b-a},$$

as follows from eq (41), (43), and (47). For $n=0$, this formula must be replaced by $\Delta r_0=-2Ar_0/(b+a)$. This spectral transformation has three highly desirable effects. First of all, the clustered eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_\mu$ are transformed into separate values $\Lambda_1,\ldots,\Lambda_\mu$. Furthermore, Λ_1 is the dominant eigenvalue of the transformed matrix B_n, and $\Lambda_1,\Lambda_2,\ldots,\Lambda_\mu$ are arranged in decreasing order.

Lastly, the undesired eigenvalues inside (a,b) are transformed into a cluster of small eigenvalues of B_n around the origin $\Lambda=0$, and with satisfactory accuracy we may assume that they coincide with the origin.

Now we switch to Lanczos' routine, replacing A by B_n in formulas (70). The corresponding density function of Dirac's type has only $(\mu+1)$ peaks, namely, the points $\Lambda_1, \Lambda_2, \ldots, \Lambda_\mu$ and the origin.

It follows again from theorem 1 that the polynomial $P_{\mu+1}(\Lambda)$ of Lanczos has the desired roots $\Lambda_1, \Lambda_2, \ldots, \Lambda_\mu$. Of course this is only approximately true. In practice it is necessary to go on computing $P_{\mu+2}, P_{\mu+3}, \ldots$, until the μ dominant roots of those polynomials do not change any more.

Instead of (72), Wielandt's transformation with $\lambda_0=0$ could be used, replacing A by $B=A^{-1}$. This is to say, the *partial differential equation* of the vibrating plate is replaced by the *integral equation* having Green's function as a kernel. In Zürich the critical angular velocities of rotating shafts were computed by such a combination of spectral transformation and Lanczos' method, and this combination was found to be perfectly safe for about the five lowest eigenvalues.

5. The Quotient-Difference Algorithm

In the general line of our investigations, the quotient-difference algorithm may be introduced as a method for constructing orthogonal polynomials $P_n(\lambda)$ and their kernel polynomials with respect to a given density function $\rho(\lambda)$ in an interval $0\leq a<b$. In [5] the following procedure was developed for this purpose. Let us replace in the basic recursion formula (7),

$$P_{n+1}(\lambda)=\lambda P_n(\lambda)-\alpha_{n+1}P_n(\lambda)-\beta_n P_{n-1}(\lambda), \qquad (\alpha_0=\beta_0=0), \tag{74}$$

the two sequences α_n, β_n of numbers by two other sequences q_n, e_n, defined by

$$q_n+e_{n-1}=\alpha_n, \qquad q_n e_n=\beta_n \qquad (q_0=e_0=0). \tag{75}$$

The q_n, e_n are determined uniquely by the α_n, β_n. Thus

$$P_{n+1}=\lambda P_n-(q_{n+1}+e_n)P_n-q_n e_n P_{n-1}, \tag{76}$$

which may be written,

$$\frac{P_{n+1}+q_{n+1}P_n}{\lambda}=P_n-e_n\frac{P_n+q_nP_{n-1}}{\lambda}.$$

Introducing the functions

$$P'_n(\lambda)=\frac{P_{n+1}+q_{n+1}P_n}{\lambda}, \tag{77}$$

this is

$$P'_n(\lambda)=P_n(\lambda)-e_nP'_{n-1}(\lambda). \tag{78}$$

From this it follows by induction that the $P'_n(\lambda)$ are again polynomials, and (74) splits into eq (77) and (78),

$$P_{n+1}(\lambda)=\lambda P'_n(\lambda)-q_{n+1}P_n(\lambda) \tag{79}$$

$$P'_n(\lambda)=P_n(\lambda)-e_nP'_{n-1}(\lambda), \tag{80}$$

for computing the two sequences P_n, P'_n *simultaneously* by recursion, starting from $P_0=1$, $P'_0=1$. We prove now that the $P'_n(\lambda)$ also build an orthogonal set, but with respect to the density $\lambda\rho(\lambda)$. Indeed, from (79),

$$\int P'_iP'_k\lambda\rho d\lambda=\int P_{i+1}P'_k\rho d\lambda+q_{i+1}\int P_iP'_k\rho d\lambda.$$

Both integrals on the right side vanish for $k<i$ because the polynomial P_i is orthogonal to any polynomial of lower degree, as follows from (20).

When this result is compared with theorem 3, it follows at once that the *polynomials* $P'_n(\lambda)$ are nothing else than the *kernel polynomials* $K_n(0, \lambda)$ up to a constant factor. From the orthogonality of

the $P'_n(\lambda)$ it follows that those polynomials satisfy again a three-term recurrence formula of type (76):

$$P'_{n+1}=\lambda P'_n-(q'_{n+1}+e'_n)P'_n-q'_n e'_n P'_{n-1}. \tag{81}$$

In order to compute the new coefficients q'_n, e'_n, we use (79) and (80):

$$P'_{n+1}=P_{n+1}-e_{n+1}P'_n=(\lambda P'_n-q_{n+1}P_n)-e_{n+1}P'_n.$$

Using (80) again,

$$\begin{aligned} P'_{n+1}&=\lambda P'_n-q_{n+1}(P'_n+e_n P'_{n-1})-e_{n+1}P'_n \\ &=\lambda P'_n-(q_{n+1}+e_{n+1})P'_n-q_{n+1}e_n P'_{n-1}. \end{aligned}$$

Comparing this with (81), we find the important rules

$$q'_{n+1}+e'_n=q_{n+1}+e_{n+1}, \qquad q'_n e'_n=q_{n+1}e_n. \tag{82}$$

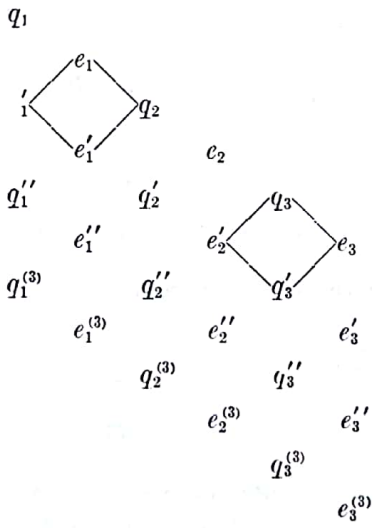
A first application is the construction of the recurrence relations of the kernel polynomials $K_n(0,\lambda)$ from the recurrence relation of the basic polynomials $P_n(\lambda)$. This can be used in order to establish formula (39), which is the principal tool in this theory of solving linear equations.

Formulas (82) are a little hard to remember. In order to facilitate their use let us work out the following table. The numbers q_n, e_n and q'_n, e'_n are recorded in two descending diagonals (see the diagram below). Then the formulas (82) may be called the "rhombus rules" because they can be expressed in the following form:

1. Rhombus centered in a q-column. The sum of the two lower elements equals the sum of the two upper elements.

2. Rhombus centered in an e-column. The product of the two lower elements equals the product of the two upper elements.

It is obvious that this scheme can be extended in the vertical direction down the page, thus yielding a two-dimensional arrangement of numbers,



with the third diagonal containing, for instance, the coefficients of the recurrence formula of polynomials $P''_n(\lambda)$ which are orthogonal with respect to the density $\lambda^2\rho(\lambda)$ and may be called "second kernel polynomials." In general the $(k+1)$th diagonal corresponds to a set of polynomials $P^{(k)}_n(\lambda)$, orthogonal with respect to the density $\lambda^k\rho(\lambda)$ (higher kernel polynomials).

By the rhombus rules this scheme is determined uniquely if the first diagonal is known (observe $e^{(k)}_0=0$) and may be computed automatically using those rules.

There is, however, a second possibility to compute the table, namely, from its first column. From

(75) and (8) it follows that

$$q_1=\alpha_1=\frac{\int\lambda P_0^2\rho d\lambda}{\int P_0^2\rho d\lambda}=\frac{\int\lambda\rho d\lambda}{\int\rho d\lambda}.$$

In the lower diagonals, density ρ must be replaced by $\lambda^k\rho(\lambda)$. Therefore,

$$q_1^{(k)}=\frac{\int\lambda^{k+1}\rho d\lambda}{\int\lambda^k\rho d\lambda}.$$

Let us introduce the *moments* of the given density function, $\rho(\lambda)$:

$$s^{(k)}=\int\lambda^k\rho(\lambda)d\lambda. \tag{83}$$

Then we have

$$q_1^{(k)}=\frac{s^{(k+1)}}{s^{(k)}}. \tag{84}$$

Let us resume. *The quotient-difference table of a given density function, ρ, is a two-dimensional arrangement of numbers. In the first column stand the quotients of two consecutive moments of ρ. The first diagonal contains the coefficients of the recurrence formula of the orthogonal polynomials with respect to ρ. Any other diagonal contains the recurrence coefficients of a higher kernel polynomial. The table may be computed either from the first column or from the first diagonal using the rhombus rules.*

EXAMPLE. Take the density $\rho(\lambda)=1$ in the interval (0,1) (Legendre polynomials). The moments are

$$s^{(k)}=\int_0^1\lambda^k d\lambda=\frac{1}{k+1}.$$

Computed from its first column, the table stands as follows:

q_1	e_1	q_2	e_2	q_3
$\frac{1}{2}$				
	$\frac{1}{6}$			
$\frac{2}{3}$		$\frac{1}{3}$		
	$\frac{1}{12}$		$\frac{1}{5}$	
$\frac{3}{4}$		$\frac{9}{20}$		$\frac{3}{10}$
	$\frac{1}{20}$		$\frac{2}{15}$	
$\frac{4}{5}$		$\frac{8}{15}$		
	$\frac{1}{30}$			
$\frac{5}{6}$				

From the first diagonal and (75),

$$\alpha_1=\frac{1}{2}, \qquad \alpha_2=\frac{1}{2}, \qquad \alpha_3=\frac{1}{2},$$

$$\beta_1=\frac{1}{12}, \qquad \beta_2=\frac{1}{15}.$$

From (74) we get the polynomials of the orthogonal set,

$$P_0=1, \qquad P_1=\lambda-\frac{1}{2}, \qquad P_2=\lambda^2-\lambda+\frac{1}{6}, \qquad P_3=\lambda^3-\frac{3}{2}\lambda^2+\frac{3}{5}\lambda-\frac{1}{20}.$$

Density of Dirac's type. Again we investigate the special case,

$$\rho(\lambda)=\sum_{j=1}^{m} k_j^2\delta(\lambda-\lambda_j), \tag{85}$$

where the λ_j are given abscissas of the λ-axis and k_j given constants not equal to zero. From theorem 1 we learn that $P_m(\lambda)$ has its zeros at the peaks $\lambda_1, \lambda_2, \ldots, \lambda_m$. But the same statement is true for any higher polynomial $P_m^{(k)}(\lambda)$, because for a fixed k the polynomials $P_i^{(k)}(\lambda)$ build the orthogonal set with respect to the density $\lambda^k\rho(\lambda)$, which has its peaks at the same points $\lambda_1, \lambda_2, \ldots, \lambda_m$. Furthermore, all these polynomials have the coefficient 1 in the highest power of λ. Therefore the polynomials $P_m, P_m', \ldots, P_m^{(k)}, \ldots$ are identical. Let us now use formula (80). In the general case it is

$$P_i^{(k+1)}(\lambda)=P_i^{(k)}(\lambda)-e_i^{(k)}P_{i-1}^{(k+1)}(\lambda).$$

For $i=m$ we have $P_m^{(k+1)}=P_m^{(k)}$; thus, the formula gives the result $e_m^{(k)}=0$ for any k.

THEOREM 8. *In the case of a density of Dirac's type having m peaks, the column of the numbers e_m in the quotient-difference table vanishes identically.* Hence in this case the table is bounded from the right by a vertical line.

The importance of the quotient-difference algorithm is due to the following theorem of Rutishauser.

THEOREM 9. *In the case of a density of Dirac's type having m peaks, the numbers in the column q_i converge to the abscissa of the ith peak. This is the ith root of the "last" polynomial $P_m(\lambda)$ of the orthogonal set.* (The peaks are supposed to be arranged in descending order.)

For the column q_1 this follows at once from (84). Indeed,

$$q_1^{(\mu)}=\frac{\int\lambda^{\mu+1}\rho d\lambda}{\int\lambda^{\mu}\rho d\lambda}=\frac{\sum_{j=1}^{m} k_j^2\cdot\lambda_j^{\mu+1}}{\sum_{j=1}^{m} k_j^2\cdot\lambda_j^{\mu}} \tag{86}$$

or

$$q_1^{(\mu)}=\lambda_1\frac{1+\left(\frac{k_2}{k_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{\mu+1}+\ldots+\left(\frac{k_m}{k_1}\right)^2\left(\frac{\lambda_m}{\lambda_1}\right)^{\mu+1}}{1+\left(\frac{k_2}{k_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{\mu}+\ldots+\left(\frac{k_m}{k_1}\right)^2\left(\frac{\lambda_m}{\lambda_1}\right)^{\mu}}.$$

Because of the arrangement $\lambda_1>\lambda_2>\ldots>\lambda_m$, the numbers $q_1^{(\mu)}$ have the limit λ_1 for $\mu\to\infty$. In the appendix we give a short proof for the second column, q_2, leaving the general case to the reader as an exercise.

Application to the eigenvalue problem. Let A be a symmetric positive definite matrix and k_0 an arbitrary vector. Denote the eigenvalues of A by $\lambda_1, \lambda_2, \ldots, \lambda_m$ and the components of k_0 in the system of principal axis by k_{0j}. In section 3 we took as a basis for developing Lanczos' rule the density,

$$\rho(\lambda)=\sum_{j=1}^{m} k_{0j}^2\delta(\lambda-\lambda_j). \tag{87}$$

Adopting this density we find for the moments (83),

$$s^{(k)}=\int \lambda^k \rho(\lambda)d\lambda=\sum_{(j)}\lambda_j^k k_{0j}^2=(A^k k_0,k_0). \tag{88}$$

Thus the moments are the well-known *Schwarzian constants* of matrix A with respect to the vector k_0. From those constants one might start the quotient-difference table by working from left to right, always using the rhombus rules. From theorem 9, it follows immediately that the desired eigenvalues of A are the limits of the q-columns. However, it turns out that this numerical procedure is in most cases utterly unstable, and therefore only a few dominant eigenvalues are obtained. In order to avoid this difficulty, let us compute the table not from its first column but from its first diagonal. The diagonal has to be constructed from formulas (75).

$$q_n=\alpha_n-e_{n-1}, \qquad e_n=\frac{\beta_n}{q_n}, \tag{89}$$

the numbers α_n,β_n being given by Lanczos' rule (70). Then one works from this first diagonal down, always remembering that the column e_m vanishes as stated in theorem (8). This procedure yields the eigenvalues as limits of the q-columns *without computation of the characteristic polynomial.*

The quotient-difference algorithm may therefore be regarded as a *two-dimensional link* of the one-dimensional sequences of Schwarzian and Lanczos constants.

The quotient-difference algorithm was developed by Rutishauser in [8]. His paper [9] is devoted to mathematical applications outside the field of eigenvalue problems. In [10] he discusses the construction of the eigenvectors. He finds them as limits of vector sequences contained in a two-dimensional arrangement of vectors, that is to say, in a three-dimensional table of numbers.

In closing, the author wishes to point out that the solution of a difficult eigenvalue problem may well be divided into three independent parts:

1. Preparation of the given matrix by spectral transformation, as pointed out in section 4.
2. Computation of fairly good estimates of eigenvalues and eigenvectors. This can be done by Lanczos' rule together with the quotient-difference algorithm, which avoids the computation of polynomials.
3. Refinement of each eigenvector individually. This can be done by the methods of section 4, including the good old power method.

6. Appendix. Rutishauser's Theorem

In order to prove theorem 9 for the second column q_2, we attach to the given density (85) the rational function

$$f(\lambda)=\frac{k_1^2}{\lambda-\lambda_1}+\frac{k_2^2}{\lambda-\lambda_2}+\cdots+\frac{k_m^2}{\lambda-\lambda_m}, \tag{90}$$

having its poles at the peaks of the density function. The moments of the density are

$$s^{(k)}=\int \lambda^k \rho(\lambda)d\lambda=\sum_{(j)}\lambda_j^k k_j^2,$$

and are therefore the coefficients of the power series expansion of $f(\lambda)$ at the point $\lambda=\infty$:

$$f(\lambda)=\frac{s^{(0)}}{\lambda}+\frac{s^{(1)}}{\lambda^2}+\frac{s^{(2)}}{\lambda^3}+\cdots. \tag{91}$$

With $z=1/\lambda$, we are therefore reduced to the problem of computing the poles of a rational function $f(z)$ from its power series

$$f(z)=\sum_{n=1}^{\infty}s^{(n)}z^n. \tag{92}$$

460176—58——4

There is a well-known algorithm available to perform this (Bernoulli, König, Aitken). It can be found in Householder's book [3, p. 104–106 and 116]. We have to build the determinants,

$$H_k^{(2)}=\begin{vmatrix} s^{(k)} & s^{(k-1)} \\ s^{(k+1)} & s^{(k)} \end{vmatrix}. \tag{93}$$

Then

$$\lambda_1\lambda_2=\lim \frac{H_{k+1}^{(2)}}{H_k^{(2)}}, \qquad k\to\infty. \tag{94}$$

Together with our elementary result,

$$\lambda_1=\lim q_1^{(k)}, \qquad k\to\infty, \tag{95}$$

this yields λ_1 and λ_2. From (93) and (84) it follows by straightforward computing that

$$\frac{H_{k+1}^{(2)}}{H_k^{(2)}}=q_1^{(k)}q_1^{(k-1)}\,\frac{q_1^{(k+1)}-q_1^{(k)}}{q_1^{(k)}-q_1^{(k-1)}}.$$

By the first rhombus rule,

$$\frac{H_{k+1}^{(2)}}{H_k^{(2)}}=q_1^{(k)}q_1^{(k-1)}\,\frac{e_1^{(k)}}{e_1^{(k-1)}},$$

and by the second rhombus rule,

$$\frac{e_1^{(k)}}{e_1^{(k-1)}}=\frac{q_2^{(k-1)}}{q_1^{(k)}};$$

thus

$$\frac{H_{k+1}^{(2)}}{H_k^{(2)}}=q_1^{(k-1)}q_2^{(k-1)}.$$

Taking into account (94) and (95), we obtain

$$\lambda_1\lambda_2=\lambda_1\cdot\lim q_2^{(k-1)}.$$

Thus $\lambda_2=\lim q_2^{(k)}$, which is Rutishauser's statement concerning the second column. The proof for the other columns must be based on the higher determinants $H_k^{(i)}$, as introduced in [3, p. 116]. The task of establishing their connection with the numbers of the quotient-difference table can be facilitated by using the theory of continued fractions as pointed out in [8].

7. References

[1] G. Szegö, Orthogonal polynomials, Am. Math. Soc. Colloquium Publications **23** (1939).
[2] D. Young, On Richardson's method for solving linear systems with positive definite matrices, J. Math. Phys. **32**, 243–255 (1953).
[3] A. S. Householder, Principles of numerical analysis, p. 149 (McGraw-Hill Book Co., Inc., New York, N. Y., 1953).
[4] E. Stiefel, Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme, Comment. Math. Helv. **29**, 157–179 (1955).
[5] M. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Research NBS **49**, 409 (1952) RP2379.
[6] E. Stiefel, Ausgleichung ohne Aufstellung der Gauss'schen Normalgleichungen, Wiss. Z. der Technische Hochschule (Dresden) **2**, 441 (1952–53).
[7] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Research NBS **45**, 255 (1950) RP2133.
[8] H. Rutishauser, Der Quotienten-Differenzen-Algorithmus, Z. angew. Math. u. Phys. **5**, 234–251 (1954).
[9] H. Rutishauser, Anwendungen des Quotienten-Differenzen-Algorithmus, Z. angew. Math. u. Phys. **5**, 496–507 (1954).
[10] H. Rutishauser, Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten-Differenzen-Algorithmus, Z. angew. Math. u. Phys. **6**, 387–401 (1955).
[11] C. Lanczos, Spectroscopic eigenvalue analysis, J. Wash. Acad. Sci. **45**, 315–323 (1955).