Faculty of Civil Engineering
Belgrade
Master Study
COMPUTATIONAL ENGINEERING

Faculty of Civil Engineering and Architecture
Niš
Doctoral Study

LECTURES

LESSON V

5. Eigensystems

5.0 Introduction

An $n \times n$ matrix $\mathbf{A}$ is said to have an *eigenvector* $\vec{\mathbf{x}}$ and corresponding *eigenvalue* λ if

$$\mathbf{A} \cdot \vec{\mathbf{x}} = \lambda \vec{\mathbf{x}} \tag{5.0.1}$$

Obviously any multiple of an eigenvector $\vec{\mathbf{x}}$ will also be an eigenvector, but we won't consider such multiples as being distinct eigenvectors. (The zero vector is not considered to be an eigenvector at all). Evidently (5.0.1) can hold only if

$$det|\mathbf{A} - \lambda \mathbf{I}| = \mathbf{0}, \tag{5.0.2}$$

which, if expanded out, is an nth degree polynomial in λ whose roots are the eigenvalues. This proves that there are always n (not necessarily distinct) eigenvalues. Equal eigenvalues coming from multiple roots are called *degenerate*. Root searching in the characteristic equation (5.0.2) is usually a very poor computational method for finding eigenvalues (see [2], pp. 449-453).

The above two equations also prove that every one of the n eigenvalues has a (not necessarily distinct) corresponding eigenvector: If λ is set to an eigenvalue, then the matrix $\mathbf{A} - \lambda \mathbf{I}$ is singular, and we know that every singular matrix has at least one nonzero vector in its null-space (consider singular value decomposition).

If you add $\tau \mathbf{x}$ to both sides of (5.0.1), you will easily see that the eigenvalues of any matrix can be changed or shifted by an additive constant τ by adding to the matrix that constant times the identity matrix. The eigenvectors are unchanged by this shift. Shifting, as we will see, is an important part of many algorithms for computing eigenvalues. We see also that there is no special significance to a zero eigenvalue. Any eigenvalue can be shifted to zero, or any zero eigenvalue can be shifted away from zero.

Definitions

A matrix is called symmetric if it is equal to its transpose,

$$\mathbf{A} = \mathbf{A}^T \quad \text{or} \quad a_{ij} = a_{ji} \tag{5.0.3}$$

It is called Hermitian or self-adjoint if it equals to the complex-conjugate its transpose (its Hermitian conjugate, denoted by "†")

$$\mathbf{A} = \mathbf{A}^\dagger \quad \text{or} \quad a_{ij} = a_{ji}^* \tag{5.0.4}$$

It is termed *orthogonal* if its transpose equals its inverse

$$\mathbf{A}^T \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{A}^T = I \tag{5.0.5}$$

and *unitary* if its Hermitian conjugate equals its inverse. Finally, a matrix is called *normal* if it *commutes* with its Hermitian conjugate,

$$\mathbf{A} \cdot \mathbf{A}^\dagger = \mathbf{A}^\dagger \cdot \mathbf{A} \tag{5.0.6}$$

For real matrices, Hermitian means the same as symmetric, unitary means the same as orthogonal, and both of these distinct classes are normal.

The reason that "Hermitian" is an important concept has to do with eigenvalues. The eigenvalues of a Hermitian matrix are all real. In particular, the eigenvalues of a real symmetric matrix are all real. Contrariwise, the eigenvalues of a real nonsymmetric matrix may include real values, but may also include pairs of conjugate values; and the eigenvalues of a complex matrix that is not Hermitian will in general be complex.

The reason that "normal" is an important concept has to do with the eigenvectors. The eigenvectors of a normal matrix with non-degenerate (i.e., distinct) eigenvalues are complete and orthogonal, spanning the n-dimensional vector space. For a normal matrix with degenerate eigenvalues, we have the additional freedom of replacing the eigenvectors corresponding to a degenerate eigenvalue by linear combinations of themselves. Using this freedom, we can always perform Gramm-Schmidt orthogonalization and find a set of eigenvectors that are complete and orthogonal, just as in the nondegenerate case. The matrix whose columns are an orthonormal set of eigenvectors is evidently unitary. A special case is that the matrix of eigenvectors of a real symmetric matrix is orthogonal, since the eigenvectors of that matrix are all real.

When a matrix is not normal, as typified by any random, nonsymmetric, real matrix, then in general we cannot find any orthonormal set of eigenvectors, nor even any pairs of eigenvectors that are orthogonal (except perhaps by rare chance). While the n non-orthonormal eigenvectors will "usually" span the n-dimensional vector space, they do not always do so; that is, the eigenvectors are not always complete. Such a matrix is said to be defective.

Left and Right Eigenvectors

While the eigenvectors of a non-normal matrix are not particularly orthogonal among themselves, they do have an orthogonality relation with a different set of vectors, which we must now define. Up to now our eigenvectors have been column vectors that are multiplied to the right of a matrix **A**, as in (5.0.1). These, more explicitly, are termed *right eigenvectors*. We could also, however, try to find vectors, which multiply **A** to the left and satisfy

$$\vec{x} \cdot \mathbf{A} = \lambda \vec{x} \tag{5.0.7}$$

These are called *left eigenvectors*. By taking the transpose of (5.0.7), one can see that every left eigenvector is the transpose of a right eigenvector of the transpose of **A** . Now by comparing to (5.0.2), and using the fact that the determinant of a matrix equals the determinant of its transpose, we also see that the left and right eigenvalues of **A** are identical.

If the matrix **A** is symmetric, then the left and right eigenvectors are just transposes of each other, that is, have the same numerical values as components. Likewise, if the matrix is self-adjoint, the left and right eigenvectors are Hermitian conjugates of each other. For the general non-normal case, however, we have the following calculation: Let

$\mathbf{X_R}$ be the matrix formed by columns from the right eigenvectors, and $\mathbf{X_L}$ be the matrix formed by rows from the left eigenvectors. Then (5.0.1) and (5.0.7) can be rewritten as

$$\mathbf{A} \cdot \mathbf{X}_R = \mathbf{X}_R \cdot diag(\lambda_1 \ldots \lambda_N); \qquad \mathbf{X}_L \cdot \mathbf{A} = diag(\lambda_1 \ldots \lambda_N) \cdot \mathbf{X}_L \tag{5.0.8}$$

Multiplying the first of these equations on the left by $\mathbf{X}_L$, the second on the right by $\mathbf{X}_R$, and subtracting the two, gives

$$(\mathbf{X}_L \cdot \mathbf{X}_R) \cdot diag(\lambda_1 \ldots \lambda_N) = diag(\lambda_1 \ldots \lambda_N) \cdot (\mathbf{X}_L \cdot \mathbf{X}_R) \tag{5.0.9}$$

This says that the matrix of dot products of the left and right eigenvectors commutes with the diagonal matrix of eigenvalues. But the only matrices that commute with a diagonal matrix of distinct elements are themselves diagonal. Thus, if the eigenvalues are non-degenerate, each left eigenvector is orthogonal to all right eigenvectors except its corresponding one, and vice versa. By choice of normalization, the dot products of corresponding left and right eigenvectors can always be made unity for any matrix with non-degenerate eigenvalues.

If some eigenvalues are degenerate, then either the left or the right eigenvectors corresponding to a degenerate eigenvalue must be linearly combined among themselves to achieve orthogonality with the right or left ones, respectively. This can always be done by a procedure akin to Gram-Schmidt orthogonalization. The normalization can then be adjusted to give unity for the nonzero dot products between corresponding left and right eigenvectors. If the dot product of corresponding left and right eigenvectors is zero at this stage, then you have a case where the eigenvectors are incomplete. Note that incomplete eigenvectors can occur only where there are degenerate eigenvalues, but do not always occur in such cases (in fact, never occur for the class of "normal" matrices).

In both the degenerate and non-degenerate cases, the final normalization to unity of all nonzero dot products produces the result: The matrix whose rows are left eigenvectors is the inverse matrix of the matrix whose columns are right eigenvectors, *if the inverse exists.*

Diagonalization of a Matrix

Multiplying the first equation in (5.0.8) by $\mathbf{X}_L$, and using the fact that $\mathbf{X}_L$ and $\mathbf{X}_R$ are matrix inverses, we get

$$\mathbf{X}_R^{-1} \cdot \mathbf{A} \cdot \mathbf{X}_R = diag(\lambda_1 \ldots \lambda_N). \tag{5.0.10}$$

This is a particular case of a *similarity transform* of the matrix $\mathbf{A}$,

$$\mathbf{A} \rightarrow \mathbf{Z}^{-1} \cdot \mathbf{A} \cdot \mathbf{Z} \tag{5.0.11}$$

for some transformation matrix $\mathbf{Z}$. Similarity transformations play a crucial role in the computation of eigenvalues, because they leave the eigenvalues of a matrix unchanged. This is easily seen from

$$\begin{aligned} det|\mathbf{Z}^{-1} \cdot \mathbf{A} \cdot \mathbf{Z} - \lambda\mathbf{I}| &= det|\mathbf{Z}^{-1} \cdot (\mathbf{A} - \lambda\mathbf{I}) \cdot \mathbf{Z}| \\ &= det|\mathbf{Z}| \; det|\mathbf{A} - \lambda\mathbf{I}| \; det|\mathbf{Z}^{-1}| \\ &= det|\mathbf{A} - \lambda\mathbf{I}| \end{aligned} \tag{5.0.12}$$

Equation (5.0.10) shows that any matrix with complete eigenvectors (which includes all normal matrices and "most" random non-normal ones) can be diagonalized by a similarity transformation, that the columns of the transformation matrix that effects the diagonalization are the right eigenvectors, and that the rows of its inverse are the left eigenvectors.

For real, symmetric matrices, the eigenvectors are real and orthonormal, so the transformation matrix is orthogonal. The similarity transformation is then also an *orthogonal transformation* of the form

$$\mathbf{A} \to \mathbf{Z}^T \cdot \mathbf{A} \cdot \mathbf{Z} \tag{5.0.13}$$

While real nonsymmetric matrices can be diagonalized in their usual case of complete eigenvectors, the transformation matrix is not necessarily real. It turns out, however, that a real similarity transformation can ”almost” do the job. It can reduce the matrix down to a form with little two-by-two blocks along the diagonal, all other elements zero. Each two-by-two block corresponds to a complex-conjugate pair of complex eigenvalues.

The ”grand strategy” of virtually all modern eigensystem routines is to nudge the matrix $\mathbf{A}$ towards diagonal form by a sequence of similarity transformations,

$$\begin{aligned}\mathbf{A} &\to \mathbf{P_1}^{-1} \cdot \mathbf{A} \cdot \mathbf{P}_1 \to \mathbf{P}_2^{-1} \cdot \mathbf{P}_1^{-1} \cdot \mathbf{A} \cdot \mathbf{P}_1 \cdot \mathbf{P}_2 \\ &\to \mathbf{P}_3^{-1} \cdot \mathbf{P}_2^{-1} \cdot \mathbf{P}_1^{-1} \cdot \mathbf{A} \cdot \mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \mathbf{P}_3 \to \quad \text{etc.}\end{aligned} \tag{5.0.14}$$

If we get all the way to diagonal form, then the eigenvectors are the columns of the accumulated transformation

$$\mathbf{X}_R = \mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \mathbf{P}_3 \cdot \ldots \tag{5.0.15}$$

Sometimes we do not want to go all the way to diagonal form. For example, if we are interested only in eigenvalues. not eigenvectors, it is enough to transform the matrix $\mathbf{A}$ to be triangular, with all elements below (or above) the diagonal zero. In this case the diagonal elements are already the eigenvalues, as you can see by mentally evaluating (5.0.2) using expansion by minors.

There are two rather different sets of techniques for implementing the strategy (5.0.14). It turns out that they work rather well in combination, so most modern eigensystem routines use both. The first set of techniques constructs individual $\mathbf{P}_i$'s as explicit ”atomic” transformations designed to perform specific tasks, for example zeroing a particular off-diagonal element (Jacobi transformation), or a whole particular row or column (Householder transformation), elimination method). In general, a finite sequence of these simple transformations cannot completely diagonalize a matrix. There are then two choices: either use the finite sequence of transformations to go most of the way (e.g., to some special form like *tridiagonal* or *Hessenberg*) and follow up with the second set of techniques about to be mentioned; or else iterate the finite sequence of simple transformations over and over until the deviation of the matrix from diagonal is negligibly small. This latter approach is conceptually simplest. However, for n greater than ~ 10, it is computationally inefficient by a roughly constant factor ~ 5.

The second set of techniques, called *factorization methods*, is more subtle. Suppose that the matrix $\mathbf{A}$ can be factored into a left factor $\mathbf{F}_L$ and a right factor $\mathbf{F}_R$. Then

$$\mathbf{A} = \mathbf{F}_L \cdot \mathbf{F}_R \quad \text{or equivalently} \quad \mathbf{F}_L^{-1} \cdot \mathbf{A} = \mathbf{F}_R \tag{5.0.16}$$

If we now multiply back together the factors in the reverse order, and use the second equation in (5.0.16) we get

$$\mathbf{F}_R \cdot \mathbf{F}_L = \mathbf{F}_L^{-1} \cdot \mathbf{A} \cdot \mathbf{F}_L \tag{5.0.17}$$

which we recognize as having effected a similarity transformation on $\mathbf{A}$ with the transformation matrix being $\mathbf{F}_L$. The **QR** method which exploits this idea will be explained later.

Factorization methods also do not converge exactly in a finite number of transformations. But the better ones do converge rapidly and reliably, and, when following an appropriate initial reduction by simple similarity transformations, they are the methods of choice. The presented considerations are very important for those dealing with dynamics of construction and seismic engineering, especiallin the phase of modelling and dynamic response computation.

Definitions and theorems regarding eigenvalue problem

For further considerations we need some theorems and definitions, as follows (see [1], pp. 211-213).

Definition 5.0.1. *Let $\mathbf{A} = [a_{ij}]$ be complex square matrix of order n. Every vector $\vec{f} \in \mathbf{C}^n$, which is different from zero-vector, is called eigenvector of matrix $\mathbf{A}$ if exists scalar $\lambda \in \mathbf{C}$ such that holds* (5.0.1). *Scalar λ in* (5.0.1) *is called corresponding eigenvalue. Having in mind that* (5.0.1) *can be presented in the form*

$$(\mathbf{A} - \lambda \mathbf{I})\vec{x} = \vec{\mathbf{0}},$$

we conclude that equation (5.0.1) *has non-trivial solutions (in $\vec{x}$) then and only then if holds* (5.0.2).

Definition 5.0.2. *If $\mathbf{A}$ is square matrix, then polynomial $\lambda \to P(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I})$ is called characteristic polynomial, and corresponding equation $P(\lambda) = 0$ its characteristic equation.*

Let $\mathbf{A} = [a_{ij}]_{n\times n}$. The characteristic polynomials can be expressed in the form

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{12} & & a_{nn}-\lambda \end{vmatrix}$$

or

$$P(\lambda) = (-1)^n(\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \dots + (-1)^{n-1}p_{n-1}\lambda + (-1)^n p_n),$$

where p_k is sum of all principal minors of order k of determinant of matrix $\mathbf{A}$, i.e.

$$p_k = \sum_{1\le i_1<i_2<\dots<i_k\le n} \det(\mathbf{A}^{i_1 i_2 \cdots i_k}_{i_1 i_2 \cdots i_k}).$$

Note that

$$p_1 = \sum_{i=1}^{n} a_{ii} = tr\ \mathbf{A} \quad \text{and} \qquad p_n = \det(\mathbf{A}).$$

Often, in place of characteristic polynomial P is used so known normed characteristic polynomial H, defined by

$$H(\lambda) = (-1)^n P(\lambda) = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \dots + (-1)^n p_n.$$

Eigenvalues of matrix $\mathbf{A}$ (i.e. zeros of polynomial P) $\lambda_i(i = 1,\dots,n)$) will be denoted as $\lambda_i(\mathbf{A})$.

Definition 5.0.3. *The set of all eigenvalues of square matrix $\mathbf{A}$ is called spectrum of that matrix and denoted with $Sp\,(\mathbf{A})$.*

Definition 5.0.4. *Spectral radius $\rho(\mathbf{A})$ of square matrix $\mathbf{A}$ is number*

$$\rho(\mathbf{A}) = \max_i |\lambda_i(\mathbf{A})|.$$

Theorem 5.0.1. *Every matrix is, in matrix sense, null of its characteristic polynomial.*

This theorem is known as Cayley-Hamilton theorem.

Theorem 5.0.2. *Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of matrix $A = [a_{ij}]$ of order n and $x \to Q(x)$ scalar polynomial of degree m. Then*

$$Q(\lambda_1), \ldots, Q(\lambda_n)$$

are eigenvalues of matrix $\mathbf{Q}(\mathbf{A})$.

Theorem 5.0.3. *Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of regular matrix $\mathbf{A}$ of order n. Then*

$$\lambda_1^{-1}, \ldots, \lambda_n^{-1}$$

are eigenvalues of matrix $\mathbf{A}^{-1}$.

Theorem 5.0.4. *Eigenvalues of triangular matrix are equal to diagonal elements.*

The following theorem gives recursive procedure for obtaining characteristic polynomial of tridiagonal matrix.

Theorem 5.0.5. *Let*

$$\mathbf{A}_k = \begin{bmatrix} b_1 & c_1 & 0 & \ldots & 0 \\ a_2 & b_2 & c_2 & & 0 \\ \vdots & & & & \\ 0 & 0 & & & b_k \end{bmatrix} \quad \textit{and} \quad H_k(\lambda) = (-1)^k \det(\mathbf{A}_k - \lambda \mathbf{I}).$$

Normed characteristic polynomial $\lambda \to H(\lambda)$ $(= H_n(\lambda))$ of matrix $\mathbf{A}(= \mathbf{A}_n)$ is to be obtained by recursive procedure

$$H_k(\lambda) = (\lambda - b_k) H_{k-1}(\lambda) - a_{k-1} c_{k-1} H_{k-2}(\lambda) \quad (k = 2, \ldots, n),$$

where $H_0(\lambda) = 1$ and $H_{k-1}(\lambda) = \lambda - b_1$.

Definition 5.0.5. *For matrix $\mathbf{B}$ one says to be similar to matrix $\mathbf{A}$ if there exists at least one regular matrix $\mathbf{C}$ such that*

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}.$$

Theorem 5.0.6. *Similar matrices have identical characteristic polynomials, and therewith identical eigenvalues.*

5.1 Localization of Eigenvalues

A lot of problems reduce to eigenvalue problem. Here we will give some results regarding localization of eigenvalues in complex space (see [1], pp. 290-292).

Theorem 5.1.1 (Gershgorin). *Let $A = [a_{ij}]$ square matrix of order n and C_i $(i = 1, \ldots n)$ discs in complex space with centers in a_{ii} and radiuses $r_i = \sum\limits_{\substack{j=i \\ j \neq i}} |a_{ij}|$, i.e.*

$$C_i = \{z \mid |z - a_{ii}| \leq r_i\} \quad (i = 1, \ldots, n).$$

If we denote with C union of these discs, then all eigenvalues of matrix $\mathbf{A}$ are in C.

Proof. *Let λ be eigenvalue of matrix $\mathbf{A}$, and $\vec{x}$ corresponding eigenvector normalized so that $\|\vec{x}\|_\infty = \max\limits_i |x_i| = |x_m| = 1$. Then $\lambda\vec{x} = \mathbf{A}\vec{x}$, i.e.*

$$(\lambda - a_{ii}) x_i = \sum_{\substack{j=1 \\ j \neq i}} a_{ij} x_j \qquad (i = 1, \ldots, n),$$

wherefrom for $i = m$ we have

$$|\lambda - a_{mm}| \leq \sum_{\substack{j=1 \\ j \neq m}}^{n} |a_{ij}| \cdot |x_j| \leq \sum_{\substack{j=1 \\ j \neq m}}^{n} |a_{ij}| = r_m.$$

■

Well, eigenvalue λ lies in the disc C. Because λ is arbitrary eigenvalue, we conclude that all eigenvalues are in union of discs, i.e. C.

Remark 5.1.1. *Regarding fact that matrix $\mathbf{A}^T$ has same eigenvalues as matrix $\mathbf{A}$, on the basis of previous theorem one can conclude that all eigenvalues of matrix $\mathbf{A}$ are located in the union of D discs*

$$D_j = \{z \mid |z - a_{jj}| \leq s_j\} \quad (j = 1, \ldots, n),$$

where $s_j = \sum\limits_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}|$.

Based on previous one concludes that all eigenvalues of matrix $\mathbf{A}$ lie in the cut of sets C and D.

Theorem 5.1.2. *If m discs from Theorem 5.1.1 form connected area which is isolated from other discs, then exact m eigenvalues of matrix A are located in this area.*

The proof of this theorem could be found in extraordinary monograph of Wilkinson [7].

Example 5.1.1.

Take matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0.1 & -0.1 \\ 0 & 2 & 0.4 \\ -0.2 & 0 & 3 \end{bmatrix}.$$

Based on theorem 5.1.1 eigenvalues are located in discs

$$C_1 = \{z \mid |z - 1| \leq 0.2\}, \quad C_2 = \{z \mid |z - 2| \leq 0.4\}, \quad C_3 = \{z \mid |z - 3| \leq 0.2\}.$$

Note that, based on remark 5.1.1, it follows that discs D_1, D_2, D_3 have radiuses $0.2, 0.1, 0.5$, respectively. By the way, the exact values of eigenvalues, given with seven figures, are $\lambda_1 = 0.9861505,\ \lambda_2 = 2.0078436, \lambda_3 = 3.0060058$, and normed characteristic polynomial is

$$H(\lambda) = \lambda^3 - 6\lambda^2 + 10.98\lambda - 5.952.$$

Theorem on localization of eigenvalues has theoretical and practical importance (for example, for determining initial values at iterative methods, for analysis at perturbation problems, etc).

For determining eigenvalues there are a lot of methods, whereby some of them enable finding of all eigenvalues, and others only some of them, for example, dominating ones, i.e. with maximum modulus. Some of methods perform only determination of coefficients of characteristic polynomial, so that some of methods for solution of algebraic equations have to be used (see Chapter 6). Such approach is not recommended, being in most cases numerically unstable, i.e. ill conditioned. Namely, because the coefficients of characteristic polynomials are, in general, subjects to round-off error, due to ill-conditioning of characteristic polynomials, the big errors in eigenvalues occur.

5.2 Methods for determination of characteristic polynomial

In this section we will induce some methods for determining of characteristic polynomial of matrix $A = [a_{ij}]$,

$$P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}). \tag{5.2.1}$$

As previously said, it is not to be recommended to use polynomial obtained in this way for determining of eigenvalues and eigenvectors of $\mathbf{A}$ except in the case of well conditioned characteristic polynomial.

1. Krylov's method . Instead of (5.2.1) we will consider normalized characteristic polynomial

$$H(\lambda) = (-1)^n P(\lambda) = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \cdots + (-1)^n p_n. \tag{5.2.2}$$

Based on Cayley-Hamilton theorem (Theorem 2.3.1) we have

$$H(\mathbf{A}) = \mathbf{A}^n - p_1\mathbf{A}^{n-1} + p_2\mathbf{A}^{n-2} - \cdots + (-1)^n p_n\mathbf{I} = 0.$$

Let now $\vec{y}^{(0)}$ be arbitrary n-dimensional vector, with which multiply right-hand side of previous equation. Then we get

$$p_1\mathbf{A}^{n-1}\vec{y}^{(0)} - p_2\mathbf{A}^{n-2}\vec{y}^{(0)} + \cdots + (-1)^n p_n\vec{y}^{(0)} = \mathbf{A}^n\vec{y}^{(0)},$$

wherefrom, using coordinate representation, we get the system of linear equations

$$\begin{bmatrix} y_1^{(n-1)} & -y_1^{(n-2)} & \dots & (-1)^{n-1}y_1^{(0)} \\ y_2^{(n-1)} & -y_2^{(n-2)} & \dots & (-1)^{n-1}y_2^{(0)} \\ \vdots & & & \\ y_n^{(n-1)} & -y_n^{(n-2)} & \dots & (-1)^{n-1}y_n^{(0)} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_n^{(n)} \end{bmatrix}, \tag{5.2.3}$$

where we put

$$\vec{y}^{(k)} = \mathbf{A}^{(0)}\vec{y}^{(0)} = \begin{bmatrix} y_1^{(k)} & y_2^{(k)} & \dots y_n^{(k)} \end{bmatrix}^T \quad (k = 1, 2, \dots, n).$$

Note that the degrees of matrix $\mathbf{A}^k$ are not to be calculated, but the recursion of form

$$\vec{y}^{(k)} = \mathbf{A}\vec{y}^{(k-1)} \qquad (k = 1, \dots, n) \tag{5.2.4}$$

should be applied.

Under condition that matrix of obtained system of linear equations (5.2.3) is regular, by solving this system we get the coefficients $p_1, p_2, \dots, p_n$. If the matrix of system is singular, the starting vector $\vec{y}^{(0)}$ has to be changed.

Let us illustrate the application of this method to the finding out the characteristic polynomial of matrix of order four.

Example 5.2.1. *Let*

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & -2 & -1 \\ -1 & 3 & -1 & 0 \\ 1 & -2 & 4 & 1 \\ 3 & 0 & 1 & 3 \end{bmatrix}.$$

If we take $\vec{y}^{(k)} = [1 \quad 0 \quad 0 \quad 0]^T$, using (5.2.4) we get

$$\vec{y}^{(1)} = \mathbf{A}\vec{y}^{(0)} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 3 \end{bmatrix}, \quad \vec{y}^{(2)} = \mathbf{A}\vec{y}^{(1)} = \begin{bmatrix} 2 \\ -7 \\ 12 \\ 19 \end{bmatrix},$$

$$\vec{y}^{(3)} = \mathbf{A}\vec{y}^{(2)} = \begin{bmatrix} -51 \\ -35 \\ 83 \\ 75 \end{bmatrix}, \quad \vec{y}^{(4)} = \mathbf{A}\vec{y}^{(3)} = \begin{bmatrix} -464 \\ -137 \\ 426 \\ 155 \end{bmatrix},$$

Thus, system (5.2.3) becomes

$$\begin{bmatrix} -51 & -2 & 3 & -1 \\ -35 & 7 & -1 & 0 \\ 83 & -12 & 1 & 0 \\ 75 & -19 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} -464 \\ -137 \\ 426 \\ 155 \end{bmatrix},$$

wherefrom we get

$$p_1 = 13,\ p_2 = 67,\ p_3 = 151,\ p_4 = 120.$$

Accordingly, we have

$$H(\lambda) = P(\lambda) = \lambda^4 - 13\lambda^3 + 67\lambda^2 - 151\lambda + 120.$$

Check the previous result out by `Mathematica` session and write a procedure for Krylov's method in `Mathematica` and `Delphi (Pascal)`.

2. Leverrier's method and Faddeev's modification.

Leverrier's method is based on known Newton's formulas for sums of degrees of all zeros of polynomial. Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of matrix $\mathbf{A}$, i.e. zeros of polynomial (5.2.2), where every zero is taken as many times as its multiplicity. Then for sums

$$s_m = \lambda_1^m + \lambda_2^m + \cdots + \lambda_n^m \quad (m = 0, 1, \ldots, n)$$

hold Newton formulas (see, for example, [21, pp. 241-242])

$$s_m - p_1 s_{m-1} + p_2 s_{m-2} - \cdots + (-1)^{m-1} p_{m-1} s_1 + (-1)^m m p_m = 0$$
$$(m = 1, \ldots, n)$$

wherefrom we have

$$\begin{aligned}
p_1 &= s_1,\\
p_2 &= -\frac{1}{2}(s_2 - p_1 s_1),\\
p_3 &= \frac{1}{3}(s_3 - p_1 s_2 + p_2 s_1),\\
&\vdots\\
p_n &= \frac{(-1)^{n-1}}{n}(s_n - p_1 s_{n-1} + \cdots + (-1)^{n-1} p_{n-1} s_1).
\end{aligned}$$

Thus, if all sums are known, we can determine coefficients of characteristic polynomials. Notice that

$$s_1 = \lambda_1 + \lambda_2 + \cdots + \lambda_n = tr\mathbf{A} = \sum_{i=1}^{n} a_{ii}.$$

Since $\lambda_1^m, \lambda_2^m, \ldots, \lambda_n^m$ are eigenvalues of matrix $\mathbf{A}^m$ (see theorem 5.0.2.), we conclude that

$$s_m = tr\mathbf{A}^m.$$

Thus, when $\mathbf{A}^m = [a_{ij}^{(m)}]_{n\times n}$ we have $s_m = \sum\limits_{i=1}^{n} a_{ii}^{(m)}$, whereby degrees of matrices are determined successively using

$$\mathbf{A}^m = \mathbf{A} \cdot \mathbf{A}^{m-1}.$$

Example 5.2.2.

For matrix from previous example we have in turn

$$\mathbf{A}^2 = \mathbf{A}\cdot\mathbf{A} = \begin{bmatrix} 2 & 16 & -17 & -8\\ -7 & 9 & -5 & 0\\ 12 & -12 & 17 & 6\\ 19 & 4 & 1 & 7 \end{bmatrix},$$

$$\mathbf{A}^3 = \mathbf{A}\cdot\mathbf{A}^2 = \begin{bmatrix} -51 & 86 & -98 & -43\\ -35 & 23 & -15 & 2\\ 83 & -46 & 62 & 23\\ 75 & 48 & -31 & 3 \end{bmatrix},$$

$$\mathbf{A}^4 = \mathbf{A}\cdot\mathbf{A}^3 = \begin{bmatrix} -464 & 348 & -411 & -174\\ -137 & 29 & -11 & 26\\ 426 & -96 & 151 & 48\\ 155 & 356 & -319 & -97 \end{bmatrix}.$$

Because of

$$\begin{aligned}
s_1 &= tr\mathbf{A} = 3+3+4+3 = 13,\\
s_2 &= tr\mathbf{A}^2 = 2+9+17+7 = 35,\\
s_3 &= tr\mathbf{A}^3 = -51+23+62+3 = 37,\\
s_4 &= tr\mathbf{A}^4 = -464+29+151-97 = -381,
\end{aligned}$$

we have, in turn

$$\begin{aligned}
&p_1 = 13,\\
&p_2 = -(35 - 13\cdot 13)/2 = 67,\\
&p_3 = (37 - 13\cdot 35 + 67\cdot 13)/3 = 151,\\
&p_4 = -(-381 - 13\cdot 37 + 67\cdot 35 - 151\cdot 13)/4 = 120.
\end{aligned}$$

A modification of Leverrier method have been invented by Faddeev. This modification requires less numerical operations, as follows.

In spite of m-th degree of matrix, $\mathbf{A}^m$, a sequence of matrices $\mathbf{A}_m$ $(m = 1, \ldots, n)$ is obtained using formulas

$$\mathbf{A}_m = \mathbf{A}\mathbf{B}_{m-1}, \quad q_m = \frac{1}{m} tr\mathbf{A}_m, \quad \mathbf{B}_m = \mathbf{A}_m - q_m\mathbf{I},$$

where for $\mathbf{B}_0$ is taken unit matrix.

Using induction one can prove that $q_m = (-1)^{m-1}p_m$ $(m = 1, \ldots, n)$.

Example 5.2.3.

For matrix from previous example, using method of Faddeev, we have successively

$$\mathbf{A}_1 = \mathbf{A},\ q_1 = tr\mathbf{A}_1 = 13,\ \mathbf{B}_1 = \mathbf{A}_1 - 13\mathbf{I}$$

$$= \begin{bmatrix} -10 & 2 & -2 & -1\\ -1 & -10 & -1 & 0\\ 1 & -2 & -9 & 1\\ 3 & 0 & 1 & -10 \end{bmatrix},$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1 = \begin{bmatrix} -37 & -10 & 9 & 5\\ 6 & -30 & 8 & 0\\ -1 & 14 & -35 & -7\\ -20 & 4 & -12 & 32 \end{bmatrix},\ q_2 = \frac{1}{2}tr\mathbf{A}_2 = -67,$$

$$\mathbf{B}_2 = \mathbf{A}_2 + 67\mathbf{I} = \begin{bmatrix} 30 & -10 & 9 & 5\\ 6 & 37 & 8 & 0\\ -1 & 14 & 32 & -7\\ -20 & 4 & -12 & 35 \end{bmatrix},$$

$$\mathbf{A}_3 = \mathbf{A}\mathbf{B}_2 = \begin{bmatrix} 124 & 12 & -9 & -6\\ -11 & 107 & -17 & 2\\ -6 & -24 & 109 & 12\\ 29 & -4 & 23 & 113 \end{bmatrix},$$

$$q_3 = \frac{1}{3}tr\mathbf{A}_3 = 151,\ \mathbf{B}_3 = \mathbf{A}_3 - 151\mathbf{I} = \begin{bmatrix} -27 & 12 & -9 & -6\\ -11 & -44 & -17 & 2\\ -6 & -24 & -42 & 12\\ 29 & -4 & 23 & -38 \end{bmatrix},$$

$$\mathbf{A}_4 = \mathbf{A}\mathbf{B}_3 = \begin{bmatrix} -120 & 0 & 0 & 0\\ 0 & -120 & 0 & 0\\ 0 & 0 & -120 & 0\\ 0 & 0 & 0 & -120 \end{bmatrix},\ q_4 = \frac{1}{4}tr\mathbf{A}_4 = -120$$

Thus, $p_1 = q_1 = 13,\ p_2 = -q_2 = 67,\ p_3 = q_3 = 151,\ p_4 = -q_4 = 120.$

3. Method of Danilevsky.

This method is based on transformation of matrix to so known Frobenius form

$$F = \begin{bmatrix} f_1 & f_2 & \cdots & f_{n-1} & f_n \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & & 1 & 0 \end{bmatrix}, \tag{5.2.5}$$

where matrices **A** and **F** are similar, i.e. there exists such regular matrix **C** that $\mathbf{F} = \mathbf{C}^{-1}\mathbf{AC}$ (see Definition 5.0.5). In regard to fact that similar matrices have identical characteristic polynomials, based on (5.2.5) it is simple to get characteristic polynomial of matrix **A**. Namely, if we develop $\det(\mathbf{F} - \lambda\mathbf{I})$ up to elements of first column, we get

$$P(\lambda) = (f_1 - \lambda)(-\lambda)^{n-1} - f_2(-\lambda)^{n-2} + \cdots + (-1)^{n-1} f_n,$$

i.e.

$$P(\lambda) = (-1)^n(\lambda^n - f_1\lambda^{n-1} - f_2\lambda^{n-2} - \cdots - f_n).$$

Thus, $p_m = (-1)^{m-1} f_m \quad (m = 1, \ldots, n)$.

Let us define for row-vector $\vec{z} = [z_1 \ z_2 \ \cdots \ z_n]$ square matrix of order n

$$G_r(\vec{z} =) \begin{bmatrix} \mathbf{I}_{n-r} & \mathbf{O}_{n-r,r} \\ \mathbf{P}_{r,n-r}(\vec{z}) & \mathbf{R}_r(\vec{z}) \end{bmatrix},$$

where $\mathbf{I}_{n-r}$ unit matrix of order $n-r$, $\mathbf{O}_{n-r,r}$ null-matrix of type $r \times (n-r)$, $\mathbf{P}_{r,n-r}(\vec{z})$ rectangular matrix of type $r \times (n-r)$, which has in first row successively elements $z_1, z_2, \ldots, z_{n-r}$, while all other elements are equal to zero, and finally, $\mathbf{R}_r(\vec{z})$ is square matrix of order r which differs from unit matric only by first row, which contains the residual elements of vector $\vec{z}$, i.e. has in turn, the elements $z_{n-r+1}, \ldots, z_n$.

Suppose that $z_{n-r+1} \neq 0$. Using theorem on inverting of block structured matrices (see [1], p.205) it is easy to find inverse matrix of $G_r(\vec{z})$

$$\mathbf{G}_r^{-1}(\vec{z}) = \begin{bmatrix} \mathbf{I}_{n-r} & \mathbf{O}_{n-r,r} \\ -\dfrac{\mathbf{P}_{r,n-r}(\vec{z})}{z_{n-r+1}} & \mathbf{R}_r^{-1}(\vec{z}) \end{bmatrix},$$

where

$$\mathbf{R}_r(\vec{z}) = \begin{bmatrix} \dfrac{1}{z_{n-r+1}} & -\dfrac{z_{n-r+2}}{z_{n-r+1}} & \cdots & -\dfrac{z_n}{z_{n-r+1}} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

Let $\mathbf{A}^{(k)}$ matrix of order n, which rows are vectors $\vec{a}_1^{(k)}, \ldots, \vec{a}_n^{(k)}$, i.e.

$$\mathbf{A}_n^{(k)} =) \begin{bmatrix} \vec{a}_1^{(k)} \\ \vdots \\ \vec{a}_n^{(k)} \end{bmatrix},$$

Put now $\mathbf{A}^{(1)} = \mathbf{A}$, where **A** is given matrix whose Frobenius form we are looking for.

By direct multiplication it can be shown that using sequence of transformations

$$\begin{aligned} \mathbf{A}^{(2)} &= \mathbf{G}_2(\vec{a}_n^{(1)})\mathbf{A}^{(1)}\mathbf{G}_2^{-1}(\vec{a}_n^{(1)}) \qquad (a_{n,n-1}^{(1)} \neq 0), \\ \mathbf{A}^{(3)} &= \mathbf{G}_3(\vec{a}_{n-1}^{(2)})\mathbf{A}^{(2)}\mathbf{G}_3^{-1}(\vec{a}_{n-1}^{(2)}) \qquad (a_{n-1,n-2}^{(2)} \neq 0), \\ &\vdots \\ \mathbf{A}^{(n)} &= \mathbf{G}_n(\vec{a}_2^{(n-1)})\mathbf{A}^{(n-1)}\mathbf{G}_n^{-1}(\vec{a}_2^{(n-1)}) \qquad (a_{2,1}^{(n-1)} \neq 0), \end{aligned}$$

one gets the Frobenius form $\mathbf{F} = \mathbf{A}^n$. Thereby

$$\mathbf{C} = \mathbf{G}_2^{-1}(\vec{a}_n^{(1)})\mathbf{G}_3^{-1}(\vec{a}_{n-1}^{(2)})\ldots\mathbf{G}_n^{-1}(\vec{a}_2^{(n-1)})$$

and

$$\mathbf{C}^{-1} = \mathbf{G}_n(\vec{a}_2^{(n-1)})\mathbf{G}_{n-1}^{-1}(\vec{a}_3^{(n-2)})\ldots\mathbf{G}_2(\vec{a}_n^{(1)}).$$

Example 5.2.4.

For matrix from previous example, we have in succession

$$\mathbf{G}_2 = \mathbf{G}_2(3,0,1,3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{A}^{(2)} = \mathbf{G}_2\mathbf{A}\mathbf{G}_2^{-1} = \begin{bmatrix} 9 & 2 & -2 & 5 \\ 2 & 3 & -1 & 3 \\ 16 & 4 & 1 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{G}_3 = \mathbf{G}_3(16,4,1,4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 16 & 4 & 1 & 4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -4 & 1/4 & -1/4 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{A}^{(3)} = \mathbf{G}_3\mathbf{A}^{(2)}\mathbf{G}_3^{-1} = \begin{bmatrix} 1 & 1/2 & -5/2 & 3 \\ -24 & 12 & -43 & 48 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{G}_4 = \mathbf{G}_4(-24,12,-43,48) = \begin{bmatrix} -24 & 12 & -43 & 48 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_4^{-1} = \begin{bmatrix} -1/24 & 1/2 & -43/24 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{F} = \mathbf{A}^{(4)} = \mathbf{G}_4\mathbf{A}^{(3)}\mathbf{G}_4^{-1} = \begin{bmatrix} 13 & -67 & 151 & -120 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

5.3. Methods for dominant eigenvalues

Very often, in many applications (i.e. in dynamic of constructions), one needs only maximal (by module) eigenvalue and corresponding eigenvector.

Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues and $\vec{x}_1, \ldots, \vec{x}_n$ corresponding eigenvectors of matrix $\mathbf{A} = [a_{ij}]_{n\times n}$. If

$$|\lambda_1| = \ldots = |\lambda_r| > |\lambda_{r+1}| \geq \ldots \geq |\lambda_n|$$

we say that $\lambda_1, \ldots, \lambda_r$ are dominant eigenvalues of matrix $\mathbf{A}$. In this section we will consider a method for determination of dominant eigenvalue and corresponding eigenvector, as well as some modifications of this method. We suppose that eigenvectors are linearly independent, forming a basis in $\mathcal{R}^n$. Therefore, the arbitrary non-zero vector $\vec{v}_0$ can be expressed as

$$\vec{v}_0 = \sum_{i=1}^{n} \alpha_i \vec{x}_i, \tag{5.3.1}$$

where α_i some scalars. Define now the iterative process

$$\vec{v}_k = \mathbf{A}\vec{v}_{k-1} \qquad (k = 1, 2, \ldots).$$

Then

$$\vec{v}_k = \mathbf{A}\vec{v}_{k-1} = \mathbf{A}^2\vec{v}_{k-2} = \ldots = \mathbf{A}^k\vec{v}_0 = \sum_{i=1}^{n} \alpha_i \mathbf{A}^k \vec{x}_i,$$

or, regarding to (5.3.1) and assertion of Theorem 5.0.2,

$$\vec{v}_k = \sum_{i=1}^{n} \alpha_i \lambda_i^k \vec{x}_i. \tag{5.3.2}$$

The special interesting case here is when one dominant eigenvalue λ_1 $(r = 1)$ exists. Assuming $\alpha_1 \neq 0$, on the basis of (5.3.2) we have

$$\vec{v}_k = \alpha_1 \lambda_1^k (\vec{x}_1 + \sum_{i=1}^{n} \frac{\alpha_i}{\alpha_1} (\frac{\lambda_i}{\lambda_1})^k \vec{x}_i) = \alpha_1 \lambda_1^k (\vec{x}_1 + \vec{\epsilon}_k),$$

where $\vec{\epsilon}_k \to 0$, when $k \to \infty$.

Introduce now notation $(\vec{y})_i$ for i-th coordinate of some vector $\vec{y}$. Then i-th coordinate of vector $\vec{v}_k$ is

$$(\vec{v}_k)_i = \alpha_1 \lambda_1^k \big((\vec{x}_1)_i + (\vec{\epsilon}_k)_i\big).$$

Because of

$$\vec{v}_{k+1} = \alpha_1 \lambda^{k+1} (\vec{x}_1 + \vec{\epsilon}_{k+1}),$$

based on previous, for every i $(1 \le i \le n)$ we have

$$\frac{(\vec{v}_{k+1})_i}{(\vec{v}_k)_i} = \lambda_1 \frac{(\vec{x}_1)_i + (\vec{\epsilon}_{k+1})_i}{(\vec{x}_1)_i + (\vec{\epsilon}_k)_i} \to \lambda_1 \quad (k \to \infty).$$

Based on this fact, the method for determination of dominant eigenvalue λ_1, knows as power method, can be formulated. Vector $\vec{v}_k$ is thereby an approximation of non-normed eigenvector which corresponds to dominant eigenvalue*. By practical realization of this method the norming of eigenvector is performed, i.e. of vector $\vec{v}_k$ after every iteration step. Norm-setting is performed by dividing vector $\vec{v}_k$ by its coordinate with maximal module. So, power method can be expressed by

$$\vec{z}_k = \mathbf{A}\vec{v}_{k-1}, \ \vec{v}_k = \vec{z}_k / \gamma_k,$$

where γ_k coordinate of vector $\vec{z}_k$ with maximal module, i.e. , $\gamma_k = (\vec{z}_k)_i$ and $|(\vec{z}_k)_i| = \|\vec{z}_k\|$. Note that $\gamma_k \to \lambda_1$ and $\vec{v}_k \to \dfrac{\vec{x}_1}{\|\vec{x}_1\|_\infty}$, when $k \to +\infty$.

* If $\vec{x}$ eigenvector, then $c\vec{x}$ $(c \neq 0)$ is also eigenvector corresponding to the same eigenvalue.

Speed of convergence of this method depends on ratio $|\lambda_1/\lambda_2|$. Namely, it holds

$$|\lambda_1 - \gamma_k| = \mathbf{O}(|\frac{\lambda_2}{\lambda_1}|^k). \tag{5.3.3}$$

Note that by deriving of this method we suppose that $\alpha_1 \neq 0$, meaning that method converges if λ_1 is dominant eigenvalue and if initial vector $\vec{v}_0$ has a component with same direction as eigenvector $\vec{x}_1$. On behavior of this method without those assumptions one can find in the monograph of Wilkinson [7, p. 570] and Parlett and Poole [11]. Practically, due to round-off errors in iterative process, the condition $\alpha_1 \neq 0$ will be satisfied after few steps, although starting assumption for vector $\vec{v}_0$ not being fulfilled.

Example 5.3.1.

Let

$$\mathbf{A} = \begin{bmatrix} -261 & 209 & -49 \\ -530 & 422 & -98 \\ -800 & 631 & -144 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 10, \lambda_2 = 4, \lambda_3 = 3$.

By taking for initial vector $\vec{v}_0 = [0 \quad 0 \quad -1]^T$, by power method we get the results given in Table 5.3.1.

Table 5.3.1

k	γ_k	$(\vec{v}_k)_1$	$(\vec{v}_k)_2$	$(\vec{v}_k)_3$
1	144.0000	0.340278	0.680556	1.
2	13.2083	0.334911	0.669821	1.
3	10.7287	0.333774	0.667549	1.
4	10.2038	0.333463	0.666926	1.
5	10.0599	0.333372	0.666744	1.
6	10.0179	0.333345	0.666690	1.
7	10.0054	0.333337	0.666674	1.
8	10.0016	0.333334	0.666669	1.
9	10.0005	0.333334	0.666667	1.
10	10.0001	0.333333	0.666667	1.
11	10.0000	0.333333	0.666667	1.

Because of linear convergence of the power method, for convergence acceleration the Aitken δ^2 method can be used. A simple method for convergence acceleration is given in [1], pp. 303-305.

5.4 Methods for subdominant eigenvalues

Suppose that eigenvalues of matrix $\mathbf{A}$ are ordered in a way

$$|\lambda_1| > |\lambda_2| > \cdots > \lambda_n.$$

In this section the methods for determination of subdominant eigenvalues, i.e. $\lambda_2, \lambda_3, \ldots, \lambda_m$ $(m < n)$ will be considered. The three methods will be explained.

1. Method of orthogonalization. Suppose, at first, that matrix $\mathbf{A}$ is symmetric, and that eigenvector $\mathbf{x_1}$ which corresponds to dominant eigenvalue λ_1 $(\lambda_1| > |\lambda_i|, \ i = 2, \ldots, n)$ has been determined by, for example, power method. Starting with arbitrary vector $\vec{z}$, let us form vector $\vec{v}_0$ which is orthogonal to vector $\vec{x}_1$. So we have (see Gram-Schmidt's method of orthogonalization)

$$\vec{v}_0 = \vec{z} - \frac{(\vec{z}, \vec{x}_1)}{(\vec{x}_1, \vec{x}_1)} \vec{x}_1. \tag{5.4.1}$$

Because of $(\vec{v}_0, \vec{x}_1) = 0$, from theoretical point of view, the series $\vec{v}_k = \mathbf{A}\vec{v}_{k-1}$ $(k = 1, 2, \ldots)$ in the power method could be used for determination of λ_2 and corresponding eigenvector $\vec{x}_2$. Nevertheless, regardless of fact that $\vec{v}_0$ does not have the component in direction of eigenvector $\vec{x}_1$ power method would, because of round-off errors, after some number of iterations, converge toward eigenvector $\vec{x}_1$. This fact was mentioned in the previous section.

It is possible to eliminate this influence of round-off errors using so known periodical "purification" of vector v_0 from component in direction of $\vec{x}_1$. That means that, after, say, r steps, we compute $\vec{v}_0$ using $\vec{v}_r$ in spite of $\vec{z}$ in (5.4.1), i.e. by means

$$\vec{v}_0 = \vec{v}_r - \frac{(\vec{v}_r, \vec{x}_1)}{(\vec{x}_1, \vec{x}_1)}\vec{x}_1.$$

In this way, if the period of "purification" is small enough so that it cannot happen significant accumulation of round-off error, by power method can be determined eigenvalue λ_2 and eigenvector $\vec{x}_2$.

By continuation of this procedure we can further determine λ_3 and $\vec{x}_3$.

Generally, if we determine $\lambda_1, \ldots, \lambda_\nu$ and corresponding vectors $\vec{x}_1, \ldots, \vec{x}_\nu$ $(\nu < m)$ it is possible to determine $\lambda_{\nu+1}$ and $\vec{x}_{\nu+1}$ by power method, by forming vector $\vec{v}_0$ orthogonal to $\vec{x}_1, \ldots, \vec{x}_\nu$. So, starting from arbitrary vector $\vec{z}$, we have

$$\vec{v}_0 = \vec{z} - \sum_{i=1}^{\nu} \frac{(\vec{z}, \vec{x}_i)}{(\vec{x}_i, \vec{x}_i)}\vec{x}_i, \tag{5.4.2}$$

meaning that vector $\vec{v}_0$ has components only in direction of residual eigenvectors, i.e.

$$\vec{v}_0 = \alpha_{\nu+1}\vec{x}_{\nu+1} + \ldots + \alpha_n\vec{x}_n.$$

Power method applied to $\vec{v}_0$ gives $\vec{x}_{\nu+1}$ and $\lambda_{\nu+1}$ in absence of round-off errors. Being not the case, it is necessary frequent "purification" of vector $\vec{v}_k$ from components in direction $\vec{x}_1, \ldots, \vec{x}_n$. In other words, after r steps, one should determine again $\vec{v}_0$ using (5.4.2), by using $\vec{v}_r$ in spite of $\vec{z}$.

Also, in the case when matrix $\mathbf{A}$ is not symmetric, but has complete system of eigenvectors, the given orthogonalizing procedure can be applied.

2. Inverse iteration method. This method is applied to general matrix $\mathbf{A}$ and is based on solution of system of equations

$$(\mathbf{A} - p\mathbf{I})\vec{v}_k = \vec{v}_{k-1}, \tag{5.4.3}$$

where p is constant, and $\vec{v}_0$ arbitrary vector. System (5.4.3) is usually to be solved by Gauss method of elimination or Cholesky method by factorization of matrix $\mathbf{B} = \mathbf{A} - p\mathbf{I}$. Note that the method of inverse iteration is equivalent to the power method applied to $\mathbf{B}$. Therefore, by applying method of inverse iteration the dominant eigenvalue of matrix $\mathbf{B}$is obtained, i.e. $\mu_\nu = 1/(\lambda_\nu - p)$ for which it holds

$$\min_j |\lambda_j - p| = |\lambda_\nu - p|.$$

Eigenvalue λ_ν is closest eigenvalue of matrix $\mathbf{A}$ to the number p. Eigenvector obtained thereby is the same one for matrices $\mathbf{B}$ and $\mathbf{A}$.

By convenient choice of parameter p all eigenvalues of matrix $\mathbf{A}$ can be, in principle, obtained.

Similar to power method, here is also suitable to perform norming of vector $\vec{v}_k$ so that we have

$$\mathbf{B}\vec{z}_k = \vec{v}_{k-1}, \qquad \vec{v}_k = \vec{z}_k/\gamma_k, \tag{5.4.4}$$

where γ_k is coordinate of vector $\vec{z}_k$ with greatest module.

Example 5.4.1.

Using method of inverse iteration for matrix

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 10 & 1 \\ 4 & 1 & 10 \end{bmatrix},$$

we will determine eigenvalue closest to number $p = 9$, as well as corresponding eigenvector.

Using factorization by Gauss method with pivoting for matrix $\mathbf{B} = \mathbf{A} - 9\mathbf{I}$, we get

$$\mathbf{LR} = \mathbf{PB}$$

where

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -4/5 & 1 & 0 \\ -1/5 & 2/3 & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} -5 & 1 & 4 \\ 0 & 9/5 & 21/5 \\ 0 & 0 & -1 \end{bmatrix}$$

and permutation matrix $\mathbf{P}$ defined by index series $\mathbf{I} = (1, 3)$.

The method of inverse iteration $(5.4.4)$ can now be expressed in the form

$$\mathbf{L}\vec{y}_k = p\vec{v}_{k-1}, \quad \mathbf{R}\vec{z}_k = \vec{y}_k, \quad \vec{v}_k = \vec{z}_k/\gamma_k,$$

by which application the results given in Table 5.4.1. are obtained.

Table 5.4.1

k	$(\vec{v}_k)_1$	$(\vec{v}_k)_2$	$(\vec{v}_k)_3$	β_k
1	0.	1.	−1.	6.
2	−0.2	1.	−0.5	9.3
3	−0.17241	1.	−0.48276	9.34483
4	−0.17200	1.	−0.48000	9.34800
5	−0.17185	1.	−0.47980	9.34835
6	−0.17184	1.	−0.47977	9.34838

For initial vector we took $\vec{v}_0 = [1 \quad 0 \quad 0]^T$. In the last column of table is given the quantity $\beta_k = p + 1/\gamma_k$, which gives approximation for corresponding eigenvalue λ. One can see that this eigenvalue has approximate value 9.34838.

3. Deflation method. The methods of this kind are composed from construction of sequence of matrices $\mathbf{A}_n \ (= \mathbf{A}),\ \mathbf{A}_{n-1}, \ldots, \mathbf{A}_1$, which order is equal to index and thereby

$$Sp(\mathbf{A}_n) \supset Sp(\mathbf{A}_{n-1}) \supset \cdots \supset Sp(\mathbf{A}_1),$$

where $Sp(\mathbf{A}_k)$ denotes spectrum of matrix $\mathbf{A}_k$.

We will describe now a special and important case of deflation method, when matrix $\mathbf{A}$ is Hermitian.

Let $\vec{x} = [x_1 \quad x_2 \ldots x_n]^T$ be eigenvector of matrix $\mathbf{A}$ corresponding to eigenvalue λ and normed

$$(\vec{x}, \vec{x}) = \vec{x} * \vec{x} = \|\vec{x}\|_E^2 = 1$$

with first coordinate x_1 being nonnegative.

Have a look over matrix

$$\mathbf{P} = \mathbf{I} - 2\vec{w}\vec{w}^*, \tag{5.4.5}$$

where the vector $\vec{w} = [w_1\ w_2\ \ldots w_n]^T$ is defined by first vector $\vec{e}_1 = [1\ 0\ \ldots\ 0]^T$ from natural basis space $\mathcal{R}^n$ in the following way:

$$\vec{w} * \vec{w} = \|\vec{x}\|_E^2 = 1 \quad w_1 \geq 0, \tag{5.4.6}$$

$$\mathbf{P}\vec{e}_1 = \vec{x}. \tag{5.4.7}$$

The matrix $\mathbf{P}$ is of form

$$\mathbf{P} = \begin{bmatrix} 1 - 2w_1\bar{w}_1 & -2w_1\bar{w}_2 & \ldots & -2w_1\bar{w}_n \\ -2w_2\bar{w}_1 & 1 - 2w_2\bar{w}_2 & & -2w_2\bar{w}_n \\ \vdots & & & \\ -2w_n\bar{w}_1 & -2w_n\bar{w}_2 & & 1 - 2w_n\bar{w}_n \end{bmatrix}.$$

Note that $\mathbf{P}^* = \mathbf{P}$, what means that matrix $\mathbf{P}$ is Hermitian, too. Moreover, regarding to (5.4.6), by direct multiplication we see that

$$\mathbf{P}^*\mathbf{P} = \mathbf{P}^2 = \mathbf{I},$$

and conclude that matrix $\mathbf{P}$ is unitary.

Based on (5.4.7) we find coordinates of vector w. So, from $1 - 2w\bar{w}_1 = x_1$ and $-2w_k\bar{x}_k \quad (k = 2, \ldots, n)$ it follows

$$w_1 = \sqrt{\frac{1 - x_1}{2}} \quad \text{and} \quad w_k = -\frac{x_k}{2w_1} \quad (k = 2, \ldots, n).$$

Note that $\bar{w}_1 = w_1 > 0$.

Now, based on (5.4.7) and $\mathbf{A}\vec{x} = \lambda\vec{x}$ we find that $\mathbf{AP}\vec{e}_1 = \mathbf{P}\vec{e}_1$, wherefrom we conclude that $\mathbf{P}^*\mathbf{AP}\vec{e}_1 = \lambda\vec{e}_1$, i.e. $\vec{e}_1$ is eigenvector of matrix $\mathbf{B} = \mathbf{P}^*\mathbf{AP} = \mathbf{PAP}$. Note, also, that first column in matrix $\mathbf{B}$ is just vector $\lambda\vec{e}_1$, i.e.

$$\mathbf{B} = \begin{bmatrix} \lambda & b_{12} & b_{13} & \cdots & b_{1n} \\ 0 & b_{22} & b_{23} & & b_{2n} \\ 0 & b_{32} & b_{33} & & b_{3n} \\ \vdots & & & & \\ 0 & b_{n2} & b_{n3} & & b_{nn} \end{bmatrix} = \begin{bmatrix} \lambda & \vec{b}_{n-1}^T \\ \vec{0}_{n-1} & \mathbf{A}_{n-1} \end{bmatrix},$$

where with $\mathbf{A}_{n-1}$ we denoted matrix of order $n-1$ which matches with enclosed block, $\vec{0}_{n-1}$ is zero-vector of order $n-1$, and, finally, $\vec{b}_{n-1}^T = [b_{12}\ b_{13} \cdots b_{1n}]^T$.

Regarding the fact that matrix $\mathbf{B}$ is similar (we say also unitary similar) to matrix $\mathbf{A}$, we conclude that

$$Sp(\mathbf{A}_{n-1}) = Sp(\mathbf{A}_n) \backslash (\lambda) \qquad (\mathbf{A}_n = \mathbf{A}).$$

In order to get matrix $\mathbf{A}_{n-2}$ we are proceeding in a similar way. In spite of matrix $\mathbf{P}$ we use matrix

$$\mathbf{P}_1 = \begin{bmatrix} 1 & \vec{0}_{n-1}^T \\ \vec{0}_{n-1} & \mathbf{Q} \end{bmatrix},$$

where matrix $\mathbf{Q}$ is of order $n-1$and of form (5.4.5), satisfying the conditions (5.4.6) and (5.4.7) regarding eigenvector $\vec{y}$ and eigenvalue ν of matrix $\mathbf{A}_{n-1}$. Because of $\mathbf{P}_1^{-1} = \mathbf{P}_1^* = \mathbf{P}_1$ we conclude that matrix $\mathbf{P}_1$ is unitary, too.

Now matrix $\mathbf{C} = \mathbf{P}_1\mathbf{BP}_1 = \mathbf{P}_1\mathbf{PAPP}_1$ has a form

$$\mathbf{P}_1 = \begin{bmatrix} \lambda & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & \mu & c_{23} & & c_{2n} \\ 0 & 0 & c_{33} & & c_{3n} \\ \vdots & & & & \\ 0 & 0 & c_{n3} & & c_{nn} \end{bmatrix} = \begin{bmatrix} \lambda & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & \mu & c_{23} & & c_{2n} \\ 0 & 0 & & & \\ \vdots & & & \mathbf{A}_{n-2} & \\ 0 & 0 & & & \end{bmatrix},$$

where matrix $\mathbf{A}_{n-2}$ is of order $n-2$. By continuing this procedure we get upper triangular matrix which is unitary similar to initial matrix $\mathbf{A}$. Having in mind that matrix $\mathbf{A}$ is Hermitian, we conclude that it is unitary similar to diagonal matrix.

The presented procedure demands before of every step determination of one eigenvalue and corresponding eigenvector, what can be done by some of previously presented methods. Thus, before the first step, one has to determine eigenvalue λ and eigenvector $\vec{x}$ of matrix $\mathbf{A}$, before the second step eigenvalue μ and eigenvector $\vec{y}$ of matrix of $n-1$ order $\mathbf{A}_{n-1}$, and so on.

It is clear that eigenvalues of matrix $\mathbf{A}$ are diagonal elements of obtained triangular matrix, i.e. $\lambda_1 = \lambda$, $\lambda_2 = \mu$, etc. It remains the question what is with eigenvectors of matrix $\mathbf{A}$? It is clear that $\vec{x}_1 = \vec{x}$. We will show how, based on obtained results, the second eigenvector of matrix $\mathbf{A}$ can be found.

If the coordinates of eigenvector $\vec{y}$ are $y_2, \ldots, y_n$, in order to find, at first, eigenvector $\vec{y}\,'$ of matrix $\mathbf{B}$, put $\vec{y}\,' = [y_1\ y_2\ \ldots y_n]^T$ and try to determine y_1.

Because of

$$\mathbf{B}\vec{y}\,' = \begin{bmatrix} \lambda & \vec{b}_{n-1}^T \\ \vec{\mathbf{0}}_{n-1} & \mathbf{A}_{n-1} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vec{y} \end{bmatrix} = \begin{bmatrix} \lambda y_1 + \vec{b}_{n-1}^T \vec{y} \\ \mathbf{A}_{n-1}\vec{y} \end{bmatrix},$$

i.e.

$$\mathbf{B}\vec{y}\,' = \begin{bmatrix} \lambda y_1 + \vec{b}_{n-1}^T \vec{y} \\ \mu \vec{y} \end{bmatrix},$$

it follows

$$y_1 + \vec{b}_{n-1}^T \vec{y} = y_1.$$

If $\lambda \neq \mu$, by virtue of previous equality we get

$$y_1 = \frac{1}{\lambda - \mu}\vec{b}_{n-1}^T \vec{y} = \frac{1}{\lambda - \mu}(b_{12}y_2 + \cdots + b_{1n}y_n).$$

Now simply find the eigenvector $\vec{x}_2$ of matrix $\mathbf{A}$, corresponding to eigenvalue $\lambda_2 \neq \mu$. Indeed, because of $\mathbf{PAP}\vec{y}\,' = \mu\vec{y}\,'$, i.e. $\mathbf{A}(\mathbf{P}\vec{y}\,') = \mu(\mathbf{P}\vec{y}\,')$ we conclude that $\vec{x}_2 = \mathbf{P}\vec{y}\,'$.

In a similar way the other eigenvectors can be determined.

5.5. Jacobi method

In this section we begin presentation of method for solving a complete problem of eigenvalues, i.e. for determination of all eigenvalues and corresponding eigenvectors. The method presented in this section originates from C.G.J. Jacobi (1846), and can be applied on Hermitian matrices. Jacobi method is based on transformation of Hermitian matrix $\mathbf{A}$ to diagonal matrix $\mathbf{D}$, which eigenvalues are diagonal elements. It is known that exists unitary matrix $\mathbf{H}$ such that

$$\mathbf{H}^*\mathbf{A}\mathbf{H} = \mathbf{D}. \tag{5.5.1}$$

Because of $\mathbf{H}^* = \mathbf{H}^{-1}$ in virtue of (5.5.1) matrices $\mathbf{A}$ and $\mathbf{D}$ are similar and have same eigenvalues.

Consider first the transformation

$$\mathbf{B} = \mathbf{R}^*\mathbf{A}\mathbf{R}, \tag{5.5.2}$$

where $\mathbf{R}$ is arbitrary unitary matrix.

Theorem 5.5.1. *Matrices* **B** *and* **A** *from* (5.5.2) *have same Schmidt norms, i.e.* $\epsilon(\mathbf{B}) = \epsilon(\mathbf{A})$.

Proof.

Matrices $\mathbf{A}$ and $\mathbf{B}$ are similar. Let λ_i $(i = 1, 2, \ldots, n)$ be their eigenvalues. Because of

$$\epsilon(\mathbf{A})^2 = \sum_{i,j\neq 1}^{n} |a_{ij}|^2 = tr(\mathbf{A}^*\mathbf{A}) = tr(\mathbf{A}^2) = \sum_{i=1}^{n} \lambda_i^2,$$

we conclude that $\epsilon(\mathbf{B}) = \epsilon(\mathbf{A})$

The basic idea in Jacobi method is construction of series of similar matrices $\{\mathbf{A}_k\}_{k\in N}$, starting from $\mathbf{A}_1 = \mathbf{A}$, with strategy of minimization of value $od(\mathbf{A}_k)$ where with od is denoted "norm" of elements of matrix $\mathbf{A}$ out of diagonal, i.e.

$$od(\mathbf{A}) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|^2.$$

Note that $od(\mathbf{A}) + d(\mathbf{A}) = \epsilon(\mathbf{A})$, where $d(\mathbf{A}) = \sum\limits_{i=1}^{n} |a_{ij}|^2$. Strategy of minimization should provide

$$od(\mathbf{A}_{k+1}) \leq od(\mathbf{A}_k) \quad \text{and} \quad \lim_{k\to+\infty} od(\mathbf{A}_k) = 0.$$

It remains the question how to generate series $\mathbf{A}_k$, i.e. which unitary matrices should be used for transformation of $\mathbf{A}_k$ to $\mathbf{A}_{k+1}$.

Look at so known rotation matrix $R = R(p, q)$ with elements

$$\begin{aligned}
&r_{pp} = e^{i\alpha}\cos\theta, && r_{pq} = e^{i\beta}\sin\theta,\\
&r_{qp} = -e^{-i\beta}\sin\theta, && r_{qq} = e^{-i\alpha}\cos\theta,\\
&r_{ij} = \delta_{ij} \ \text{(in other cases)},
\end{aligned}$$

where θ, α, β are real numbers. By direct multiplication we show that $\mathbf{R}^*\mathbf{R} = \mathbf{I}$, i.e. the rotation matrix is unitary. We call usually this matrix elementary matrix of rotation with angle θ in the plane (p, q).

In spite of being applicable for Hermitian matrices, due to simplicity we will present Jacobi method for the case when $\mathbf{A}$ is real symmetric matrix. In favor of this is the fact that problem of eigenvalues of Hermitian matrix $\mathbf{A}$ of order n can reduce to problem of eigenvalue of symmetric matrix of order $2n$. Namely, matrix $\mathbf{A}$ can be decomposed to two real matrices $\mathbf{S}$ and $\mathbf{K}$ in form $\mathbf{A} = \mathbf{S} + i\mathbf{K}$, where $\mathbf{S}$ is symmetric, and $\mathbf{K}$ angular-symmetric matrix. Eigenvalue λ and eigenvector $\vec{v} = \vec{x} + i\vec{y}$ of matrix $\mathbf{A}$ satisfy equality

$$\begin{bmatrix} \mathbf{S} & -\mathbf{K} \\ \mathbf{K} & \mathbf{S} \end{bmatrix} \cdot \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \lambda \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix},$$

whereby if λ is simple eigenvalue of matrix $\mathbf{A}$ then it is double for matrix of order $2n$ which appears in (5.5.3).

From previously noted reasons we suppose that $\mathbf{A}$ is real symmetric matrix. In rotation matrix we can take $\alpha = \beta = 0$, so that this matrix is now orthogonal. Therefore, we have

$$\begin{aligned}
&r_{pp} = \cos\theta = c, && r_{pq} = \sin\theta = s,\\
&r_{pq} = -\sin\theta = -s, && r_{qq} = \cos\theta = c,\\
&r_{ij} = \delta_{ij} \ \text{(in other cases)}.
\end{aligned}$$

Note that elements of matrix $\mathbf{B}$ in transformation

$$\mathbf{B} = \mathbf{R}^T\mathbf{A}\mathbf{R} \tag{5.5.4}$$

coincide with corresponding elements of matrix $\mathbf{A}$, beside those located in $p-$th and $q-$th row and column. Moreover, we have

$$\begin{bmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix}, \tag{5.5.5}$$

wherefrom, regarding to the statement of theorem 5.5.1, we find

$$b_{pp}^2 + 2b_{pq}^2 + b_{qq}^2 = a_{pp}^2 + 2a_{pq}^2 + a_{qq}^2.$$

Since, again according to theorem 5.5.1, $\epsilon(\mathbf{A}) = \epsilon(\mathbf{B})$ we have

$$od(\mathbf{B}) = \epsilon(\mathbf{B}) - d(\mathbf{B}) = \epsilon(\mathbf{A})\, d(\mathbf{B}),$$

i.e.

$$\begin{aligned} od(\mathbf{B}) &= od(\mathbf{A}) + d(\mathbf{A}) - d(\mathbf{B}), \\ &= od(\mathbf{A}) + a_{pp}^2 + a_{qq}^2 - (b_{pp}^2 + b_{qq}^2) \\ &= od(\mathbf{A}) + 2b_{pq}^2 - 2a_{pq}^2, \end{aligned}$$

wherefrom we see that $od(\mathbf{B})$ will be minimized if $b_{pq} = 0$. Using this condition we will derive $c = \cos\theta$ and $s = \sin\theta$ in rotation matrix.

Since, based on (5.5.5),

$$b_{pq} = (c^2 - s^2)a_{pq} + cs(a_{pp} - a_{qq}),$$

we derive

$$(\cos^2\theta - \sin^2\theta)a_{pq} + \cos\theta\sin\theta(a_{pp} - a_{qq}) = 0,$$

i.e.

$$\tau = \cot(2\theta) = \frac{a_{qq} - a_{pp}}{2a_{pq}} \qquad (a_{pq} \neq 0).$$

If $a_{pq} = 0$, we have $c = 1$ and $s = 0$, what means that R is unit matrix. Put $\tan\theta = t$. Because of $\cot(2\theta) = (1-t^2)/2t = \tau$, for determination of t the following quadratic equation should be solved

$$t^2 + 2\tau t - 1 = 0.$$

Usually one takes the solution with lower absolute value, i.e.

$$t = \frac{sgn\tau}{|\tau| + \sqrt{\tau^2 + 1}},$$

what ensures the condition that rotation angle is $|\theta| \leq \pi/4$. Then, the elements of rotation matrix are

$$c = \frac{1}{\sqrt{1 + t^2}} \qquad \text{and} \qquad s = tc.$$

With in this way determined matrix of rotation R, "norm" of off-diagonal elements of matrix $\mathbf{B}$ in (5.5.4) is minimized, whereby element on the position (p, q) is zeroed. This procedure is repeated, i.e. the series of similar matrices is generated,

$$\mathbf{A}_{k+1} = \mathbf{R}_k^T \mathbf{A}_k \mathbf{R}_k \qquad (k = 1, 2, \ldots), \tag{5.5.6}$$

starting with $\mathbf{A}_1 = \mathbf{A}$, where remains the open question how to chose rotation matrix $\mathbf{R}_k = \mathbf{R}_k(p, q)$, i.e. pairs (p, q) on every step in (5.5.6). At classical Jacobi method for (p, q) is taken position of dominant element, i.e. off-diagonal element of matrix $\mathbf{A}_k = [a_{ij}^{(k)}]$ with greater absolute value. Thus, (p, q) is to be determined from the condition

$$|a_{pq}^{(k)}| = \max_{i<j} |a_{ij}^{(k)}|.$$

The elements of matrix $\mathbf{A}_k$ which change during transformation (5.5.6) are only those located in p-th and q-th row and column:

$$\left.\begin{cases} a_{ip}^{(k+1)} = a_{pi}^{(k+1)} = ca_{ip}^{(k)} - sa_{iq}^{(k)} \\ a_{iq}^{(k+1)} = a_{qi}^{(k+1)} = sa_{ip}^{(k)} + ca_{iq}^{(k)} \end{cases}\right\} \quad (i \neq p, q),$$

$$\begin{aligned} a_{pp}^{(k+1)} &= c^2 a_{pp}^{(k)} - 2csa_{pq}^{(k)} + s^2 a_{qq}^{(k)}, \\ a_{qq}^{(k+1)} &= s^2 a_{pp}^{(k)} + 2csa_{pq}^{(k)} + c^2 a_{qq}^{(k)}, . \\ a_{pq}^{(k+1)} &= a_{qp}^{(k+1)} = 0 \end{aligned}$$

For such constructed method we have $\mathbf{A}_k \to \mathbf{D} = [\lambda_i \delta_{ij}]$, when $k \to +\infty$. Indeed, regarding to inequality

$$2(a_{pq}^{(k)})^2 \geq \frac{1}{N} od(\mathbf{A}_k) \qquad (N = \frac{1}{2} n(n-1)),$$

we have

$$od(\mathbf{A}_{k+1}) = od(\mathbf{A}_k) - 2(a_{pq}^{(k)})^2 \leq (1 - \frac{1}{N}) od(\mathbf{A}_k),$$

i.e.

$$od(\mathbf{A}_k) \leq (1 - \frac{1}{N})^{k-1} od(\mathbf{A}),$$

wherefrom it follows $\lim\limits_{k \to +\infty} od(\mathbf{A}_k) = 0$, i.e. $\lim\limits_{k \to +\infty} od(\mathbf{A}_k) = \mathbf{D}$. Convergence of Jacobi method is quadratic in a sense that there exists constant M (> 0) such that for enough large k

$$od(\mathbf{A}_{k+N}) \leq M(od(\mathbf{A}_k))^2.$$

By virtue of (5.5.6) we can write

$$\mathbf{A}_{k+1} = \mathbf{H}_k^T \mathbf{A}_k \mathbf{H}_k, \tag{5.5.7}$$

where $\mathbf{H}_k = \mathbf{R}_1 \mathbf{R}_2 \ldots \mathbf{R}_k$. For k large enough, matrix $\mathbf{H}_k$ can be treated as good enough approximation for unitary matrix $\mathbf{H}$ which appears in (5.5.1). Jacobi iterative process (5.5.6) usually stops when $od\mathbf{A}_{k+1} \leq \delta^2$ where δ is exactness given in advance.

In classical Jacobi method good deal of time could be spent for determination of position of dominant element. One modification of classic Jacobi method, known as cyclic Jacobi method , takes for (p, q) successively the pairs:

$$(1,2), (1,3), \ldots, (1,n); (2,3), \ldots, (2,n); \ldots, (n-1,n); (1,2), \ldots$$

It can be shown that this method has also quadrat convergence.

If k is large enough in (5.5.7), so that criteria for process termination is fulfilled, we can consider that

$$\mathbf{H}_k^T \mathbf{A}_k \mathbf{H}_k = \mathbf{A}_{k+1} = D = [\lambda_i \delta_{ij}]_{n \times n},$$

wherefrom it follows

$$\mathbf{A}\mathbf{H}_k = \mathbf{H}_k D,$$

what means that columns of matrix $\mathbf{H}_k$ are eigenvectors of matrix $\mathbf{A}$.

Moreover, these eigenvectors form orthogonal set. Matrix $\mathbf{H}_k = [h_{ij}^{(k)}]_{n \times n}$ can be generated recursively using rotation matrix

$$\mathbf{H}_k = \mathbf{H}_{k-1} \mathbf{R}_k(p, q) \qquad (\mathbf{H}_0 = \mathbf{I}).$$

In scalar form we have

$$\left.\begin{cases} h_{ip}^{(k+1)} = c\, h_{ip}^{(k)} - s\, h_{iq}^{(k)} \\ h_{iq}^{(k+1)} = s\, h_{ip}^{(k)} + c\, h_{iq}^{(k)} \end{cases}\right\} \quad (i = 1, 2, \ldots, n),$$

$$h_{ij}^{(k+1)} = h_{ij}^{(k)} \quad \text{in} \quad \text{other} \quad \text{cases}.$$

Example 5.5.1.

We will apply classic Jacobi method on matrix **A** from example (5.4.1). Put

$$\mathbf{A}_1 = \mathbf{A} = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 10 & 1 \\ 4 & 1 & 10 \end{bmatrix}.$$

Dominant element is on position $(p, q) = (1, 3)$. Then

$$\tau = \frac{a_{33} - a_{11}}{2a_{13}} = \frac{3}{4}, \; t = \frac{1}{2}, \; c = c_1 = \frac{2}{\sqrt{5}}, \; s = s_1 = \frac{1}{\sqrt{5}},$$

so that

$$\mathbf{R}_1 = \mathbf{R}_1(1,3) = \begin{bmatrix} 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \\ -1/\sqrt{5} & 0 & 2/\sqrt{5} \end{bmatrix},$$

$$\mathbf{H}_1 = \mathbf{R}_1, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 1/\sqrt{5} & 0 \\ 1/\sqrt{5} & 10 & 3/\sqrt{5} \\ 0 & 3/\sqrt{5} & 12 \end{bmatrix}.$$

By continuing this procedure we get the following results:

a) For elementary rotation matrices:

Table 5.5.1

k	(p, q)	c_k	s_k
2	$(2, 3)$	0.89376	0.44855
3	$(1, 2)$	0.99852	0.05431
4	$(1, 3)$	0.99982	0.01872

b) For series $\mathbf{H}_k$ and $\mathbf{A}_k$ (matrices $\mathbf{A}_k$ are symmetric; elements of lower triangle are not given):

$$\mathbf{H}_2 = \begin{bmatrix} 0.89443 & -0.20060 & 0.39970 \\ 0. & 0.89376 & 0.44855 \\ -0.44721 & -0.40119 & 0.79940 \end{bmatrix},$$

$$\mathbf{A}_3 = \begin{bmatrix} 2.00000 & 0.39970 & 0.20060 \\ & 9.32668 & 0.00000 \\ & & 12.67332 \end{bmatrix},$$

$$\mathbf{H}_3 = \begin{bmatrix} 0.90400 & -0.15172 & 0.39970 \\ -0.04854 & 0.89244 & 0.44855 \\ -0.42476 & -0.42489 & 0.79940 \end{bmatrix},$$

$$\mathbf{A}_4 = \begin{bmatrix} 1.97826 & 0.00000 & 0.20030 \\ & 9.34842 & 0.01089 \\ & & 12.67332 \end{bmatrix},$$

$$\mathbf{H}_4 = \begin{bmatrix} 0.89636 & -0.15172 & 0.41655 \\ -0.05693 & 0.89244 & 0.44756 \\ -0.43965 & -0.42489 & 0.79131 \end{bmatrix},$$

$$\mathbf{A}_5 = \begin{bmatrix} 1.97451 & -0.00020 & 0.00000 \\ & 9.34842 & 0.01089 \\ & & 12.67707 \end{bmatrix}.$$

All results are rounded to five decimals. From $\mathbf{A}_5$ and $\mathbf{H}_4$ we have eigenvalues

$$\lambda_1 \cong 1.97451, \quad \lambda_2 \cong 9.34842, \quad \lambda_3 \cong 12.67707,$$

and corresponding eigenvectors

$$\vec{x}_1 = \begin{bmatrix} 0.89636 \\ -0.05693 \\ -0.43965 \end{bmatrix}, \ \vec{x}_2 = \begin{bmatrix} -0.15172 \\ 0.89244 \\ -0.42489 \end{bmatrix}, \ \vec{x}_3 = \begin{bmatrix} 0.41655 \\ 0.44756 \\ 0.79131 \end{bmatrix}.$$

Note that in example 5.4.1 we determined by inverse matrix method eigenvalue λ_2 and corresponding eigenvector $\vec{x}_2$ (normed regarding to coordinate with maximal absolute value).

Remark 5.5.1.

When solution of complete eigenvalue problem is not demanded, it is not necessary to generate matrix $\mathbf{H}_k$. For example, if needed only one eigenvector, say $\vec{x}_m$, then it can be simply obtained by applying rotation on vector $\vec{e}_m$, which $m-$th coordinate equals to one, and all other are equal to zero.

5.6. Givens' and Householder's method

As we have seen in previous section, Jacobi method transforms symmetric (or, generally Hermitian) matrix to the diagonal one after infinity number of steps. Off-diagonal elements (a_{pq} and a_{qp}) which are zeroed on the distinct step of of Jacobi method, can in later stages become such that notable deviate from zero, what is especially expressed when matrix order is high. This fact slows algorithm. In this section we will present two methods at which this disadvantage is eliminated. First of these methods is developed by W.J. Givens (1954), and second by A.S. Householder (1958). The methods are such that by definitive number of steps transform initial real symmetric matrix to symmetric, tridiagonal matrix. In the next section we will consider problem of eigenvalues for symmetric tridiagonal matrix. Both methods can be simply applied to Hermitian matrices. Moreover, their application to general matrices leads to matrix reduction to so known Hessenberg form (a_{ij}=0 for $i \geq j+2$).

As in previous section, we will present noted methods for the case of matrix $\mathbf{A}$ being real and symmetric.

1. Givens' method. This reduction method is based on successive application of so known Givens transformation, whereby after finite number of rotations matrix $\mathbf{A}$ reduces to tridiagonal matrix.

we will call elements of matrix $\mathbf{A} = [a_{ij}]_{n\times n}$ for which indices it holds $|i-j| > 1$ off-tri-diagonal elements. We will introduce for them such arrangement that we denote a_{ij} as s-th off-tri-diagonal element if the pair (i, j) is s-th element of cycling index sequence

$$(1,3),(1,4),\ldots,(1,n),(2,4),(2,5),\ldots,(2,n),\ldots,(n-2,n).$$

Note that this sequence contains totally

$$M = (n-2) + (n-1) + \cdots + 1 = \frac{1}{2}(n-1)(n-2)$$

pairs.

Let $\mathbf{A}_1 = \mathbf{A}$ and

$$\mathbf{A}_{k+1} = \mathbf{G}_k^T \mathbf{A}_k \mathbf{G}_k, \qquad (k = 1, 2, \ldots, M) \tag{5.6.1}$$

where matrices $\mathbf{G}_k$, so known two-dimensional rotation, are chosen so that first k off-tri-diagonal elements of matrix $\mathbf{A}_{k+1}$ are equal to zero.

Theorem 5.6.1. *(Givens) Let* $\mathbf{A}$ *be real symmetric matrix, series* $\mathbf{A}_k = [a_{ij}^{(k)}]_{n\times n}$ *defined by* (5.6.1) *and* $(p-1,q)$ *k-th pair of cyclic indexed series of off-tri-diagonal elements. If rotation matrix* $\mathbf{G}_k = [g_{ij}^{(k)}]_{n\times n}$ *is defined as*

$$\begin{aligned} &g_{pp}^{(k)} = g_{qq}^{(k)} = c, \;\; g_{pq}^{(k)} = -g_{pq}^{(k)} = -s, \\ &g_{ij}^{(k)} = \delta_{ij} \quad \textit{(in other cases)}, \end{aligned}$$

where

$$\begin{aligned} &c = \cos\theta = \frac{1}{s}a_{p-1,p}^{(k)}, \;\; s = \sin\theta = \frac{1}{s}a_{p-1,q}^{(k)}, \\ &s = \sqrt{(a_{p-1,p}^{(k)})^2 + (a_{p-1,q}^{(k)})^2}, \end{aligned}$$

or using $\mathbf{G}_k = \mathbf{I}$, *if* $a_{p-1,q}^{(k)} = 0$, *we have*

(a) Matrices $\mathbf{A}_{k+1}$ *are real and symmetric;*

(b) First k *off-tri-diagonal elements of matrix* $\mathbf{A}_{k+1}$ *are equal to zero* $(k = 1, 2, \ldots, M)$;

(c) Matrix $\mathbf{A}_{M+1}$ *is tri-diagonal.*

The proof can be carried out by mathematical induction. Note the basic distinction between Givens and Jacobi method. Namely, at Jacobi method, with rotation matrix $\mathbf{R}_k = \mathbf{R}_k(p,q)$ is nullified only element on position (p,q). At Givens method we have that in matrix $\mathbf{A}_k$ first $k-1$ off-tri-diagonal elements equal to zero. By Givens rotation $\mathbf{G}_k = \mathbf{G}_k(p,q)$, corresponding elements in matrix $\mathbf{A}_k$ remain unchanged, i.e. equal to zero. But, because of

$$a_{p-1,q}^{(k+1)} = -a_{p-1,p}^{(k)}\sin\theta + a_{p-1,q}^{(k)}\cos\theta,$$

with choice

$$\tan\theta = \frac{a_{p-1,q}^{(k)}}{a_{p-1,p}^{(k)}}$$

we have also k-th element being equal to zero, i.e. $a_{p-1,q}^{(k+1)} = 0$. Exactly this choice of $\tan\theta$ is taken for determination of elements c and s in rotation matrix $\mathbf{G}_k$.

Based on (5.6.1) we have

$$\begin{aligned} &a_{pp}^{(k+1)} = c^2 a_{pp}^{(k)} + 2\,c\,s\,a_{pq}^{(k)} + s^2 a_{qq}^{(k)}, \\ &a_{pq}^{(k+1)} = a_{qp}^{(k+1)} = (c^2 - s^2)a_{pq}^{(k)} + c\,s\,(a_{qq}^{(k)} - a_{pp}^{(k)}),. \\ &a_{qq}^{(k+1)} = s^2 a_{pp}^{(k)} - 2\,c\,s a_{pq}^{(k)} + c^2 a_{qq}^{(k)}, \\ &\left.\begin{cases} a_{ip}^{(k+1)} = a_{pi}^{(k+1)} = c\,a_{ip}^{(k)} + s a_{iq}^{(k)} \\ a_{iq}^{(k+1)} = a_{qi}^{(k+1)} = -s a_{ip}^{(k)} + c\,a_{iq}^{(k)} \end{cases}\right\} \quad (i \neq p, q), \\ &a_{ij}^{(k+1)} = a_{ji}^{(k+1)} = a_{ij}^{(k)} \quad \text{(in other cases)}\,. \end{aligned}$$

Note that Givens algorithm requires altogether M extractions of roots and approximatively $\frac{4}{3}n^3$ multiplications.

Example 5.6.1.

Apply Givens method to reduction of matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}.$$

to tridiagonal form. Based on previous, it is necessary $M = 3$ steps. Givens rotations are thereby determined by

Table 5.6.1

k	(p,q)	c_k	q_k
1	$(2,3)$	0.707107	0.707107
2	$(2,4)$	0.816496	0.577350
3	$(3,4)$	0.397360	0.917663

and corresponding sequence of symmetric matrices $\mathbf{A}_k \quad (k = 2,3,4)$ (elements of lower triangle are not given):

$$\mathbf{A}_2 = \begin{bmatrix} 1. & 1.414214 & 0. & 1. \\ & 7. & 2. & 9.899495 \\ & & 1. & 4.242641 \\ & & & 20. \end{bmatrix},$$

$$\mathbf{A}_3 = \begin{bmatrix} 1. & 1.732051 & 0. & 0. \\ & 20.666667 & 4.082493 & 9.428090 \\ & & 1. & 2.309401 \\ & & & 6.333333 \end{bmatrix},$$

$$\mathbf{A}_3 = \begin{bmatrix} 1. & 1.732051 & 0. & 0. \\ & 20.666667 & 10.274023 & 0. \\ & & 7.175493 & 0.364642 \\ & & & 0.157895 \end{bmatrix}.$$

All results are rounded to six decimals.

2. Householder's method. This method is based on usage of sequence of orthogonal transformations of form

$$\mathbf{H} = \mathbf{I} - 2\vec{w}\vec{w}^T, \qquad \vec{w}^T\vec{w} = 1$$

with convenient chosen vectors $\vec{w}$. It is not difficult to show that these matrices are orthogonal.

Let $\mathbf{A}$ be real symmetric matrix of order n. Put $\mathbf{A}_1 = \mathbf{A}$ and define the series

$$\text{(5.6.2)} \qquad \mathbf{A}_{k+1} = \mathbf{H}_k^T \mathbf{A}_k \mathbf{H}_k, \qquad (k = 1, 2, \ldots, n-2)$$

where

$$\text{(5.6.3)} \qquad \mathbf{H}_k = \mathbf{I} - 2\vec{w}\vec{w}^T.$$

Special choice of vector $\vec{w}$ on every step in Householder method provides for matrix $\mathbf{A}_{n-1}$ to be tridiagonal. Note that all matrices in the series $\mathbf{A}_k$ are real and symmetric.

The strategy of Householder's method is that after first step the off-diagonal elements in the first row (and column) are zeroed, after second step elements in second row (and column), etc. Thereby, previously zeroed elements are not changed. Thus, for matrix $\mathbf{A}_k = [a_{ij}^{(k)}] \quad (k = 2, \ldots, n-1)$ we have

$$a_{ij}^{(k)} = 0 \qquad (1 \le i \le k-1 \wedge |i-j| > 1).$$

To secure such transformation, for $\vec{w}$ in (5.6.3) one shall take:

$$\vec{w} = \vec{0}, \quad \text{if} \qquad m_k = \sum_{j=k+2}^{n} (a_{ij}^{(k)})^2 = 0$$

and

$$\vec{w} = \beta\vec{v}, \quad \text{if} \qquad m_k \neq 0.$$

Thereby, for coordinates of vector $\vec{v} = [v_1 \ldots\ v_n]^T$ one shell take

$$v_i = 0 \ (i \le k), \quad v_{k+1} = 2Sy^2, \quad v_i = a_{ki}^{(k)} \quad (i \ge k+2),$$

where

$$S^2 = \sum_{j=k+1}^{n} (a_{ij}^{(k)})^2 \qquad \big(S = sgn(a_{k,k+1}^{(k)})\sqrt{S^2},\ \text{if}\ \ (a_{k,k+1}^{(k)} \neq 0\big),$$

$$y = \frac{1}{2K}(S + a_{k,k+1}^{(k)}),\ 2K^2 = S^2 + a_{k,k+1}^{(k)}S,\ \beta = \frac{1}{2Sy}.$$

Determination of matrix $\mathbf{A}_{k+1}$ in (5.6.2) can be simplified, regarding the fact that

$$\begin{aligned}\mathbf{A}_{k+1} &= (\mathbf{I} - 2\vec{w}\vec{w}^T)\mathbf{A}_k(\mathbf{I} - 2\vec{w}\vec{w}^T)\\ &= \mathbf{A}_k - 2\vec{w}\vec{w}^T\mathbf{A}_k - 2\mathbf{A}_k - 2\vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T\mathbf{A}_k\vec{w}\vec{w}^T,\end{aligned}$$

i.e.

$$\mathbf{A}_{k+1} = \mathbf{A}_k - 2\beta^2(\vec{v}\vec{u}^T + \vec{u}\vec{v}^T),$$

where

$$\vec{u} = \vec{\xi} - a\vec{v}, \qquad \xi = \mathbf{A}_k\vec{v}, \qquad a = \beta^2\vec{v}^T\vec{\xi}.$$

Note that Householder's method needs $n-2$ root extractions and nearly $\frac{2}{3}n^3$ multiplications, what is two times lesser than number of multiplications in Givens algorithm.

Example 5.6.2.

Apply Householder's method on transformation of matrix $\mathbf{A}$ from previous example. Now one needs only two steps (n-2=2).

First step (k=1): We have $m_1 = 2 \neq 0,\ S^2 = 3,\ S = 1.732051,\ K = 1.538189,\ y = 0.888074, \beta = 0.325058,\ a = 10.479274,$

$$\vec{v} = \begin{bmatrix} 0. \\ 2.732051 \\ 1. \\ 1. \end{bmatrix},\ \vec{\xi} = \begin{bmatrix} 4.732051 \\ 12.464102 \\ 24.196152 \\ 40.928203 \end{bmatrix},\ \vec{u} = \begin{bmatrix} 4.732051 \\ -16.165808 \\ 13.716878 \\ 30.448929 \end{bmatrix},$$

$$\mathbf{A}_2 = \begin{bmatrix} 0. & -1.732051 & 0. & 0. \\ & 20.666667 & -1.503206 & -10.163460 \\ & & 0.202565 & 0.666667 \\ & & & 7.130768 \end{bmatrix};$$

Second step (k=2): Now we have $m_2 = 103.295919 \neq 0,\ S^2 = 105.555556,\ S = -10.274023,\ K = 7.778160,\ y = -0.757070, \beta = 0.064283,\ a = 3.818322,$

$$\vec{v} = \begin{bmatrix} 0. \\ 0. \\ -11.777230 \\ -10.163460 \end{bmatrix},\ \vec{\xi} = \begin{bmatrix} 0. \\ 120.999532 \\ -9.161295 \\ -80.324767 \end{bmatrix},\ \vec{u} = \begin{bmatrix} 0. \\ 120.999532 \\ 35.819733 \\ -41.507243 \end{bmatrix},$$

$$\mathbf{A}_3 = \begin{bmatrix} 1. & -1.732051 & 0. & 0. \\ & 20.666667 & 10.274023 & 0. \\ & & 7.175439 & -0.364642 \\ & & & 0.157895 \end{bmatrix};$$

Elements of lower triangle in symmetric matrices $\mathbf{A}_2$ and $\mathbf{A}_3$ are not given.

Note that obtained tridiagonal matrix differs from this one obtained by Givens algorithm. Of course, this two matrices are similar, because operation of similarity with diagonal matrix $\mathbf{D} = diag(1, -1, -1, 1)$ exists.

Mention that by Givens and Househoher's method the Hermitian matrices transform to Hermitian tridiagonal matrices. Moreover, general complex matrices transform to Hessenberg form.

5.7. Eigenvalue problem for symmetric tridiagonal matrices

Let $\mathbf{A}$ be real symmetric tridiagonal matrix of order n which non-zero elements will be denoted as

$$a_{ii} = b_i \quad (i = 1, \ldots, n),$$
$$a_{i,i-1} = a_{i-1,i} = c_i \quad (i = 2, \ldots, n).$$

With $p_k(\lambda)$ denote main minor of order k of matrix $\mathbf{A} - \lambda\mathbf{I}$, i.e.

$$p_k(\lambda) = \begin{vmatrix} b_1 - \lambda & c_2 & & & & \\ c_2 & b_2 - \lambda & c_3 & & \mathbf{0} & \\ \ddots & \ddots & \ddots & & & \\ & & & & & \\ & \mathbf{0} & & c_{k-1} & b_{k-1} - \lambda & c_k \\ & & & & c_k & b_k - \lambda \end{vmatrix}$$

and define $p_0(\lambda) = 1$. Note that $p_1(\lambda) = b_1 - \lambda$.

By developing of determinant $p_k(\lambda)$ up to elements of last row we get

$$p_k(\lambda) = (b_k - \lambda)p_{k-1}(\lambda) - c_k^2 p_{k-2}(\lambda).$$

The value of characteristic polynomial of matrix $\mathbf{A}$ can be simple evaluated, based on previous, using three member recurrent relation

$$\begin{aligned} &p_k(\lambda) = (b_k - \lambda)p_{k-1}(\lambda) - c_k^2 p_{k-2}(\lambda) \quad (k = 2, \ldots, n), \\ &p_0(\lambda) = 1, \quad p_1(\lambda) = b_1 - \lambda. \end{aligned} \tag{5.7.1}$$

A simple method for determination of eigenvalues of symmetric tridiagonal matrices is based on usage of recurrent relation (5.7.1), method of interval bisection, and statement of the following theorem, which is simple to prove:

Theorem 5.7.1. *(Givens) Let all elements $c_k \neq 0$ of symmetric tridiagonal matrix $\mathbf{A}$ of order n. Then it holds:*

(1) Zeros of every polynomial p_k $(k = 2, \ldots, n)$ are real, different, and divided by zeros of polynomial p_{k-1};

(2) If $p_n(\lambda) \neq 0$, number of eigenvalues of matrix $\mathbf{A}$ less than λ is equal to number of sign change of $s(\lambda)$ in the series

$$p_0(\lambda),\ p_1(\lambda),\ \ldots, p_n(\lambda). \tag{5.7.2}$$

If some $p_k(\lambda) = 0$, then on this place in series (5.7.2) can be taken arbitrary sign, regarding to $p_{k-1}(\lambda)p_{k+1}(\lambda) < 0$.

Remark that in theorem there exists condition $c_k \neq 0$ for every $k = 2, \ldots, n$. If, for example, for some $k = m$, $c_m = 0$, then problem simplifies, because it splits in two problems of lower order (m and $n - m$). Namely, matrix $\mathbf{A}$ becomes

$$\mathbf{A} == \begin{bmatrix} \mathbf{A}' & \mathbf{0} \\ \mathbf{0} & \mathbf{A}'' \end{bmatrix},$$

where $\mathbf{A}'$ and $\mathbf{A}''$ are tridiagonal symmetric matrices of order m and $n - m$, respectively, and in this case is

$$det(\mathbf{A} - \lambda\mathbf{I}) = det(\mathbf{A}' - \lambda\mathbf{I})\, det(\mathbf{A}'' - \lambda\mathbf{I}).$$

Using multiple values for λ it is possible by systematic application of Theorem 5.7.1 to determine disjunct intervals in which lie eigenvalues of matrix $\mathbf{A}$. Thus, if we find that

$$s(\lambda_1) = s_1 \quad \text{and} \qquad s(\lambda_2) = s_2 = s_1 + 1 \quad (\lambda_1 < \lambda_2),$$

based on Theorem 5.7.1 we have that in interval (λ_1, λ_2) lies one only eigenvalue of matrix $\mathbf{A}$. Then for its determination the simple method of halving of interval (bisection method) can be used, by contraction of this starting interval up to desired exactness.

For determination of intervals in which lie eigenvalues it can be used also theorem of Gershgorin (Section 5.1), so that those intervals are

$$[b_1 - |c_2|,\ b_1 + |c_2|],$$
$$[b_i - |c_i| - |c_{i+1}|,\ b_i + |c_i| + |c_{i+1}|],\quad (i = 2, \ldots, n-1),$$
$$[b_n - |c_n|,\ b_n + |c_n|].$$

Unfortunately, these intervals are not disjunct, and in general case contain not one only eigenvalue of matrix $\mathbf{A}$.

Example 5.7.1.

For given matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & & \\ 1 & 3 & 2 & \\ & 2 & 5 & 3 \\ & & 3 & 7 \end{bmatrix}.$$

we have

$$p_0(\lambda) = 1,\ p_1(\lambda) = 1 - \lambda,\ p_2(\lambda) = (3 - \lambda)p_1(\lambda) - p_0(\lambda),$$
$$p_3(\lambda) = (5 - \lambda)p_2(\lambda) - 4p_1(\lambda),\ p_4(\lambda) = (7 - \lambda)p_3(\lambda) - 9p_2(\lambda).$$

Let $\lambda = 0$. Then we have $p_0(0) = 1,\ p_1(0) = 1,\ p_2(0) = 2,\ p_3(0) = 6,\ p_4(0) = 24$. Thus, in the series $(5.7.2)$ are $+ + + + +$, what means that there is no sign change, i.e. $s(0) = 0$. According to Theorem $5.7.1$, matrix $\mathbf{A}$ does not have negative eigenvalues, i.e. it is positive-definitive.

Taking in sequence for λ values $1, 2, 4, 5, 7, 9, 10$ we get the results given in table 5.7.1.

Table 5.7.1

λ	$p_0(\lambda)$	$p_1(\lambda)$	$p_2(\lambda)$	$p_3(\lambda)$	$p_4(\lambda)$	$s(\lambda)$
1	1	0	−1	−4	−15	1
2	1	−1	−2	−2	8	2
4	1	−3	2	14	24	2
5	1	−4	7	16	−31	3
7	1	−6	23	−22	−207	3
9	1	−8	47	−156	−111	3
10	1	−9	62	−274	264	4

Based on values of $s(\lambda)$ we conclude that in each interval $(0, 1), (1, 2), (4, 5), (9, 10)$ is located one eigenvalue of matrix $\mathbf{A}$. These eigenvalues with six figures are

$$\lambda_1 \cong 0.322548,\ \lambda_1 \cong 1.745761,\ \lambda_1 \cong 4.536620,\ \lambda_1 \cong 9.395071.$$

Note that these are zeroes of Laguerre polynomial L_4.

5.8. LR and QR algorithms

This section is devoted to so known factorization methods. First such method for solution of problem of eigenvalues for arbitrary matrix $\mathbf{A}$ was described by H. Rutishauser ([14]) in the year 1958, which called it **LR** transformation. Method consists in construction of series of matrices $\{\mathbf{A}_k\}_{k \in N}$, starting from $\mathbf{A}_1 = \mathbf{A}$, in the following way: Matrix $\mathbf{A}$ factorizes to lower triangular matrix $\mathbf{L}_k$ with unit diagonal and upper triangular matrix $\mathbf{R}_k$, i.e.

$$\mathbf{A}_k = \mathbf{L}_k \mathbf{R}_k, \tag{5.8.1}$$

and then the following member is determined by multiplication of obtained factors in opposite sequence, i.e.

$$\mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{L}_k. \tag{5.8.2}$$

Note that matrices $\mathbf{A}_{k+1}$ and $\mathbf{A}_k$ are similar, because they are connected with transformation of similarity

$$\mathbf{A}_{k+1} = \mathbf{L}_k^{-1}\mathbf{A}_k\mathbf{L}_k. \tag{5.8.2}$$

Factorization of (5.8.1) can be performed by Gauss method of elimination.

If we put

$$\mathbf{L}^{(k)} = \mathbf{L}_1 \ldots \mathbf{L}_k \quad \text{and} \quad \mathbf{R}^{(k)} = \mathbf{R}_k \ldots \mathbf{R}_1,$$

based on (5.8.2) we have

$$\mathbf{L}_k\mathbf{A}_{k+1} = \mathbf{A}\mathbf{L}^{(k)},$$

wherefrom it follows

$$\mathbf{L^{(k)}R^{(k)}} = \mathbf{L}^{(k-1)}\mathbf{A}_k\mathbf{R}^{(k-1)} = \mathbf{A}\mathbf{L}^{(k-1)}\mathbf{R}^{(k-1)}.$$

By iterating the last equality, we get

$$\mathbf{L}^{(k)}\mathbf{R}^{(k)} = \mathbf{A}^2\mathbf{L}^{(k-2)}\mathbf{R}^{(k-2)} = \ldots \mathbf{A}^k,$$

what means that $\mathbf{L^{(k)}R^{(k)}}$ is factorization of matrix $\mathbf{A}^k$. Using this facts, Rutishauser (see, also [7]) showed that under certain conditions series of matrices $\{\mathbf{A}_k\}$ converges towards some upper triangular matrix, which elements on the main diagonal give eigenvalues of matrix $\mathbf{A}$. Usually, $\mathbf{LR}$ method is applied to matrices previously reduced to upper Hessenberg form ($a_{ij} = 0$ for $i \geq j + 2$). If, by means of some method, general matrix reduced to lower Hessenberg form we apply $\mathbf{LR}$ method to transposed matrix, which has the same eigenvalues. All matrices in series $\{\mathbf{A}_k\}$ have Hessenberg form. Acceleration of convergence of series $\{\mathbf{A}_k\}$ can be done by convenient shifting p_k, so that, in spite of $\mathbf{A}_k$ we factorize $\mathbf{B}_k = \mathbf{A}_k - p_k\mathbf{I} = \mathbf{L}_k\mathbf{R}_k$, whereby $\mathbf{A}_{k+1} = p_k\mathbf{I} + \mathbf{R}_k\mathbf{L}_k$.

Unfortunately, $\mathbf{LR}$ algorithm has several disadvantages (see monograph of Wilkinson [7]). For example, factorization does not exist for every matrix. One better factorization method was developed by J.G.F. Francis ([15]) and V.N. Kublanovskaya ([16]), where matrix $\mathbf{L}$ is replaced with unitary matrix $\mathbf{Q}$. So one gets $\mathbf{QL}$ algorithm defined by

$$\mathbf{A}_k = \mathbf{Q}_k\mathbf{R}_k, \; \mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{Q}_k \quad (k = 1, 2, \ldots), \tag{5.8.3}$$

starting from $\mathbf{A}_1 = \mathbf{A}$. Note that $\mathbf{A}_{k+1} = \mathbf{Q} * \mathbf{A}_k\mathbf{Q}_k$.

If we put

$$\mathbf{Q}^{(k)} = \mathbf{Q}_1 \ldots \mathbf{Q}_k \quad \text{and} \quad \mathbf{R}^{(k)} = \mathbf{R}_k \ldots \mathbf{R}_1, \tag{5.8.4}$$

similar as $\mathbf{LR}$ method, we find

$$\mathbf{Q}^{(k)}\mathbf{A}_{k+1} = \mathbf{A}\mathbf{Q}^{(k)} \quad \text{and} \quad \mathbf{Q}^{(k)}\mathbf{R}_k = \mathbf{A}^k. \tag{5.8.5}$$

Theorem 5.8.1. *If matrix $\mathbf{A}$ regular, then exists decomposition $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is unitary, and $\mathbf{R}$ upper triangular matrix. Moreover, if diagonal elements of matrix $\mathbf{R}$ are positive, decomposition is unique.*

$\mathbf{QR}$ factorization (5.8.3) can be performed by using unitary matrices of form $\mathbf{I} - 2\vec{w}\vec{w}^*$. So, in order to transform $\mathbf{A}_k$ to $\mathbf{R}_k$, i.e. reduction of columns to $\mathbf{A}_k$, we have

$$(\mathbf{I} - 2\vec{w}_{n-1}\vec{w}_{n-1}^*) \ldots (\mathbf{I} - 2\vec{w}_{n-1}\vec{w}_{n-1}^*)\mathbf{A}_k = \mathbf{R}_k. \tag{5.8.6}$$

The matrix $\mathbf{Q}_k$ is then

$$\mathbf{Q}_k = (\mathbf{I} - 2\vec{w}_1\vec{w}_1^*)\ldots(\mathbf{I} - 2\vec{w}_{n-1}\vec{w}_{n-1}^*). \tag{5.8.7}$$

QR algorithm is efficient if initial matrix has (upper) Hessenberg form. Then, previously mentioned unitary matrices reduce to two-dimensional rotations. All matrices $\mathbf{A}_k$ are of Hessenberg form. Thus, eigenvalue problem for general matrix is most convenient to be solved in two steps. At first, reduce matrix to Hessenberg form, and then apply the **QR** algorithm.

In special case, when initial matrix is tridiagonal, matrices $\mathbf{A}_k$ in **QR** algorithm are also tridiagonal. In that case, using conveniently chosen shift p_k, **QR** algorithm becomes very efficient for solving eigenvalue problem of tridiagonal matrices.

By introducing of shift p_k, the formulae (5.8.3) become

$$\mathbf{A}_k - p_k\mathbf{I} = \mathbf{Q}_k\mathbf{R}_k,\ \mathbf{A}_{k+1} = p_k\mathbf{I} + \mathbf{R}_k\mathbf{Q}_k \quad (k = 1, 2, \ldots). \tag{5.8.8}$$

Suppose that $\mathbf{A}(= \mathbf{A}_1)$ is symmetric tridiagonal real matrix. All others matrices $\mathbf{A}_k = [a_{ij}^{(k)}]_{n\times n}$ are the same. In order to simplify, introduce notation from section 5.7., i.e.

$$\begin{gathered} a_{ii}^{(k)} = b_i^{(k)} \qquad (i = 1, \ldots, n), \\ a_{i,i-1}^{(k)} = a_{i-1,i}^{(k)} = c_i^{(k)} \quad (i = 2, \ldots, n), \end{gathered}$$

and suppose $c_i^{(1)} \neq 0 \quad (i = 2, \ldots, n)$. Then matrix $\mathbf{A}_1$ has all different eigenvalues.

There are two ways for choosing of shift p_k. The first one is to take for p_k value of matrix element located in lower right corner of matrix $\mathbf{A}_k$, i.e. $p_k = b_n^{(k)}$. The second way of p_k selection is that for p_k is taken such eigenvalue of matrix of type 2×2

$$\begin{bmatrix} b_{n-1}^{(k)} & c_n^{(k)} \\ c_n^{(k)} & b_n^{(k)} \end{bmatrix} \tag{5.8.9}$$

which is closer to the value $b_n^{(k)}$. Such choice

$$p_k = b_n^{(k)} + d - sgn(d)\sqrt{d^2 + (c_n^{(k)})^2}, \quad d = \frac{1}{2}(b_{n-1}^{(k)} - b_n^{(k)}) \tag{5.8.10}$$

emanates from Wilkinson and gives faster convergence of **QR** algorithm in comparison to choice $p_k = b_n^{(k)}$. In both cases we have convergence such that

$$c_n^{(k)}c_{n-1}^{(k)} \to 0 \quad (k \to \infty).$$

If $c_n^{(k)}$ becomes neglectful small (i.e. machine zero), we can take $b_n^{(k)}$ as one eigenvalue of matrix $\mathbf{A}$ and put last row and column in $\mathbf{A}_k$ away, so that procedure is to be continued with problem of dimension $n-1$. Nevertheless, if $c_{n-1}^{(k)}$ is neglectfully small and $c_n^{(k)}$ significant, we can at once determine two eigenvalues of matrix $\mathbf{A}_k$, and those are actually eigenvalues of matrix (5.8.9). The process is to be continued by dropping last two rows and columns and applying algorithm (5.8.8) on the problem of dimension $n-2$. In this way, **QR** algorithm becomes very efficient because it produces deflation of matrix order.

In practical application of **QR** algorithm, decomposition (5.8.3) or (5.8.8) is not necessary explicitly to calculate. Namely, it is possible directly to obtain (5.8.5), i.e.

$$\mathbf{A}_{k+1} = \mathbf{Q}^{(k)^T}\mathbf{A}\mathbf{Q}^{(k)},$$

where now $\mathbf{Q}^{(k)}$ is orthogonal matrix given by (5.8.4). Note that matrix $\mathbf{Q}^k$ converges to orthogonal matrix of eigenvectors of matrix $\mathbf{A}$ (compare with the case of Jacobi method, section 5.5.).

As already given, unitary (orthogonal) matrices which appear in (5.8.6) and (5.8.7) reduce to two-dimensional rotations.

Let $\mathbf{Z}_p = \mathbf{Z}_p^{(k)} = (\mathbf{I} - 2\vec{w}_p\vec{w}_p^T)$, $\vec{w}_p = [w_{p1} \ldots \vec{w}_{pn}]^T$, where for coordinates of vector $\vec{w}_p$ we took

$$w_{pp} = \sin\frac{\theta}{2},\ w_{p,p+1} = \cos\frac{\theta}{2},\ w_{pi} = 0\ (i \neq p, p+1).$$

Based on (5.4.8) we have

$$\mathbf{Z}_p = \begin{array}{c} \left[\begin{array}{cccccccc} 1 & & & & & & & \\ & \ddots & & & & \mathbf{0} & & \\ & & 1 & & & & & \\ & & & c & s & & & \\ & & & s & -c & & & \\ & & & & & 1 & & \\ & \mathbf{0} & & & & & \ddots & \\ & & & & & & & 1 \end{array}\right] \begin{array}{l} \\ \\ \\ \leftarrow p \\ \leftarrow p+1 \\ \\ \\ \\ \end{array} \\ \begin{array}{cc} \uparrow & \uparrow \\ p & p+1 \end{array} \end{array}$$

where $c = \cos\theta$ and $s = \sin\theta$.

In order to simplify, in matrix $\mathbf{A}_k$ omit iteration index, i.e. take that $b_i^{(k)} = b_i$ and $c_i^{(k)} = c_i$. Chose now the angle $\theta = \theta_1^{(k)} = \theta_1$, so that elements in first column under main diagonal in matrix $\mathbf{Z}_1 A_k$ are equal to zero. Than we have $\cot\theta_1 = b_1/c_2$, i.e.

$$c = \frac{b_1}{\sqrt{b_1^2 + c_2^2}} \text{ and } s = \frac{c_2}{\sqrt{b_1^2 + c_2^2}}.$$

In matrix

$$\mathbf{Z}_1\mathbf{A}_k\mathbf{Z}_1 = \begin{bmatrix} b_1' & c_2' & d_1 & 0 & \cdots & 0 & 0 \\ c_2' & b_2' & c_3' & 0 & & 0 & 0 \\ d_1 & c_3' & b_3 & c_4 & & 0 & 0 \\ 0 & 0 & c_4 & b_4 & & 0 & 0 \\ \vdots & & & & & 0 & 0 \\ 0 & 0 & 0 & 0 & & b_{n-1} & c_n \\ 0 & 0 & 0 & 0 & & c_n & b_n \end{bmatrix}$$

elements denoted with prime are these ones elements of matrix $\mathbf{A}_k$ which change during transformation. Matrix $\mathbf{A}_{k+1}$ is obtained by

$$\mathbf{A}_{k+1} = \mathbf{Z}_{n-1} \cdots \mathbf{Z}_2\mathbf{Z}_1\mathbf{A}_k\mathbf{Z}_1\mathbf{Z}_2 \cdots \mathbf{Z}_{n-1},$$

where $\mathbf{Z}_2 \cdots \mathbf{Z}_{n-1}$ are constructed in similar way as $\mathbf{Z}_1$, so that matrix $\mathbf{A}_{k+1}$ becomes tridiagonal. The product of all orthogonal (two-dimensional) rotations

$$\mathbf{Z} = \prod_{k=1}^{\infty}(\mathbf{Z}_1^{(k)}\mathbf{Z}_2^{(k)} \ldots \mathbf{Z}_{n-1}^{(k)})$$

gives a matrix of eigenvectors. Namely, here we have

$$\prod_{j=1}^{k}(\mathbf{Z}_1^{(j)}\mathbf{Z}_2^{(j)} \ldots \mathbf{Z}_{n-1}^{(j)}) = \mathbf{Q}_{(k)}.$$

Now we can formulate recursively $\mathbf{QR}$ algorithm for determining one eigenvalue λ and corresponding orthonormed vector $\vec{x} = \mathbf{Z}\vec{e}_1$ ($\vec{x}^T\vec{x} = 1,\ \vec{e}_1 = [1\ 0\ \cdots\ 0]^T$) of real symmetric tridiagonal matrix.

Let $\lambda^{(k)}$ and $\vec{y}^{\,(k)=[y_1^{(k)}y_2^{(k)}\cdots y_n^{(k)}]^T}$ be approximations of eigenvalue λ and eigenvector $\vec{x}$ in k-th iterative step $(k = 1, 2, \ldots)$.

Starting from $\vec{y}^{\,(1)} = \vec{e}_1$, i.e. $y_1^{(1)} = 1$, $y_i^{(1)} = 0$ $(i = 2, \ldots, n)$, k-th iterative step can be expressed in the following way ([17]):

For $p = 1, 2, \ldots, n-1$ we determinate

$$\begin{aligned}
\alpha &:= [(\bar{c}_p^{(k)})^2 + (d_p^{(k)})^2]^{1/2},\ c := \bar{c}_p^{(k)}/\alpha,\ s := d_p^{(k)}/\alpha,\\
b_p^{(k+1)} &:= c^2\bar{b}_p^{(k)} + 2\,c\,s\tilde{c}_{p+1}^{(k)} + s^2 b_{p+1}^{(k)},\\
\bar{b}_{p+1}^{(k)} &:= s^2\bar{b}_p^{(k)} - 2\,c\,s\tilde{c}_{p+1}^{(k)} + c^2 b_{p+1}^{(k)},\\
c_p^{(k+1)} &:= c\bar{c}_p^{(k)} + s\,d_p^{(k)} = \alpha,\\
c_p^{(k)} &:= (\bar{b}_p^{(k)} - b_{p+1}^{(k)})cs + \tilde{c}_{p+1}^{(k)}(s^2 - c^2),\\
\tilde{c}_{p+2}^{(k)} &:= -c\,c_{p+2}^{(k)},\ d_{p+1}^{(k)} := s\,c_{p+2}^{(k)},\\
y_p^{(k+1)} &:= c\bar{y}_p^{(k)} + sy_{p+1}^{(k)},\ \bar{y}_{p+1}^{(k)} := s\,\bar{y}_p^{(k)} - c\,y_{p+1}^{(k)},
\end{aligned}$$

using

$$\begin{aligned}
d_1^{(k)} &:= c_2^{(k)},\ \bar{c}_1^{(k)} := b_1^{(k)} - \lambda^{(k)},\\
\bar{b}_1^{(k)} &:= \bar{b}_1^{(k)},\ \tilde{c}_2^{(k)} := c_2^{(k)},\ \bar{y}_1^{(k)} := y_1^{(k)},
\end{aligned}$$

and $\lambda_k = p_k$, where p_k is eigenvalue of matrix (5.8.9) determined by (5.8.10).

The iterative process interrupts when, for example, $c_n^{(k)}$ becomes small enough. As told previously, in this case we take that $\lambda := b_n^{(k)}$ and $\vec{x} := \vec{y}^{(k)}$, eliminate last row and last column in matrix $\mathbf{A}_{k+1}$, and then repeat complete iterative process over this matrix of order $n-1$. Then we determine second eigenvalue and corresponding eigenvector (see method of deflation, Section 5.4), and so on. We act in the same way if $c_{n-1}^{(k)}$ is neglectful, and $c_n^{(k)}$ significant. In this case we determine two eigenvalues simultaneously.

Remark that series $\bar{b}_p^{(k)}$, $\bar{c}_p^{(k)}$, $\tilde{c}_p^{(k)}$, $\bar{y}_p^{(k)}$ do not demand additional space in memory. Namely, they can be stored in the same locations for series $b_p^{(k)}$, $c_p^{(k)}$, $y_p^{(k)}$.

Similar to **QR** algorithm, it is developed **QL** algorithm ([18]), where **L** is lower triangular matrix, and **Q** unitary matrix. Also, it has been developed so known implicit **QL** algorithm ([19]).

5.9. Software eigenpackages

Some general guidelines for solving eigenproblems are summarized below [23].

- When only the largest and (or) the smallest eigenvalue of a matrix is required, the power method can be employed.
- Although it is rather inefficient, the power method can be used to solve for intermediate eigenvalues.
- The direct method is not a good method for solving linear eigenproblems. However, it can be used for solving nonlinear eigenproblems.
- For serious eigenproblems, the QR method is recommended.
- Eigenvectors corresponding to a known eigenvalue can be determined by one application of the shifted inverse power method.

Almost all software routines in use nowadays trace their ancestry back to routines published in Wilkinson and Reinsch's boock *Handbook for Automatic Computation, Vol. II, Linear Algebra* [2]. A public-domain implementation of the *Handbook* routines in `FORTRAN` is the `EISPACK` set of programs [3]. The routines presented in majority of most frequently used software packages are translations of either the *Handbook* or `EISPACK` routines, so understanding these will take a lot of the way towards understanding those canonical packages.

IMSL [4] and NAG [5] each provide proprietary implementations in FORTRAN of what are essentially the *Handbook* routines.

Many commercial software packages contain eigenproblem solvers. Some of the more prominent packages are Matlab and Mathcad. More sophisticated packages, such as Mathematica, Macsyma, and Maple also contain eigenproblem solvers. The book *Numerical Recepies* [2] contains subroutines and advice for solving eigenproblems.

A good ”eigenpackage” will provide separate routines, or separate paths through sequences of routines, for the following desired calculations

- all eigenvalues and no eigenvectors
- all eigenvalues and some corresponding eigenvectors
- all eigenvalues and all corresponding eigenvectors.

The purpose of these distinctions is to save compute time and storage; it is wasteful to calculate eigenvectors that you don’t need. Often one is interested only in the eigenvectors corresponding to the largest few eigenvalues, or largest few in the magnitude, or few that are negative. The method usually used to calculate ”some” eigenvectors is typically more efficient than calculating all eigenvectors if you desire fewer than about a quarter of the eigenvectors.

A good eigenpackage also provides separate paths for each of the above calculations for each of the following special forms of the matrix:

- real, symmetric, tridiagonal
- real, symmetric, banded (only a small number of sub- and super-diagonals are nonzero)
- real, symmetric
- real, nonsymmetric
- complex, Hermitian
- complex, non-Hermitian

Again, the purpose of these distinctions is to save time and storage by using the least general routine that will serve in any particular application.

Good routines for the following paths are available:

- all eigenvalues and eigenvectors of a real, symmetric, tridiagonal matrix
- all eigenvalues and eigenvectors of a real, symmetric, matrix
- all eigenvalues and eigenvectors of a complex, Hermitian matrix
- all eigenvalues and no eigenvectors of a real, nonsymmetric matrix.

5.10. Generalized and Nonlinear Eigenvalue Problems

Many eigenpackages also deal with the so-called *generalized eigenproblem* [6],

$$\mathbf{A} \cdot \vec{\mathbf{x}} = \lambda \mathbf{B} \cdot \vec{\mathbf{x}} \tag{5.10.1}$$

where **A** and **B** are both matrices. Most such problems, where **B** is nonsingular, can be handled by the equivalent

$$(\mathbf{B}^{-1} \cdot \mathbf{A}) \cdot \vec{x} = \lambda \vec{x} \tag{5.10.2}$$

Often **A** and **B** are symmetric and **B** is positive definite. The matrix $\mathbf{B}^{-1} \cdot \mathbf{A}$ in (5.10.2) is not symmetric, but we can recover a symmetric eigenvalue problem by using the Cholesky decomposition $\mathbf{B} = \mathbf{L} \cdot \mathbf{L}^T$. Multiplying equation (5.10.1) by $\mathbf{L}^{-1}$ we get

$$\mathbf{C} \cdot (\mathbf{L}^T \cdot \vec{x}) = \lambda (\mathbf{L}^T \cdot \vec{x}) \tag{5.10.3}$$

where

$$\mathbf{C} = \mathbf{L}^{-1} \cdot \mathbf{A} (\mathbf{L}^{-1})^T \tag{5.10.4}$$

The matrix $\mathbf{C}$ is symmetric and its eigenvalues are the same as those of the original problem (5.10.1); its eigenfunctions are $\mathbf{L}^T \cdot \vec{x}$. The efficient way to form $\mathbf{C}$ is first to solve the equation

$$\mathbf{Y} \cdot \mathbf{L}^T = \mathbf{A} \tag{5.10.5}$$

for the lower triangle of the matrix $\mathbf{Y}$. Then solve

$$\mathbf{L} \cdot \mathbf{C} = \mathbf{Y} \tag{5.10.6}$$

for the lower triangle of the symmetric matrix $\mathbf{C}$.

Another generalization of the standard eigenvalue problem is to problems nonlinear in the eigenvalue λ, for example,

$$(\mathbf{A}\lambda^2 + \mathbf{B}\lambda + \mathbf{C}) \cdot \vec{x} = 0 \tag{5.10.7}$$

This can be turned into a linear problem by introducing an additional unknown eigenvector $\mathbf{y}$ and solving the $2N \times 2N$ eigensystem,

$$\begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{A}^{-1} \cdot \mathbf{C} & -\mathbf{A}^{-1} \cdot \mathbf{B} \end{bmatrix} \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \lambda \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}.$$

This technique generalizes to higher-order polynomials in λ. A polynomial of degree M produces a linear $MN \times MN$ eigensystem, as given in [7].

Bibliography (Cited references and further reading)

[1] Milovanović , G.V., *Numerical Analysis I*, Naučna knjiga, Beograd, 1988 (Serbian).

[2] Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., *Numerical Recepies - The Art of Scientific Computing.* Cambridge University Press, 1989.

[3] Smith, B.T., et al., *Matrix Eigensystem Routines - EISPACK Guide , 2nd ed., vol 6 of Lecture Notes in Computer Science,*, Springer, New York, 1976.

[4] *IMSL Math/Library Users Manual* , IMSL Inc., 2500 City West Boulevard, Houston TX 77042

[5] *NAG Fortran Library*, Numerical Algorithms Group, 256 Banbury Road, Oxford OX27DE, U.K., Chapter F02.

[6] Golub, G.H., and Van Loan, C.F., *Matrix Computation*, Johns Hopkins University Press, Baltimore 1989.

[7] Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, 1965.

[8] Acton, F.S., *Numerical Methods that Work*, corrected edition, Mathematical Association of America, Chapter 13, Washington, 1970.

[9] Horn, R.A., and Johnson, C.R., *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[10] Milovanović, G.V. and Djordjević, Dj.R., *Programiranje numeričkih metoda na FORTRAN jeziku.* Institut za dokumentaciju zaštite na radu ”Edvard Kardelj”, Niš, 1981 (Serbian).

[11] Parlett, B.N. and Poole, W.G., *A geometric theory for the QR, LU, and power iterations.* SIAM J. Numer. Anal. 10(1973), 389-412.

[12] Bart, W., Martin, R.S., Wilkonson,J.H. *Calculation of the eigenvalues of a symmetric tridiagonal matrix by the bisection method.* Numer Math. 9(1967), 386-393.

[13] Wilkinson, J.H. & Reisch, C., *Handbook for Automatic Computation. Vol. II Linear Algebra.* Springer Verlag, Berlin-Heidelberg-New York, 1971.

[14] Rutishauser, H., *Solution of eigenvalue problem with the LR-transformation.* Appl. Math. Ser. Nat. Bur. Stand. 49(1958), 47-81.

[15] Francis, J.G.F., *The QR transformation - a unitary analogue to the LR transformation.* Comput.J. 4(1961/62), 265-271, 332-345.

[16] Kublanovskaya, V.N., *O nekotoryh algorifmah dlja rešenija polnoǐ problemy sobstvennyh značeniǐ.* Ž. Vyčisl. Mat. i Mat. Fiz. 1(1961), 555-570.

[17] Golub, G.H. & Welsch, J.H., *Calculation of Gauss quadrature rules.* Math. Comp. 23(1969), 221-230.

[18] Bowdler, H., Martin, R.S., Reinsch, C. Wilkinson, J.H., *The QR and QL algorithms for symmetric matrices.* Numer. Math. 11(1968), 293-306.

[19] Dubrulle, A., Martin, R.S., Wilkinson, J.H., *The implicit QR algorithm.* Numer. Math. 12(1968), 377-383.

[20] Stoer, J., and Bulirsch, R., *Introduction to Numerical Analysis*, Springer, New York, 1980.

[21] Mitrinović, D.S. and Djoković, D.Ž., *Polinomi i matrice.* ICS, Beograd, 1975.

[22] Wilkinson, J.H., and Reinsch, C., *Linear Algebra, vol. II of Handbook for Automatic Computation*, Springer, New York, 1971.

[23] Hoffman, J.D., *Numerical Methods for Engineers and Scientists.* Taylor & Francis, Boca Raton-London-New York-Singapore, 2001.