Chapter 2

EIGENVALUES AND EIGENVECTORS

1. Definitions

Let $A \in \mathbf{C}^{n\times n}, \lambda \in \mathbf{C}$ and $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$, such that the following relation holds:

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{1}$$

Then λ is called *eigenvalue* of A and $\mathbf{x}$ is called *eigenvector* corresponding to λ. The set of the eigenvalues of a matrix A is called *spectrum* of A and the largest absolute value $\rho(A)$ among the eigenvalues of A is called *spectral radius* of A.

The linear system (1), which can be written also in the form

$$(A - \lambda I)\mathbf{x} = \mathbf{0}, \tag{2}$$

admits nonzero solutions if and only if

$$\det(A - \lambda I) = 0. \tag{3}$$

By expanding $\det(A - \lambda I)$ as a polynomial in λ we obtain

$$\det(A - \lambda I) = P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n,$$

with

$$a_0 = (\text{-}1)^n, \qquad a_i = (\text{-}1)^{n-i}\sigma_i, \quad i = 1, \ldots, n,$$

where σ_i is the sum of the determinants of all the $\binom{n}{i}$ principal submatrices of A of order i. In detail:

$$a_1 = (\text{-}1)^{n-1}\mathrm{tr}\, A, \qquad a_n = \det A,$$

where we have denoted as $\mathrm{tr}\, A$ the *trace* of A, that is the sum of the principal entries of A.

From the well known formulas which relate the coefficients of an algebraic equation of degree n to the sum and the product of its roots we have:

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}\, A \quad \text{e} \quad \prod_{i=1}^{n} \lambda_i = \det A. \tag{4}$$

The polynomial $P(\lambda)$ is called *characteristic polynomial* of A and the equation $P(\lambda) = 0$ is called *characteristic equation* of A.

Due to the fundamental theorem of algebra, the characteristic equation admits n roots over the field of complex numbers, taking into account their multiplicities. Therefore a matrix of order n has, taking into account their multiplicities, n eigenvalues over the complex field.

Since any nonzero solution of the homogeneous linear system (2)is an eigenvector, the eigenvectors corresponding to an eigenvalue λ are determined up to a multiplicative constant $\alpha \neq 0$, that is if $\mathbf{x}$ is an eigenvector of A, $\alpha\mathbf{x}$ is eigenvector of A as well, for the same eigenvalue.

2.1 Example. The characteristic polynomial of the matrix

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$$

is the expansion of the determinant

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{bmatrix} = \lambda^2 - 2\lambda - 8.$$

The corresponding characteristic equation

$$\lambda^2 - 2\lambda - 8 = 0$$

has the roots $\lambda_1 = -2$ e $\lambda_2 = 4$, which are the eigenvalues of the matrix A. The eigenvectors corresponding to $\lambda_1 = -2$ are computed by solving the system (2):

$$\begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}.$$

From the first equation we obtain:

$$x_1 + x_2 = 0, \quad \text{da cui} \quad x_1 = -x_2,$$

therefore any vector

$$\mathbf{x}_1 = \alpha \begin{bmatrix} 1 \\ \text{-1} \end{bmatrix}$$

with $\alpha \neq 0$, is an eigenvector of the matrix A corresponding to the eigenvalue $\lambda_1 = -2$. The eigenvectors corresponding to $\lambda_2 = 4$ are computed by solving the system

$$\begin{bmatrix} \text{-3} & 3 \\ 3 & \text{-3} \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}.$$

From the first equation we obtain:

$$-x_1 + x_2 = 0, \quad \text{therefore} \quad x_1 = x_2,$$

thus any vector

$$\mathbf{x}_1 = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

with $\alpha \neq 0$, is an eigenvector of the matrix A corresponding to the eigenvalue $\lambda_2 = 4$. ■

2. Properties of the eigenvalues

– The eigenvalues of a diagonal or triangular (upper or lower) matrix A are equal to the principal entries. This is due to the fact that the matrix $A - \lambda I$ is diagonal or triangular as well, therefore its determinant is given by the product of the principal entries.

– If λ is an eigenvalue of a nonsingular matrix A and $\mathbf{x}$ is an eigenvector corresponding to λ, then $\lambda \neq 0$ and $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\mathbf{x}$. In detail, from

$$A\mathbf{x} = \lambda\mathbf{x}$$

we have

$$\mathbf{x} = \lambda A^{-1}\mathbf{x}$$

and therefore

$$\lambda \neq 0 \quad \text{and} \quad A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}.$$

– If λ is an eigenvalue of a matrix A, then $\overline{\lambda}$ is an eigenvalue of A^H and λ is an eigenvalue of A^T. In detail, in the first case, since

$$\det A^H = \overline{\det A},$$

we have

$$0 = \det(A - \lambda I) = \overline{\det(A - \lambda I)^H} = \overline{\det(A^H - \overline{\lambda} I)},$$

and

$$\det(A^H - \overline{\lambda} I) = 0.$$

One can proceed analogously in the second case.

– If λ is an eigenvector of a unitary matrix A, i.e. a matrix such that $A^H A = AA^H = I$, then $|\lambda| = 1$. In detail from the equation $A\mathbf{x} = \lambda\mathbf{x}$ we have

$$(A\mathbf{x})^H = (\lambda\mathbf{x})^H.$$

$$\mathbf{x}^H A^H = \overline{\lambda}\mathbf{x}^H,$$

and finally

$$\mathbf{x}^H A^H A\mathbf{x} = \overline{\lambda}\lambda\, \mathbf{x}^H\mathbf{x}.$$

Since A is unitary, we have

$$\mathbf{x}^H\mathbf{x} = \overline{\lambda}\lambda\mathbf{x}^H\mathbf{x},$$

therefore, since $\mathbf{x}^H\mathbf{x} \neq 0$,

$$\overline{\lambda}\lambda = |\lambda|^2 = 1.$$

2.2 Example. The matrix G introduced in example 1.1:

$$G = \begin{bmatrix} \cos\phi & \text{-}\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \qquad \phi \in \mathbf{R},$$

is unitary, thus its eigenvalues have absolute value 1. In detail, from the characteristic equation

$$\lambda^2 - 2\lambda\cos\phi + 1 = 0$$

we have

$$\lambda_1 = \cos\phi + \mathbf{i}\sin\phi \quad \text{e} \quad \lambda_2 = \cos\phi - \mathbf{i}\sin\phi.$$ ■

Matrix polynomial are of relevant interest. Let

$$p(x) = \alpha_0 x^k + \alpha_1 x^{k-1} + \ldots + \alpha_k,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k \in \mathbf{C}$, be a polynomial of degree k in the variable x, and let $A \in \mathbf{C}^{n\times n}$. A *polynomial in the matrix* A is a matrix defined as

$$p(A) = \alpha_0 A^k + \alpha_1 A^{k-1} + \ldots + \alpha_k I$$

(see exercise 1.3). If λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, then $p(\lambda)$ is an eigenvalue of $p(A)$ and $\mathbf{x}$ is a corresponding eigenvector. In detail we have:

$$A^i\mathbf{x} = A^{i-1}A\mathbf{x} = A^{i-1}\lambda\mathbf{x} = \lambda A^{i-1}\mathbf{x} = \lambda A^{i-2}A\mathbf{x} = \ldots = \lambda^i\mathbf{x},$$

therefore

$$\begin{aligned} p(A)\mathbf{x} &= \alpha_0 A^k\mathbf{x} + \alpha_1 A^{k-1}\mathbf{x} + \ldots + \alpha_k\mathbf{x} = \alpha_0\lambda^k\mathbf{x} + \alpha_1\lambda^{k-1}\mathbf{x} + \ldots + \alpha_k\mathbf{x} \\ &= p(\lambda)\mathbf{x}. \end{aligned}$$

2.3 Example. The matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

has the eigenvalues

$$\lambda_1 = 1 - \sqrt{2}, \quad \lambda_2 = 1, \quad \lambda_3 = 1 + \sqrt{2},$$

whose corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}.$$

The matrix

$$B = 3A^2 - A + 2I = \begin{bmatrix} 7 & 5 & 3 \\ 5 & 10 & 5 \\ 3 & 5 & 7 \end{bmatrix},$$

has the eigenvalues

$$\mu_i = 3\lambda_i^2 - \lambda_i + 2, \quad i = 1, 2, 3,$$

therefore

$$\mu_1 = 10 - 5\sqrt{2}, \quad \mu_2 = 4, \quad \mu_3 = 10 + 5\sqrt{2};$$

the eigenvectors of B are the same of A. ∎

2.4 Example. The matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

satisfies the equation

$$A^2 - 5A + 4I = 0. \tag{5}$$

So, if λ is an eigenvalue of A, $\lambda^2 - 5\lambda + 4$ is an eigenvalue of the zero matrix. Therefore the equation $\lambda^2 - 5\lambda + 4 = 0$ holds, and the eigenvalues of A are $\lambda_1 = 1$ e $\lambda_2 = 4$. Since one of these eigenvalues has multiplicity 2, and the trace of the matrix, which is equal to the sum of all eigenvalues, is 6, λ_1 must have multiplicity 2.

The equation (5) can be helpful for computing the inverse of A, because from

$$A(5I - A) = 4I$$

one obtains

$$A^{-1} = \frac{1}{4}(5I - A) = \frac{1}{4} \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$ ∎

More generally, the following theorem holds.

2.5 Theorem (Cayley-Hamilton). *Let $A \in \mathbf{C}^{n\times n}$ and let $P(\lambda)$ be its characteristic polynomial. Then*

$$P(A) = 0.$$

Proof. For $\lambda \in \mathbf{C}$ let $C = A - \lambda I$. Let B be the adjoint matrix of C, B satisfies the equation (see the exercise 1.48)

$$CB = (\det C)I. \tag{6}$$

The entries of B are determinants of submatrices of size $n-1$ of the matrix $A - \lambda I$, so they are polynomials in λ of degree at most $n-1$. The matrix B can be expressed in this way:

$$B = \lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \ldots + B_{n-1},$$

where B_j, $j = 0, 1, \ldots, n-1$ are n. From (6) we have

$$\begin{aligned}(\det C)I &= (A - \lambda I)(\lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \ldots + B_{n-1})\\ &= -\lambda^n B_0 + \lambda^{n-1}(AB_0 - B_1) + \lambda^{n-2}(AB_1 - B_2) + \ldots + AB_{n-1}.\end{aligned}$$

Moreover

$$\det C = \det(A - \lambda I) = P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n,$$

thus, by comparing the powers of λ with the same exponent, we have:

$$\begin{aligned} a_0 I &= -B_0\\ a_1 I &= AB_0 - B_1\\ a_2 I &= AB_1 - B_2\\ &\vdots\\ a_n I &= AB_{n-1}.\end{aligned}$$

After multiplying these equations by $A^n, A^{n-1}, \ldots, I$ and taking the sums of both left and right sides, we have finally

$$\begin{aligned}&a_0A^n + a_1A^{n-1} + a_2A^{n-2} + \ldots + a_nI\\ &\quad = -A^nB_0 + A^{n-1}(AB_0 - B_1) + A^{n-2}(AB_1 - B_2) + \ldots + AB_{n-1} = 0. \blacksquare\end{aligned}$$

As a consequence of the Cayley-Hamilton theorem, every matrix A annihilates its characteristic polynomial $P(\lambda)$, and any polynomial which is divided by $P(\lambda)$.

The *monic* polynomial (i.e. with leading coefficient 1) $\psi(\lambda)$ of minimal degree which is annihilated by A is called *minimal polynomial* of A, and is a factor of $P(\lambda)$ and of any polynomial $p(\lambda)$ which is annihilated by A as well. This can be easily be seen by dividing $p(\lambda)$ by $\psi(\lambda)$,

$$p(\lambda) = \psi(\lambda)s(\lambda) + r(\lambda),$$

where the degree of $r(\lambda)$ is smaller than the degree of $\psi(\lambda)$. Since

$$0 = p(A) = \psi(A)s(A) + r(A)$$

and $\psi(A) = 0$, then $r(A) = 0$. But $\psi(\lambda)$ is the polynomial of minimal degree vanishing in A, so $r(\lambda)$ must be identically zero.

Due to the fact that $\psi(\lambda)$ is a factor of $P(\lambda)$, the zeros of $\psi(\lambda)$ must be eigenvalues of A. Conversely, each eigenvalue of A is a zero of $\psi(\lambda)$, because, if μ is an eigenvalue of A, $\psi(\mu)$ is an eigenvalue of $\psi(A)$, but $\psi(A) = 0$, so $\psi(\mu) = 0$.

Now we know that $\psi(\lambda)$ has the following form:

$$\psi(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_p)^{n_p},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are all the distinct eigenvalues of A, and $n_1 + n_2 + \ldots + n_p \leq n$. If the matrix A has n distinct eigenvalues, then

$$P(\lambda) = (-1)^n \psi(\lambda).$$

2.6 Example. The matrix A introduced in the example 2.4 annihilates the polynomial

$$\psi(\lambda) = \lambda^2 - 5\lambda + 4,$$

which is its minimal polynomial, because no constant α exists such that $A + \alpha I = 0$, i.e. no polynomial of degree 1 vanishing in A exists. ■

3. Properties of the eigenvalues

2.7 Theorem. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Let $\lambda_1, \ldots, \lambda_m$, $m \leq n$, be m distinct eigenvalues of $A \in \mathbf{C}^{n \times n}$, and let $\mathbf{x}_1, \ldots, \mathbf{x}_m$ be corresponding eigenvectors. We proceed by induction on m.

For $m = 1, \mathbf{x}_1 \neq 0$, thus $\mathbf{x}_1$ is linearly independent.

For $m > 1$, by way of contradiction, we assume that a linear combination of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ exists such that

$$\sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \mathbf{0}, \tag{7}$$

where not all the α_i's are zero, that is $\alpha_j \neq 0$ for some j. In that case at least one index k exists, $k \neq j$, such that $\alpha_k \neq 0$, otherwise $\mathbf{x}_j = 0$. By multiplying both sides of (7) by A, we have

$$\mathbf{0} = A \sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, A\mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, \lambda_i \, \mathbf{x}_i, \tag{8}$$

and by multiplying both sides of (7) by λ_j, we have

$$\mathbf{0} = \lambda_j \sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, \lambda_j \, \mathbf{x}_i. \tag{9}$$

After subtracting side by side the equation (9) from the equation (8), we have finally:

$$\mathbf{0} = \sum_{i=1}^{m} \alpha_i (\lambda_i - \lambda_j) \mathbf{x}_i = \sum_{\substack{i=1 \\ i \neq j}}^{m} \alpha_i (\lambda_i - \lambda_j) \mathbf{x}_i.$$

But this is a zero linear combination of the $m-1$ eigenvectors $\mathbf{x}_i \neq \mathbf{0}$, $\quad i = 1, \ldots, m, \; i \neq j$, where $\lambda_i - \lambda_j \neq 0$ for $i \neq j$ and not all the α_i's, for $i \neq j$ are zero, since $\alpha_k \neq 0$: this is a contradiction because the $m-1$ vectors are linearly independent due to the inductive assumption. ■

It follows from theorem 2.7 that if a matrix A of order n has n distinct eigenvalues, then A has n linearly independent eigenvectors. If the matrix A has not n distinct eigenvalues, than A may or may not have n linearly independent eigenvectors. The following example shows the latter case.

2.8 Example. The matrix

$$A = \begin{bmatrix} 4 & 1 \\ -1 & 2 \end{bmatrix},$$

has characteristic polynomial $(\lambda - 3)^2$, and 3 is the unique eigenvalue, with multiplicity 2. Since all the corresponding eigenvalues have the form

$$\alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \alpha \neq 0,$$

the matrix A cannot have 2 linearly independent eigenvectors. ∎

Matrices with n distinct eigenvalues are not the only matrices with n linearly independent eigenvectors. For instance, the identity matrix I_n has eigenvalue 1 with multiplicity n, and the vectors $\mathbf{e}_i \in \mathbf{C}^n$, $i = 1, \ldots, n$, of the canonical basis of $\mathbf{C}^n$ can be chosen as eigenvectors.

In case several linearly independent eigenvectors correspond to the same eigenvalue, then they span a vector subspace: all the nonzero vectors in this subspace are eigenvectors corresponding to the same eigenvalue, as stated by the following theorem.

2.9 Theorem. *Let $A \in \mathbf{C}^{n\times n}$, and let $\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_k$ k be linearly independent eigenvectors, all corresponding to the same eigenvalue λ of A. Then a vector $\mathbf{y} \in \mathbf{C}^n$, $\mathbf{y} \neq \mathbf{0}$, with the form*

$$\mathbf{y} = \sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j$$

is an eigenvector of A.

Dim. We have

$$A\mathbf{y} = A\sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j = \sum_{j=1}^{k} \alpha_j A\mathbf{x}_j = \sum_{j=1}^{k} \alpha_j \lambda \mathbf{x}_j = \lambda \sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j = \lambda \, \mathbf{y}.$$ ∎

2.10 Definition. The multiplicity of an eigenvalue λ as a root of the characteristic equation, is denoted with $\sigma(\lambda)$, and is called *algebraic multiplicity* of λ. The maximum number of linearly independent eigenvectors corresponding to λ is denoted with $\tau(\lambda)$ and is called *geometric multiplicity* of λ. ∎

The geometric multiplicity $\tau(\lambda)$ is equal to the dimension of the vector subspace spanned by the eigenvectors corresponding to λ, that is the vector subspace

$$N(A - \lambda I) = \{ \ \mathbf{x} \in \mathbf{C}^n \ : \ (A - \lambda I)\mathbf{x} = \mathbf{0} \ \},$$

the null space of $A - \lambda I$. It is clear that

$$1 \le \sigma(\lambda) \le n \quad \text{and} \quad 1 \le \tau(\lambda) \le n.$$

2.11 Theorem. *The following inequality holds:*

$$\tau(\lambda) \leq \sigma(\lambda).$$

Proof. Let μ be an eigenvalue of A with algebraic multiplicity $\sigma = \sigma(\mu)$ and geometric multiplicity $\tau = \tau(\mu)$. From the relation (7) of Chapter 1 one derives

$$\text{rank of } (A - \mu I) = n - \ \dim N(A - \mu I) = n - \tau,$$

so all the principal submatrices with orders grater than $n - \tau$ of the matrix $A - \mu I$ are singular. Since the coefficient of the degree i term of the characteristic polynomial of a matrix is given, but for the sign, by the sum over the determinants of all the principal submatrices with size $n - i$, then the characteristic polynomial of $A - \mu I$ must have the form

$$p(\lambda) = \det[(A - \mu I) - \lambda I] = a_0 \lambda^n + a_1 \lambda^{n-1} + \ldots + a_k \lambda^{n-k},$$

where $k \leq n - \tau$. Therefore the equation $p(\lambda) = 0$ has the root $\lambda = 0$ with multiplicity $n - k \geq \tau$. If we set $x = \lambda + \mu$, we have

$$\begin{aligned} \det[(A - \mu I) - \lambda I] &= \det(A - xI) \\ &= a_0 (x - \mu)^n + a_1 (x - \mu)^{n-1} + \ldots + a_k (x - \mu)^{n-k}, \end{aligned}$$

thus the multiplicity of μ as a root of the characteristic equation is $\sigma \geq \tau$.

■

2.12 Example. The matrix $A \in \mathbf{R}^{n \times n}$, defined as follows

$$a_{ij} = \begin{cases} 1 & \text{per } j = i, \\ 1 & \text{per } j = i + 1, \\ 0 & \text{altrimenti}, \end{cases}$$

that is

$$A = \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \\ & & & & 1 \end{bmatrix},$$

has eigenvalue 1 with algebraic multiplicity n, and the corresponding eigenvalues are $\mathbf{x} = \alpha \mathbf{e}_1, \alpha \neq 0$. For this matrix we have therefore $\tau(1) = 1$ e $\sigma(1) = n$. ■

If we consider the identity matrix I_n, which has eigenvalue 1 with algebraic multiplicity n, we have $\tau(1) = \sigma(1) = n$.

4. Similarity transformations

Given a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ of $\mathbf{C}^n$ and a matrix $A \in \mathbf{C}^{n\times n}$, there a unique linear application $\mathcal{L}: \mathbf{C}^n \to \mathbf{C}^n$ exists, defined over the vectors of the basis as

$$\mathcal{L}(\mathbf{u}_j) = \sum_{i=1}^{n} a_{ij}\, \mathbf{u}_i. \tag{10}$$

The application $\mathcal{L}$ can be extended straightforward, by linearity, to all vectors $\mathbf{x} \in \mathbf{C}^n$. Let us consider the matrices U and W, whose columns are the vectors $\mathbf{u}_j$ and $\mathcal{L}(\mathbf{u}_j)$, respectively. Then the equation (10) can be represented in the form:

$$W = UA. \tag{11}$$

The same linear application $\mathcal{L}$ can be represented by referring two different bases $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$: two different matrices, A and B, would be obtained as representations. If V and Z are the matrices whose columns are the vectors $\mathbf{v}_j$ and $\mathcal{L}(\mathbf{v}_j)$ respectively, the following equation, analogue to (11), holds:

$$Z = VB. \tag{12}$$

Now we want to find the relation between the matrices A and B. Assume that the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, $i = 1, \ldots, n$, satisfy the relations

$$\mathbf{v}_j = \sum_{i=1}^{n} s_{ij}\, \mathbf{u}_i, \quad j = 1, 2, \ldots, n, \tag{13}$$

that, in a more compact way, can be rewritten as:

$$V = US,$$

where the nonsingular matrix S is the *change of basis* matrix. By substituting into (12) we have

$$Z = USB. \tag{14}$$

Moreover, since $\mathcal{L}$ is linear, from (13) we have:

$$\mathcal{L}(\mathbf{v}_j) = \mathcal{L}\,(\sum_{i=1}^{n} s_{ij}\, \mathbf{u}_i) = \sum_{i=1}^{n} s_{ij}\ \mathcal{L}(\mathbf{u}_i)$$

that is

$$Z = WS. \tag{15}$$

Substituting (11) into (15) we have:

$$Z = UAS$$

and, by comparing with (14), since U is nonsingular, we have:

$$AS = SB$$

and

$$A = SBS^{-1}.$$

If the two bases $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ e $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are orthonormal, the matrices U and V are unitary, so the matrix S is unitary as well, because

$$I = V^H V = (US)^H (US) = S^H U^H US = S^H S.$$

In this case the matrices A and B satisfy the relation

$$A = SBS^H.$$

2.13 Definition. Two matrices $A, B \in \mathbf{C}^{n\times n}$ are called *similar* if a nonsingular matrix S exists such that

$$A = SBS^{-1}.$$

The transformation from A into B is called *similarity transformation.* If the matrix S is unitary, the transformation is called *unitary similarity transformation.* ■

We remark that similarity is an equivalence relation, since it enjoys the properties of reflexivity, symmetry and transitivity.

Given a matrix A, let us consider the linear application $\mathcal{L}_A$ represented by A with respect to the canonical basis $\mathbf{e}_i,\ i = 1, \ldots, n$ di $\mathbf{C}^n$; then for all vectors $\mathbf{x} \in \mathbf{C}^n$ we have

$$\mathcal{L}_A(\mathbf{x}) = A\mathbf{x}.$$

Thus from (1), reformulated in terms of linear applications, we find that the eigenvectors $\mathbf{x}$ of A are the vectors which are transformed by the application $\mathcal{L}_A$ in vectors which are multiple of themselves. Thus every eigenvector lies on a straight line which is invariant under the linear application $\mathcal{L}_A$. Therefore the properties of the eigenvalues and of the eigenvectors are essential properties of the linear application, even if we derive them from the matrix which represents it with respect to a particular basis. This fact is expressed by the following theorem.

2.14 Theorem. *Two similar matrices have the same eigenvalues with the same algebraic and geometric multiplicities.*

Proof. Let A and B be similar matrices, that is $A = SBS^{-1}$ for some invertible matrix S. We have:

$$\begin{aligned}\det(A - \lambda I) &= \det(SBS^{-1} - \lambda SS^{-1}) = \det\left[S(B - \lambda I)S^{-1}\right] \\ &= \det S \det(B - \lambda I)\det(S^{-1}) = \det(B - \lambda I)\end{aligned}$$

Therefore both A and B have the same characteristic polynomial, the same eigenvalues with the same algebraic multiplicities. If $\mathbf{x}$ is an eigenvector of A corresponding to the eigenvalue λ, we have:

$$SBS^{-1}\mathbf{x} = \lambda\mathbf{x}$$

and

$$BS^{-1}\mathbf{x} = \lambda S^{-1}\mathbf{x}.$$

As a consequence, the vector $\mathbf{y} = S^{-1}\mathbf{x}$ is an eigenvector of B corresponding to λ. Moreover, since S^{-1} is nonsingular, if $\mathbf{x}_i,\ i = 1, \ldots, \tau(\lambda)$, are linearly independent eigenvectors of A, also $\mathbf{y}_i = S^{-1}\mathbf{x}_i,\ i = 1, \ldots, \tau(\lambda)$ are linearly independent. So A e B have the same eigenvalues with the same geometric multiplicities. ∎

The above theorem shows also that if two matrices are similar, they have the same trace and the same determinant.

2.15 Definition. A matrix A similar to a diagonal matrix D is called *diagonalizable.* ∎

2.16 Theorem. *A matrix A of order n is diagonalizable if and only if it has n linearly independent eigenvectors. Moreover the columns of the matrix S, such that $S^{-1}AS$ is diagonal, are the eigenvectors of A.*

Proof. First we assume that A has n linearly independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. Let D be the diagonal matrix with λ_i as i-th principal entry, and S the matrix whose i-th column is $\mathbf{x}_i$. From the equation

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i, \quad i = 1, 2, \ldots, n,$$

we have also that

$$AS = SD. \tag{16}$$

Since S is nonsingular, due to the linear independence of its columns, the inverse S^{-1} exists; thus from (16) we have

$$A = SDS^{-1}.$$

Conversely, let $A = SDS^{-1}$, where D is a diagonal matrix with the eigenvalues of A as principal entries. So we have also $AS = SD$. If we call $\mathbf{s}_1, \ldots, \mathbf{s}_n$ the columns of S, we have:

$$A\,[\mathbf{s}_1 \,|\, \mathbf{s}_2 \,|\, \ldots \,|\, \mathbf{s}_n] = [\lambda_1\mathbf{s}_1 \,|\, \lambda_2\mathbf{s}_2 \,|\, \ldots \,|\, \lambda_n\mathbf{s}_n]$$

and

$$A\mathbf{s}_i = \lambda_i\mathbf{s}_i, \quad i = 1, 2, \ldots, n.$$

Therefore the n columns of S are eigenvectors of A, which turn out to be linearly independent, since S is nonsingular. ∎

2.17 Example. The matrices A and $B = 3A^2 - A + 2I$ introduced in the example 2.3 have three distinct eigenvalues, are diagonalized by the same similarity transformation, since they share the same set of three linearly independent eigenvectors. If we set

$$S = \begin{bmatrix} 1 & 1 & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 1 & -1 & 1 \end{bmatrix},$$

we have

$$S^{-1} = \begin{bmatrix} 1/4 & -\sqrt{2}/4 & 1/4 \\ 1/2 & 0 & -1/2 \\ 1/4 & \sqrt{2}/4 & 1/4 \end{bmatrix}$$

and

$$A = S \begin{bmatrix} 1-\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+\sqrt{2} \end{bmatrix} S^{-1},$$

$$B = S \begin{bmatrix} 10-5\sqrt{2} & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 10+5\sqrt{2} \end{bmatrix} S^{-1},$$

Also the matrix A introduced in the example 2.4 has three linearly independent eigenvectors, although it does not have three distinct eigenvalues, therefore it is diagonalizable. If we set

$$S = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 1 \end{bmatrix},$$

we have

$$S^{-1} = \frac{1}{3}\begin{bmatrix} -1 & 2 & -1 \\ 2 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

and

$$A = S\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} S^{-1}. \qquad \blacksquare$$

5. Canonical forms

It follows, from theorems 2.7 e 2.16, that if a matrix has all distinct eigenvalues, then it is diagonalizable, because it has n linearly independent eigenvectors. If a matrix does not have all distinct eigenvalues, then it may not be diagonalizable and this happens when one, at least, eigenvalue of A has the geometric multiplicity smaller than the algebraic one. In this regard, the following theorem plays a fundamental role (for the proof see [3]).

2.18 Theorem (Jordan canonical or normal form). *Let $A \in \mathbf{C}^{n\times n}$ and $\lambda_i,\ i = 1,\ldots,p$, be its distinct eigenvalues, with algebraic and geometric multiplicities $\sigma(\lambda_i)$ and $\tau(\lambda_i)$ respectively. Then A is similar to a block diagonal matrix*

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix},$$

where the square block J_i, corresponding to the eigenvalue λ_i, has order $\sigma(\lambda_i)$, and is block diagonal itself:

$$J_i = \begin{bmatrix} C_i^{(1)} & & & \\ & C_i^{(2)} & & \\ & & \ddots & \\ & & & C_i^{(\tau(\lambda_i))} \end{bmatrix}, \quad i = 1, 2, \ldots, p,$$

where each one of the $\tau(\lambda_i)$ blocks has the form

$$C_i^{(j)} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbf{C}^{\nu_i^{(j)}\times\nu_i^{(j)}}, \quad j = 1, 2, \ldots, \tau(\lambda_i),$$

with the integers $\nu_i^{(j)}$ *such that*

$$\sum_{j=1}^{\tau(\lambda_i)} \nu_i^{(j)} = \sigma(\lambda_i).$$

The matrix J is called Jordan canonical (or normal) form of the matrix A. It is unique, up to the ordering of its blocks. ■

If the eigenvalues λ_i di A are all distinct, the blocks J_i are all of order one, therefore the matrix is diagonalizable. If, on the contrary, the eigenvalues are not all distinct but A has n linearly independent eigenvalues, then the blocks J_i are diagonal, and also in this case the matrix is diagonalizable.

2.19 Example. The knowledge of all the eigenvalues, with their multiplicities, is not sufficient to determine the structure of the Jordan canonical fornm, as this example shows. Both the matrices

$$A_1 = \begin{bmatrix} -4 & 7 & 4 & -8 & 6 & -3 \\ -5 & 7 & 5 & -8 & 6 & -3 \\ -4 & 4 & 6 & -7 & 6 & -3 \\ -3 & 3 & 3 & -4 & 6 & -3 \\ -2 & 2 & 2 & -4 & 6 & -2 \\ -1 & 1 & 1 & -2 & 2 & 1 \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} -4 & 12 & -10 & 8 & -6 & 4 \\ -5 & 12 & -9 & 8 & -6 & 4 \\ -4 & 8 & -6 & 8 & -6 & 4 \\ -3 & 6 & -16 & 8 & -5 & 4 \\ -2 & 4 & -4 & 4 & -2 & 4 \\ -1 & 2 & -2 & 2 & -2 & 4 \end{bmatrix}$$

have the eigenvalue $\lambda = 2$, with algebraic multiplicity 6 and geometric multiplicity 3. Their Jordan canonical forms are the following:

$$A_1 = SJ'S^{-1} = S \begin{bmatrix} 2 & 1 & 0 & & & \\ 0 & 2 & 1 & & & \\ 0 & 0 & 2 & & & \\ & & & 2 & 1 & \\ & & & 0 & 2 & \\ & & & & & 2 \end{bmatrix} S^{-1}$$

and

$$A_2 = SJ''S^{-1} = S \begin{bmatrix} 2 & 1 & & & & \\ 0 & 2 & & & & \\ & & 2 & 1 & & \\ & & 0 & 2 & & \\ & & & & 2 & 1 \\ & & & & 0 & 2 \end{bmatrix} S^{-1}.$$

In both cases

$$S = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 5 & 4 & 3 & 2 & 1 \\ 4 & 4 & 4 & 3 & 2 & 1 \\ 3 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \qquad S^{-1} = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}.$$

■

If the matrix A has real entries, then there exists a *Jordan real canonical form* of A, analogous to the one defined in 2.18, where the blocks $C_i^{(j)}$ corresponding to real eigenvalues have the same form described in theorem 2.18, while the blocks $C_i^{(j)}$ corresponding to nonreal eigenvalues are modified in this way: for every pair $\lambda_i = a_i + \mathbf{i}b_i$ and $\overline{\lambda}_i = a_i - \mathbf{i}b_i$ of conjugate complex eigenvalues of A the submatrices $C_i^{(j)}$ are block bidiagonal with the form

$$C_i^{(j)} = \begin{bmatrix} E_i & I_2 & & & \\ & E_i & I_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & I_2 \\ & & & & E_i \end{bmatrix},$$

where

$$E_i = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}.$$

2.20 Example. The matrix

$$A = \begin{bmatrix} 8 & -16 & 13 & -3 \\ 6 & -12 & 10 & -2 \\ 4 & -9 & 9 & -3 \\ 2 & -5 & 5 & -1 \end{bmatrix}$$

has the Jordan canonical form

$$A = SJS^{-1} = S \begin{bmatrix} 1+\mathbf{i} & 1 & 0 & 0 \\ 0 & 1+\mathbf{i} & 0 & 0 \\ 0 & 0 & 1-\mathbf{i} & 1 \\ 0 & 0 & 0 & 1-\mathbf{i} \end{bmatrix} S^{-1},$$

where

$$S = \begin{bmatrix} 4-3\mathbf{i} & 2-\mathbf{i} & 4+3\mathbf{i} & 2+\mathbf{i} \\ 3-3\mathbf{i} & 2-\mathbf{i} & 3+3\mathbf{i} & 2+\mathbf{i} \\ 2-2\mathbf{i} & 2-\mathbf{i} & 2+2\mathbf{i} & 2+\mathbf{i} \\ 1-\mathbf{i} & 1-\mathbf{i} & 1+\mathbf{i} & 1+\mathbf{i} \end{bmatrix}$$

and

$$S^{-1} = \frac{1}{2}\begin{bmatrix} 1-\mathbf{i} & -1+2\mathbf{i} & -\mathbf{i} & 0 \\ 0 & -1 & 2-\mathbf{i} & -1+2\mathbf{i} \\ 1+\mathbf{i} & -1-2\mathbf{i} & \mathbf{i} & 0 \\ 0 & -1 & 2+\mathbf{i} & -1-2\mathbf{i} \end{bmatrix}.$$

The matrix A has real entries, thus it can be represented by the Jordan real canonical form

$$A = ZJ_RZ^{-1} = Z\begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix}Z^{-1},$$

where

$$Z = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \qquad Z^{-1} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

■

The minimal polynomial of A can be obtained from the Jordan canonical form. In detail, if we set $A = SJS^{-1}$, where J is the Jordan real canonical form of A, for any integer k we have

$$A^k = \underbrace{SJS^{-1}SJS^{-1}\dots SJS^{-1}}_{k \text{ times}} = SJ^kS^{-1}$$

and, for any polynomial $P(A)$:

$$P(A) = SP(J)S^{-1}.$$

Thus, for the minimal polynomial $\psi(\lambda)$ of A, we have

$$\psi(A) = S\psi(J)S^{-1},$$

therefore the minimal polynomials of A and of J are the same. Due to the block diagonal structure of J we have also

$$\psi(J) = \begin{bmatrix} \psi(J_1) & & & \\ & \psi(J_2) & & \\ & & \ddots & \\ & & & \psi(J_p) \end{bmatrix},$$

and, as a consequence,

$$\psi(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \dots (\lambda - \lambda_p)^{n_p},$$

must have exponents n_i, $i = 1, 2, \dots, p$, such that $(\lambda - \lambda_i)^{n_i}$ is the minimal polynomial of J_i, that is the integers n_i must be the smallest ones which allows the polynomial $(\lambda - \lambda_i)^{n_i}$ to be simultaneously annihilated in all the submatrices $C_i^{(j)}$, $j = 1, 2, \dots, \tau(\lambda_i)$. This happens if and only if n_i is the largest among the orders of the submatrices $C_i^{(j)}$, for $j = 1, 2, \dots, \tau(\lambda_i)$.

2.21 Example. The minimal polynomial of the matrix A_1 introduced in the example 2.19 is

$$\psi(\lambda) = (\lambda - 2)^3,$$

while the minimal polynomial of the matrix A_2 of the same example is

$$\psi(\lambda) = (\lambda - 2)^2.$$

■

Among all the similarity transformations which relate the matrices B and $A = SBS^{-1}$, a relevant interest has to be given to those ones defined by a unitary S, i.e. the matrices satisfying the equations $S^H S = SS^H = I$. The following theorem shows how it is possible, by means of a unitary similarity, to transform any matrix into an upper triangular one.

2.22 Theorem (Schur canonical or normal form). *Let $A \in \mathbf{C}^{n\times n}$ and let $\lambda_1, \dots, \lambda_n$ be its eigenvalues. Then a unitary matrix U and an upper triangular matrix T exist such that*

$$A = UTU^H.$$

The principal entries of T are the eigenvalues λ_i.

Proof. We proceed by induction on the order n. For $n = 1$ the thesis holds, $T = [\lambda_1]$ and $U = [1]$ can be chosen. For $n > 1$, let $\mathbf{x}_1$ be a normalized eigenvector for the eigenvalue λ_1, and let S be the space spanned by $\mathbf{x}_1$. If we denote with $\mathbf{y}_2, \dots, \mathbf{y}_n$ the vectors of an orthonormal basis of the space $S^\perp$, the matrix

$$Q = [\,\mathbf{x}_1 \,|\, \mathbf{y}_2 \,|\, \dots \,|\, \mathbf{y}_n\,]$$

is unitary, and $Q^H \mathbf{x}_1 = \mathbf{e}_1$. Let us consider the matrix

$$B = Q^H A Q$$

whose first column is

$$B\mathbf{e}_1 = Q^H A Q \mathbf{e}_1 = Q^H A \mathbf{x}_1 = Q^H \lambda_1 \mathbf{x}_1 = \lambda_1 Q^H \mathbf{x}_1 = \lambda_1 \mathbf{e}_1$$

therefore B can be partitioned in the following way:

$$B = \begin{bmatrix} \lambda_1 & \mathbf{c}^H \\ \mathbf{0} & A_1 \end{bmatrix},$$

where $\mathbf{c} \in \mathbf{C}^{n-1}$ and $A_1 \in \mathbf{C}^{(n-1)\times(n-1)}$. By the inductive assumption a unitary matrix $U_1 \in \mathbf{C}^{(n-1)\times(n-1)}$ exists such that

$$A_1 = U_1 A_2 U_1^H,$$

where $A_2 \in \mathbf{C}^{(n-1)\times(n-1)}$ is upper triangular. Then we have

$$A = QBQ^H = Q \begin{bmatrix} \lambda_1 & \mathbf{c}^H \\ \mathbf{0} & A_1 \end{bmatrix} Q^H = Q \begin{bmatrix} \lambda_1 & \mathbf{c}^H \\ \mathbf{0} & U_1 A_2 U_1^H \end{bmatrix} Q^H.$$

If we denote with $U_2 \in \mathbf{C}^{n\times n}$ the unitary matrix

$$U_2 = \begin{bmatrix} 1 & \mathbf{0}^H \\ \mathbf{0} & U_1 \end{bmatrix},$$

we obtain:

$$A = QU_2 \begin{bmatrix} \lambda_1 & \mathbf{c}^H U_1 \\ \mathbf{0} & A_2 \end{bmatrix} U_2^H Q^H.$$

The matrix $U = QU_2$ is unitary as well, since it is the product of unitary matrices, so we have

$$A = U \begin{bmatrix} \lambda_1 & \mathbf{c}^H U_1 \\ \mathbf{0} & A_2 \end{bmatrix} U^H.$$

But A_2 is upper triangular, so the proof is concluded. ∎

2.23 Example. The matrix

$$A = \frac{1}{2}\begin{bmatrix} 5 & -5 & 1 & -1 \\ 5 & -5 & 3 & 1 \\ -1 & -1 & -1 & -1 \\ 3 & -1 & 1 & 1 \end{bmatrix}$$

has the eigenvalue $\lambda_1 = \mathbf{i}$ with the corresponding normalized eigenvector.

$$\mathbf{x}_1 = \frac{1}{2}\,[1\,,\,1\,,\,\mathbf{i}\,,\,-\mathbf{i}]^T.$$

Then three more vectors $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbf{C}^4$ can be considered, such that they complete $\mathbf{x}_1$ in a basis of $\mathbf{C}^4$:

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Then starting from the vectors $\mathbf{x}_i,\ i = 1, ..., 4$, by using the Gram-Schmidt orhogonalization, the following three vectors $\mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4$ are built:

$$\mathbf{y}_2 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ -\mathbf{i} \\ \mathbf{i} \end{bmatrix}, \quad \mathbf{y}_3 = \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{y}_4 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix},$$

so that the matrix

$$Q = [\mathbf{x}_1 \mid \mathbf{y}_2 \mid \mathbf{y}_3 \mid \mathbf{y}_4] = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ \mathbf{i} & -\mathbf{i} & 1 & 1 \\ -\mathbf{i} & \mathbf{i} & 1 & 1 \end{bmatrix}$$

is unitary. Then we have

$$B = Q^H A Q = \begin{bmatrix} \mathbf{i} & 0 & -2 & 3+\mathbf{i} \\ 0 & -\mathbf{i} & -2 & 3-\mathbf{i} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} T_1 & C \\ O & A_1 \end{bmatrix},$$

where $T_1 \in \mathbf{C}^{2\times 2}$ is upper triangular (more precisely in this case T_1 is diagonal). The procedure must be applied again to the matrix

$$A_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

which has the eigenvalue $\mathbf{i}$ too, with the normalized eigenvector

$$\mathbf{z}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \textbf{-i} \\ 1 \end{bmatrix}.$$

By using Gram-Schmidt the vector

$$\mathbf{z}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{i} \\ 1 \end{bmatrix}$$

is computed, so that the matrix

$$Q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \textbf{-i} & \mathbf{i} \\ 1 & 1 \end{bmatrix}$$

is unitary, and we have finally

$$B_1 = Q_1^H A_1 Q_1 = \begin{bmatrix} \mathbf{i} & 0 \\ 0 & \textbf{-i} \end{bmatrix}.$$

The Schur canonical form of A is therefore

$$A = QBQ^H = Q \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix} \begin{bmatrix} T_1 & CQ_1 \\ O & Q_1^H A_1 Q_1 \end{bmatrix} \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix}^H Q^H$$

$$= U \begin{bmatrix} \mathbf{i} & 0 & (3+3\mathbf{i})/\sqrt{2} & (3-\mathbf{i})/\sqrt{2} \\ 0 & \textbf{-i} & (3+\mathbf{i})/\sqrt{2} & (3-3\mathbf{i})/\sqrt{2} \\ 0 & 0 & \mathbf{i} & 0 \\ 0 & 0 & 0 & \textbf{-i} \end{bmatrix} U^H,$$

where U is the unitary matrix

$$U = Q \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & (1+\mathbf{i})/\sqrt{2} & (1-\mathbf{i})/\sqrt{2} \\ 1 & 1 & (\text{-}1-\mathbf{i})/\sqrt{2} & (\text{-}1+\mathbf{i})/\sqrt{2} \\ \mathbf{i} & \textbf{-i} & (1-\mathbf{i})/\sqrt{2} & (1+\mathbf{i})/\sqrt{2} \\ \textbf{-i} & \mathbf{i} & (1-\mathbf{i})/\sqrt{2} & (1+\mathbf{i})/\sqrt{2} \end{bmatrix}.$$

■

As in the case of the Jordan canonical form, also in the case of the Schur canonical form, if the matrix A is real the *Schur real canonical (normal) form.*

2.24 Theorem. *If $A \in \mathbf{R}^{n\times n}$, an orthogonal matrix $U \in \mathbf{R}^{n\times n}$ and a block upper triangular matrix $T \in \mathbf{R}^{n\times n}$ exist, such that $A = UTU^T$, where T has the form*

$$T = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix},$$

where the blocks R_{jj} per $j = 1, 2, \ldots, m$ have order 1 o 2. If λ_j is a real eigenvalue of A, then R_{jj} has order 1 and coincides with $[\lambda_j]$, if λ_j is not real, the the block R_{jj} has order 2 and eigenvalues λ_j and $\overline{\lambda}_j$. The sum of the orders of all the blocks R_{jj}, $j = 1, 2, \ldots, m$ is exactly n.

Proof. The proof is by induction, as already done for theorem 2.22. If the eigenvalue λ_1 is real, the same reasoning done for the complex case can be repeated. If $\lambda_1 = \mu_1 + \mathbf{i}\nu_1$, μ_1, $\nu_1 \in \mathbf{R}$, $\nu_1 \neq 0$, the corresponding eigenvector $\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1$, $\mathbf{x}_1$, $\mathbf{y}_1 \in \mathbf{R}^n$ is considered, where the vector $\mathbf{x}_1$ is assumed to be normalized. Since

$$A(\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1) = A\mathbf{x}_1 + \mathbf{i}A\mathbf{y}_1 = (\mu_1\mathbf{x}_1 - \nu_1\mathbf{y}_1) + \mathbf{i}(\mu_1\mathbf{y}_1 + \nu_1\mathbf{x}_1),$$

then

$$A[\,\mathbf{x}_1 \,|\, \mathbf{y}_1\,] = [\,\mathbf{x}_1 \,|\, \mathbf{y}_1\,] \begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix}. \tag{17}$$

The vectors $\mathbf{x}_1$ e $\mathbf{y}_1$ are linearly independent: if they were dependent, a constant $\alpha \neq 0$ would exist such that $\mathbf{y}_1 = \alpha\mathbf{x}_1$, and therefore

$$\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1 = \mathbf{x}_1 + \mathbf{i}\alpha\mathbf{x}_1 = (1 + \mathbf{i}\alpha)\,\mathbf{x}_1,$$

thus the real vector $\mathbf{x}_1$ would be a real eigenvector of A corresponding to the complex eigenvector λ_1, and this is not possible, because A is real.

The normalized vector $\mathbf{z}_1$ is computed, orthogonal to the vector $\mathbf{x}_1$, by setting

$$\mathbf{z}_1 = \beta\mathbf{x}_1 + \gamma\mathbf{y}_1, \quad \gamma = \frac{1}{\sqrt{\mathbf{y}_1^T\mathbf{y}_1 - (\mathbf{x}_1^T\mathbf{y}_1)^2}}, \quad \beta = -\gamma(\mathbf{x}_1^T\mathbf{y}_1).$$

Therefore

$$[\,\mathbf{x}_1 \,|\, \mathbf{z}_1\,] = [\,\mathbf{x}_1 \,|\, \mathbf{y}_1\,]\, W, \quad \text{dove } W = \begin{bmatrix} 1 & \beta \\ 0 & \gamma \end{bmatrix}. \tag{18}$$

Then an orthogonal matrix $Q \in \mathbf{R}^{n\times n}$ is computed, with $\mathbf{x}_1$ and $\mathbf{z}_1$ as the first two columns:

$$Q = [\,\mathbf{x}_1 \,|\, \mathbf{z}_1 \,|\, \mathbf{y}_3 \,|\, \ldots \,|\, \mathbf{y}_n\,].$$

The proof goes on in the same way as in the case of theorem 2.22. For the first two columns of the matrix $B = Q^T AQ$ we have from (17) e (18):

$$B\,[\,\mathbf{e}_1\,|\,\mathbf{e}_2\,] = Q^T A\,[\,\mathbf{x}_1\,|\,\mathbf{z}_1\,] = Q^T A\,[\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]W$$

$$= Q^T[\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]\begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix} W = Q^T[\,\mathbf{x}_1\,|\,\mathbf{z}_1\,]\,W^{-1}\begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix} W.$$

Since Q is orthognal, the first two columns of B can be written in the following way:

$$B\,[\,\mathbf{e}_1\,|\,\mathbf{e}_2\,] = \begin{bmatrix} I_2 \\ \\ O \end{bmatrix} W^{-1}\begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix} W = \begin{bmatrix} R_{11} \\ \\ O \end{bmatrix} \begin{matrix} \}\quad 2 \text{ righe} \\ \\ \}\ n-2 \text{ righe,} \end{matrix}$$

where the block

$$R_{11} = W^{-1}\begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix} W$$

is real and has eigenvalues λ_1 and $\overline{\lambda}_1$. The proof can be continued by using the inductive assumption as in the proof of theorem 2.22. ■

2.25 Example. We want to determine the Schur canonical form of the matrix

$$A = \frac{1}{2}\begin{bmatrix} 5 & -5 & 1 & -1 \\ 5 & -5 & 3 & 1 \\ -1 & -1 & -1 & -1 \\ 3 & -1 & 1 & 1 \end{bmatrix}$$

introduced in the example 2.23. The matrix A has the eigenvalue $\lambda_1 = \mathbf{i}$ with the corresponding eigenvector

$$\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ \mathbf{i} \\ -\mathbf{i} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathbf{i}\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

In this case the eigenvectors $\mathbf{x}_1$ e $\mathbf{y}_1$ are orthonormal. Then let us consider the two orthonormal vectors

$$\mathbf{y}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{y}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

The matrix

$$U = [\,\mathbf{x}_1\,|\,\mathbf{y}_1\,|\,\mathbf{y}_3\,|\,\mathbf{y}_4\,]$$

is orthogonal, and it is such that

$$A = U \begin{bmatrix} 0 & 1 & 5 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} U^T.$$

Thus the Schur real canonical form has been found. ■

A special and important case is given by the hermitian matrices.

2.26 Theorem. *Let A be a hermitian matrix of order n, that is $A = A^H$, and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Then a unitary matrix U exists such that*

$$A = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} U^H,$$

so the matrix A is diagonalizable. Moreover the eigenvalues λ_i, $i = 1, \ldots, n$ are real, and the columns of U are a set of orthonormal vectors.

Proof. By theorem 2.22 we have $T = U^H AU$, where T is an upper triangular matrix and U is unitary. Since $A = A^H$, we have

$$T^H = (U^H AU)^H = U^H A^H U = U^H AU = T,$$

so the triangular matrix T is diagonal with real principal entries and by theorem 2.16 the columns of U, which are orthonormal since U is unitary, are eigenvectors of A. ■

If the matrix A is real and symmetric, the matrix U is real too, therefore it is orthogonal.

2.27 Example. The matrix

$$A = \begin{bmatrix} 1 & \mathbf{i} & 0 \\ -\mathbf{i} & 2 & -\mathbf{i} \\ 0 & \mathbf{i} & 1 \end{bmatrix}$$

Has eigenvalues $\lambda_1 = 0$, $\lambda_2 = 1$ e $\lambda_3 = 3$, with corresponding eigenvectors

$$\mathbf{x}_1 = \alpha_1 \begin{bmatrix} 1 \\ \mathbf{i} \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \alpha_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \alpha_3 \begin{bmatrix} 1 \\ -2\mathbf{i} \\ 1 \end{bmatrix}, \quad \alpha_1,\ \alpha_2,\ \alpha_3 \neq 0,$$

which form a set of orthogonal vectors, and are also normalized if we set $\alpha_1 = 1/\sqrt{3}$, $\alpha_2 = 1/\sqrt{2}$ e $\alpha_3 = 1/\sqrt{6}$. Thus, in this case, the matrix

$$U = [\,\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \mathbf{x}_3\,],$$

that is

$$U = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \mathbf{i}\sqrt{2} & 0 & \text{-}2\mathbf{i} \\ \sqrt{2} & \text{-}\sqrt{3} & 1 \end{bmatrix},$$

is unitary, and the following relation holds:

$$A = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} U^H. \qquad \blacksquare$$

A wider class which includes, as special instances, hermitian and unitary matrices, is the class of normal matrices, which are those matrices satisfying the equation $A^H A = AA^H$. These matrices are particularly important, because they are all and only those matrices which can be diagonalized by unitary similarity transformarmations. Indeed, the following theorem holds.

2.28 Theorem. *A matrix $A \in \mathbf{C}^{n\times n}$ is normal, i.e. $A^H A = AA^H$, if and only if a unitary matrix U exists such that*

$$A = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} U^H,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. The columns of the matrix U are eigenvectors of the matrix A, so a set of n orthonormal eigenvectors exists.

Proof. Let us assume that A is normal. By theorem 2.22 A unitary matrix U exists such that

$$T = U^H AU,$$

where T is upper triangular. We have:

$$T^H T = U^H A^H UU^H AU = U^H A^H AU,$$
$$TT^H = U^H AUU^H A^H U = U^H AA^H U.$$

Since A is normal, it follows that

$$T^H T = TT^H, \tag{19}$$

so T is normal as well. Now we show, by induction on n, that T is diagonal. This is trivial when $n = 1$. If $n > 1$, since T is upper triangular, the entry p_{11} of the matrix $P = T^H T = T T^H$ can be written in the following two ways:

$$p_{11} = \bar{t}_{11} t_{11} = |\lambda_1|^2 \quad \text{and} \quad p_{11} = \sum_{j=1}^{n} t_{1j} \bar{t}_{1j} = |\lambda_1|^2 + \sum_{j=2}^{n} |t_{1j}|^2,$$

therefore

$$t_{1j} = 0, \quad \text{per} \quad j = 2, \ldots, n,$$

thus all the entries in the first row of T are zero, with the exception of t_{11}. If we denote with T_{n-1} the submatrix obtained from T by deleting the first row and the first column, from (19) it follows that

$$T_{n-1}^H T_{n-1} = T_{n-1} T_{n-1}^H.$$

By the induction assumption T_{n-1} is diagonal, so T is diagonal as well. Conversely, let A be diagonalizable by means of a unitary similarity:

$$A = U D U^H,$$

with D diagonal. We have:

$$A^H A = U D^H U^H U D U^H = U D^H D U^H,$$

$$A A^H = U D U^H U D^H U^H = U D D^H U^H.$$

Since D is diagonal, $D^H D$ and $D D^H$ are diagonal as well, and their principal entries are $\overline{\lambda}_i \lambda_i$ in both cases; therefore $D^H D = D D^H$ and finally

$$A^H A = A A^H. \qquad \blacksquare$$

When A is a real normal matrix, its Schur real canonical form is

$$A = U T U^T,$$

where T and U are real, U is orthogonal, and T is block diagonal, with blocks of order 1 or 2.

2.29 Example. The matrix $A \in \mathbf{R}^{4\times 4}$

$$A = \begin{bmatrix} 4 & -5 & 0 & 3 \\ 0 & 4 & -3 & -5 \\ 5 & -3 & 4 & 0 \\ 3 & 0 & 5 & 4 \end{bmatrix}$$

is normal, because $A^T A = AA^T$, so it is diagonalizable by means of a unitary similarity. By setting

$$U = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -\mathbf{i} & \mathbf{i} & 1 \\ 1 & -\mathbf{i} & \mathbf{i} & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

one has

$$A = U \begin{bmatrix} 12 & & & \\ & 1+5\mathbf{i} & & \\ & & 1-5\mathbf{i} & \\ & & & 2 \end{bmatrix} U^H.$$

A can be represented also in terms of its Schur real canonical form. Since

$$\begin{bmatrix} 1+5\mathbf{i} & 0 \\ 0 & 1-5\mathbf{i} \end{bmatrix} = V \begin{bmatrix} 1 & -5 \\ 5 & 1 \end{bmatrix} V^H,$$

where V is the unitary matrix

$$V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\mathbf{i} \\ 1 & \mathbf{i} \end{bmatrix},$$

one has

$$A = Z \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 1 & -5 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} Z^T,$$

where the orthogonal matrix Z has the following form:

$$Z = U \begin{bmatrix} 1 & \mathbf{0}^H & 0 \\ \mathbf{0} & V & \mathbf{0} \\ 0 & \mathbf{0}^H & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & \sqrt{2} & 0 & 1 \\ -1 & 0 & -\sqrt{2} & 1 \\ 1 & 0 & -\sqrt{2} & -1 \\ 1 & -\sqrt{2} & 0 & 1 \end{bmatrix}.$$

■

A consequence of the Schur theorem is a complete characterization of the eigenvalues of matrix polynomials. This result is expressed by the following theorem, whose proof is immediate.

2.30 Theorem. *Let $A = UTU^H$ be the Schur canonical form of the matrix A. If $p(x)$ is a polynomial in x, then $p(A) = Up(T)U^H$ and the eigenvalues of $p(A)$ are all and only those numbers $p(\lambda)$, where λ is an eigenvalue of A.*
■

Let $p(x)$ and $q(x)$ be two polynomials in x, such that $q(\lambda) \neq 0$ for each eigenvalue λ of A, and let us consider the rational function $f(x) = p(x)/q(x)$. By theorem 2.30, the matrix $q(A)$ is nonsingular, so it is possible to define

$$f(A) = [q(A)]^{-1}p(A).$$

For the matrix $f(A)$ the following property holds, which extends the result stated by theorem 2.30:

$$f(A) = Uf(T)U^H. \tag{20}$$

6. Some properties of positive definite matrices

2.31 Theorem. *Let A be a hermitian matrix of order n and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Then A is positive definite if and only if $\lambda_i > 0$, $i = 1, \ldots, n$.*

Proof. First we will show that if A is positive definite, then its eigenvalues are positive. Since A is hermitian, it has real eigenvalues. If λ is an eigenvalue and $\mathbf{x} \neq \mathbf{0}$ is a corresponding eigenvector, one obtains from the equation $A\mathbf{x} = \lambda\mathbf{x}$, by multiplying on the left by $\mathbf{x}^H$,

$$\mathbf{x}^H A\mathbf{x} = \lambda\mathbf{x}^H\mathbf{x}.$$

Since A is positive definite, the left hand side is positive, so, due to the condition $\mathbf{x}^H\mathbf{x} > 0$, one has $\lambda > 0$.

Viceversa, since A is hermitian, it can be expressed as $A = UDU^H$, with U unitary and D diagonal, having as principal entries the eigenvalues λ_i, $i = 1, \ldots n$, di A. If $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$, one has:

$$\mathbf{x}^H A\mathbf{x} = \mathbf{x}^H UDU^H\mathbf{x} = \mathbf{y}^H D\mathbf{y}, \tag{21}$$

where the vector $\mathbf{y} = U^H\mathbf{x}$ cannot be $\mathbf{0}$, because U is nonsingular. From (21):

$$\begin{aligned} \mathbf{x}^H A\mathbf{x} &= (\overline{y}_1, \ldots, \overline{y}_n) \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \\ &= \lambda_1\overline{y}_1 y_1 + ... + \lambda_n\overline{y}_n y_n = \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2 > 0 \end{aligned}$$

since all the eigenvalues λ_i are positive and at least one entry $|y_i|$ is nonzero. ■

The determinant of a matrix is the product of all its eigenvalues: from this property and from theorem 2.31 it follows that the determinant of a positive definite matrix is positive. Moreover, by theorem 2.31 we have that the inverse of a positive definite matrix is positive definite as well. More in detail, the inverse A^{-1} of a hermitian positive definite A is hermitian, and the eigenvalues of A^{-1} are positive because they are the inverses of the eigenvalues of A.

2.32 Example. The hermitian matrix

$$A = \begin{bmatrix} 1 & \mathbf{i} & 0 \\ -\mathbf{i} & 2 & -2\mathbf{i} \\ 0 & 2\mathbf{i} & 5 \end{bmatrix}$$

is positive definite, because for any vector $\mathbf{x} \neq \mathbf{0}$ one has

$$\mathbf{x}^H A\mathbf{x} = |x_1 + \mathbf{i}x_2|^2 + |x_2 - 2\mathbf{i}x_3|^2 + |x_3|^2 > 0,$$

and its characteristic polynomial is

$$P(\lambda) = -\lambda^3 + 8\lambda^2 - 12\lambda + 1. \tag{22}$$

Look at the graphic of $P(\lambda)$ in figure 2.1, and compute the values assumed by this polynomial in 1, 2, 7:

$$P(0) = 1,\ P(1) = -4,\ P(2) = 1,\ P(7) = -34,$$

thus it is clear that the polynomial has 3 real zeros in the intervals

$$(0, 1),\ (1, 2) \text{ e } (2, 7),$$

and this means that all the eigenvalues A are positive. ■

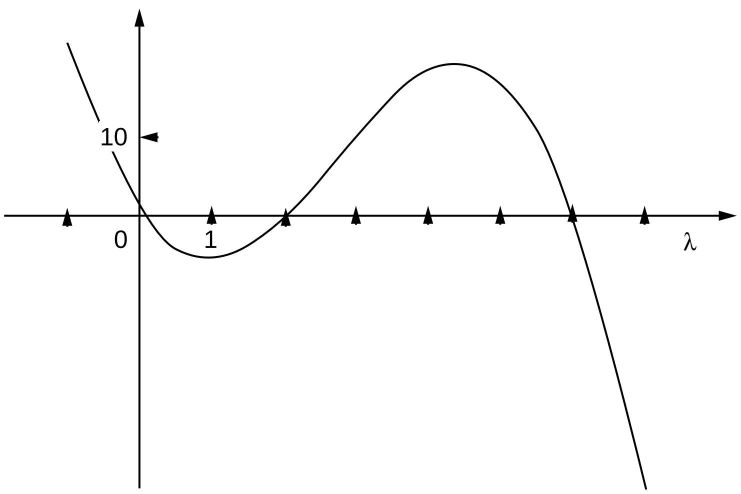

Fig. 2.1 - Graphic of the polynomial(22).

2.33 Theorem. *A hermitian matrix A is positive definite if and only if the determinants of all the leading principal submatrices of A (notice that also the determinant of A is included) are positive.*

Proof. If A is positive definite, then the thesis directly follows from theorem 1.14. Conversely, let us assume that the determinants of all the leading principal submatrices of A are positive and use induction on n. For $n = 1$ the result is obvious. For $n > 1$, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Since, by assumption, the product of the λ_i's is positive, it will be shown that an even number of negative eigenvalues cannot exist, and therefore all the λ_i's are positive. So let us assume, by way of contradiction, that m negative eigenvalues exist, with $m \geq 2$, m even (one can assume, without violating generality, that such eigenvalues are the first m). Let U be a unitary matrix such that $A = UDU^H$, where D is the diagonal matrix whose principal entries are the λ_i's, arranged in the required order. Then two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{C}^n$, can be found, such that

$$\mathbf{x}, \mathbf{y} \neq \mathbf{0},\ x_n = 0,\ y_{m+1} = y_{m+2} = \ldots = y_n = 0,\ \mathbf{y} = U^H \mathbf{x}.$$

In fact, by partitioning U^H and the vectors $\mathbf{x}, \mathbf{y}$ in the following way

$$U^H = \begin{bmatrix} V & \mathbf{v} \\ & \\ W & \mathbf{w} \end{bmatrix} \begin{matrix} \} \quad m \text{ rows} \\ \\ \} \ n - m \text{ rows} \end{matrix}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \\ 0 \end{bmatrix} \begin{matrix} \} \ n-1 \text{ entries} \\ \\ \} \quad 1 \text{ entry} \end{matrix} \qquad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \\ \mathbf{0} \end{bmatrix} \begin{matrix} \} \quad m \text{ entries} \\ \\ \} \ n-m \text{ entries} \end{matrix}$$

from the equation $\mathbf{y} = U^H \mathbf{x}$ one obtains

$$\begin{aligned} V\mathbf{x}_1 &= \mathbf{y}_1 \\ W\mathbf{x}_1 &= \mathbf{0}. \end{aligned}$$

Since in the matrix W the number $(n-1)$ of columns is larger than the number of rows $(n - m, m \geq 2)$, a zero linear combination, with some nonzero coefficient, of the columns of W exists. Consider the vector $\mathbf{x}_1 \neq \mathbf{0}$ made up of these coefficients, such that $W\mathbf{x}_1 = \mathbf{0}$ and $\mathbf{y}_1 = V\mathbf{x}_1 \neq \mathbf{0}$, otherwise the first $n - 1$ columns of U^H would be linearly dependent. For the quadratic form related to A the following equation follows

$$\mathbf{x}^H A \mathbf{x} = \mathbf{x}^H U D U^H \mathbf{x} = \mathbf{y}^H D \mathbf{y} = \sum_{i=1}^{n} \lambda_i \, |y_i|^2 = \sum_{i=1}^{m} \lambda_i \, |y_i|^2 < 0,$$

which gives a contradiction, because

$$\mathbf{x}^H A \mathbf{x} = \mathbf{x}_1^H A_{n-1} \mathbf{x}_1,$$

where $A_{n-1} \in \mathbf{C}^{(n-1)\times(n-1)}$ is the leading principal submatrix of order $n - 1$, which is positive definite by the inductive assumption. ∎

7. Eigenvalue localization

In this section three theorems will be given which allow to determine special subsets of the complex plane containing some or all the eigenvalues of a matrix.

2.34 Definition. Let $A \in \mathbf{C}^{n\times n}$. The circles of the complex plane

$$K_i = \{ \, z \in \mathbf{C} \; : \; |z - a_{ii}| \le \sum_{\substack{j=1 \\ j\ne i}}^{n} |a_{ij}| \, \}, \; i = 1, 2, \ldots, n,$$

with centers a_{ii} and radii $r_i = \sum\limits_{\substack{j=1 \\ j\ne i}}^{n} |a_{ij}|$ are called *Gershgorin circles*. The following theorem holds

2.35 Theorem (1st Gershgorin theorem). *All the eigenvalues of the matrix A of order n belong to the union*

$$\bigcup_{i=1,\ldots,n} K_i.$$

Proof. Let λ be an eigenvalue of A and $\mathbf{x}$ be a corresponding eigenvector, i.e.

$$A\mathbf{x} = \lambda\,\mathbf{x}, \;\; \mathbf{x} \ne \mathbf{0}.$$

Therefore one has:

$$\sum_{j=1}^{n} a_{ij}\, x_j = \lambda\, x_i, \;\; i = 1, \ldots, n,$$

and

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j\ne i}}^{n} a_{ij}\, x_j, \;\; i = 1, \ldots, n. \tag{23}$$

Let x_p be an entry of $\mathbf{x}$ with maximum modulus, that is

$$|x_p| = \max_{j=1,\ldots,n} |x_j| \ne 0, \tag{24}$$

and, by setting $i = p$ in (23), one has:

$$(\lambda - a_{pp})x_p = \sum_{\substack{j=1 \\ j\ne p}}^{n} a_{pj}\, x_j$$

and

$$|\lambda - a_{pp}|\,|x_p| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|\,|x_j|. \tag{25}$$

So, from (24),

$$|\lambda - a_{pp}|\,|x_p| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|\,|x_p|.$$

Finally after dividing both sides by $|x_p| > 0$, the following inequality is obtained:

$$|\lambda - a_{pp}| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|, \tag{26}$$

i.e. $\lambda \in K_p$. It is worthwhile to remark that, since the value of p is unknown a priori, it is only possible to localize λ in the union of all the circles K_i. ■

By applying the theorem above to the matrix A^T, which has the same eigenvalues as A, it results that the eigenvalues of A belong also to the union of the circles

$$H_i = \{\, z \in \mathbf{C} \;:\; |z - a_{ii}| \le \sum_{\substack{j=1\\ j\ne i}}^{n} |a_{ji}|\,\},\ \ i = 1, 2, \ldots, n,$$

thus the eigenvalues of A belong to the intersection

$$\Big(\bigcup_{i=1,\ldots,n} K_i\Big) \bigcap \Big(\bigcup_{i=1,\ldots,n} H_i\Big).$$

2.36 Example. Let us consider the matrix

$$A = \begin{bmatrix} 15 & -2 & 2 \\ 1 & 10 & -3 \\ -2 & 1 & 0 \end{bmatrix} \tag{27}$$

whose circles are the following

$$\begin{aligned} K_1 &= \{z \in \mathbf{C} \;:\; |z - 15| \le 4\}, \\ K_2 &= \{z \in \mathbf{C} \;:\; |z - 10| \le 4\}, \\ K_3 &= \{z \in \mathbf{C} \;:\; |z| \le 3\}, \end{aligned}$$

represented in fig. 2.2.

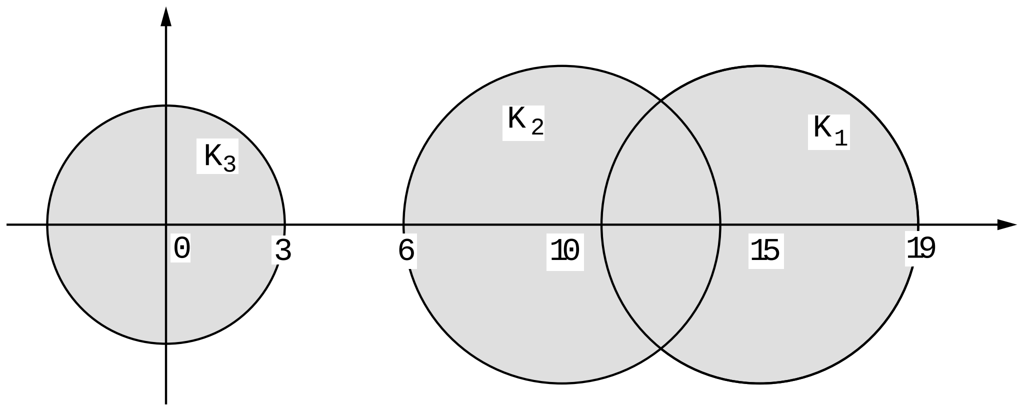

Fig. 2.2 - Gershgorin circles of the matrix A in (27).

The eigenvalues are in the grey zones. Now let us consider the circles

$$H_1 = \{z \in \mathbf{C} \ : \ |z - 15| \leq 3\},$$
$$H_2 = \{z \in \mathbf{C} \ : \ |z - 10| \leq 3\},$$
$$H_3 = \{z \in \mathbf{C} \ : \ |z| \leq 5\},$$

of the matrix A^T and represented in fig. 2.3.

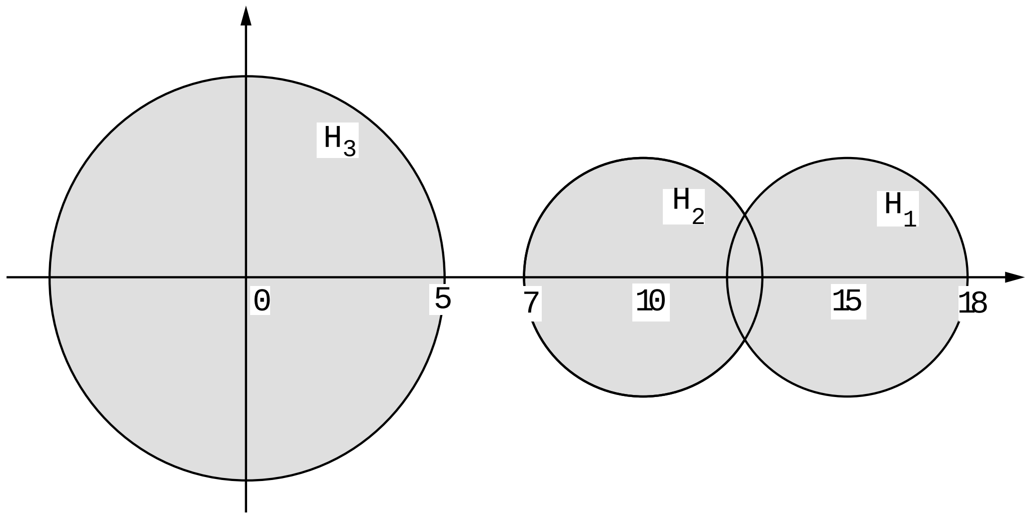

Fig. 2.3 - Gershgorin circles of the matrix A^T in (27).

Thus the eigenvalues of A belong to the intersection of the two unions of circles, represented in fig. 2.4. ∎

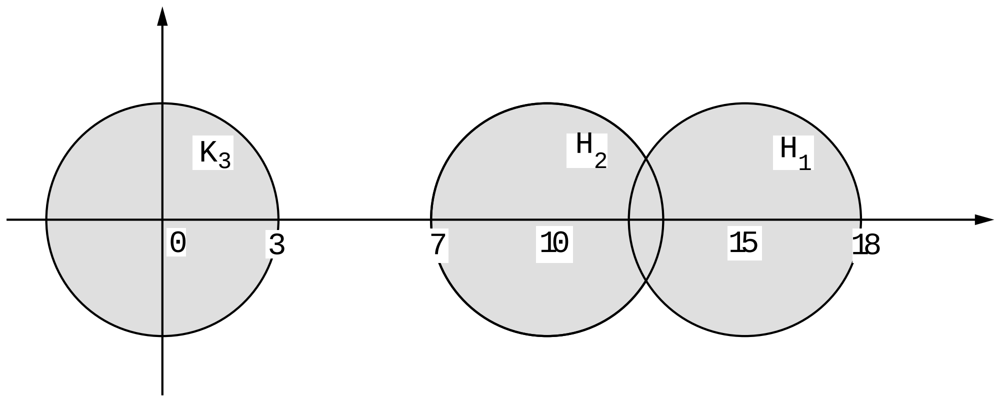

Fig. 2.4 - Intersection of the unions represented in figures 2.2 and 2.3.

2.37 Teorema (2nd Gershgorin theorem). *If the union M_1 of k Gershgorin circles is disjoint from the union M_2 of the $n-k$ circles left, then k eigenvalues belong to M_1 and $n-k$ eigenvalues belong to M_2.*

Proof. We can assume, without violating genarality, that the circles included in M_1 are the first k, that is

$$M_1 = \bigcup_{i=1,\dots,k} K_i \quad \text{e} \quad M_2 = \bigcup_{i=k+1,\dots,n} K_i.$$

Let D e R be the matrices defined as follows:

$$d_{ij} = \begin{cases} a_{ij} & \text{se } i = j, \\ 0 & \text{se } i \neq j, \end{cases} \qquad r_{ij} = \begin{cases} 0 & \text{se } i = j, \\ a_{ij} & \text{se } i \neq j, \end{cases}$$

so $A = D + R$. The matrix

$$A(t) = D + tR, \quad t \in [0, 1],$$

whose entries are continuous functions of t, has eigenvalues which are continuous functions of t, because they are the zeros of the characteristic polynomial whose coefficients are continuous functions of the entries of $A(t)$. In fact the zeros of a polynomial are continuous functions of its coefficients (see [5]). For any $t \in [0, 1]$ the first k Gershgorin circles of $A(t)$ are contained in M_1 because they have the same centers as the circles K_i, $1, \dots, k$, and radii increasing with t, and analogously the $n-k$ Gershgorin circles of $A(t)$ left are contained in M_2. Since the union of the first k Gershgorin circles of $A(t)$ is disjoint from the union of the Gershgorin circles left, if t varies continuously in the interval [0,1], the eigenvalues of $A(t)$ cannot pass from one set to another set which is disjoint from the first one. For $t = 0$, M_1 and

M_2 contain k and $n-k$ eigenvalues of $A(t)$ respectively, because $A(0) = D$ and the eigenvalues are the centers of the Gershgorin circles (which are the principal entries of A). Therefore for any $t \in [0,1]$, and in particular for $t = 1$ and $A(1) = A$, M_1 contains k eigenvalues and M_2 contains $n-k$ eigenvalues. ∎

Concerning the three eigenvalues of the matrix A in the example 2.36, one belongs K_3, while the other two are in $H_1 \cup H_2$. The two eigenvalues contained in $H_1 \cup H_2$ may be real or not, and have moduli in the interval 7, 18. The eigenvalue in K_3 is real, because if it had a nonzero imaginary part its conjugate should be eigenvalue of A too, being a zero of a polynomial with real coefficients.

For irreducible matrices another theorem can be stated which gives more information about the localization of the eigenvalues.

2.38 Theorem (3rd Gershgorin theorem). *Let the matrix A of order n be irreducible. If an eigenvalue of A lies on the boundary of each Gershgorin circle which contains it, then it lies on the boundaries of all Gershgorin circles. In particular the statement applies to the eigenvalues lying on the boundary of the union of all the circles.*

Proof. Let $\mathbf{x}$ an eigenvector corresponding to the eigenvalue λ, and let x_p be one of its entries with maximum modulus:

$$|x_p| = \max_{j=1,\ldots,n} |x_j|.$$

Reasoning in the same way as in the proof of theorem 2.35, one has that $\lambda \in K_p$. Since, by assumption, λ lies on the boundary of K_p, (26) must be an equality:

$$|\lambda - a_{pp}| = \sum_{\substack{j=1 \\ j \neq p}}^{n} |a_{pj}|.$$

Then it follows that also (25) must be an equality, so $|x_j| = |x_p|$, for all the indices j such that $a_{pj} \neq 0$. From the assumption of irreducibility, al least one index $r, r \neq p$, exists such that $a_{pr} \neq 0$, and since

$$|x_r| = \max_{j=1,\ldots,n} |x_j|,$$

the same argument used for the index p applies to the index r. In this way one concludes that λ lies on the boundary of K_r and moreover that $|x_j| = |x_r|$, for all the indices j such that $a_{rj} \neq 0$. This conclusion can be extended to all the indices, as the irreducibility of A ensures the existence of a directed path connecting each pair of nodes of the directed graph related to the matrix A. ∎

2.39 Example. The matrix

$$F = \begin{bmatrix} 0 & \cdot & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{bmatrix}$$

is called *Frobenius matrix.* If one computes $\det(F - \lambda I)$ by means of the Laplace expansion applied to the last row, the following result is found

$$\det(F - \lambda I) = (-1)^n \Big(\lambda^n + \sum_{i=0}^{n-1} a_i \, \lambda^i \Big).$$

Moreover the minimal polynomial concides, but for a sign factor, with the characeristic polynomial. In fact, by way of contradiction, if the minimal polynomial were

$$\psi(\lambda) = \lambda^k + \alpha_0 \lambda^{k-1} + \ldots + \alpha_{k-1}, \quad \text{con } k < n,$$

then, by multiplying $\psi(F)$ by the first vector of the canonical basis $\mathbf{e}_1$, one would find

$$\begin{aligned} \psi(F)\mathbf{e}_1 &= F^k \mathbf{e}_1 + \alpha_0 F^{k-1} \mathbf{e}_1 + \ldots + \alpha_{k-1} \mathbf{e}_1 \\ &= \mathbf{e}_{k+1} + \alpha_0 \mathbf{e}_k + \ldots + \alpha_{k-1} \mathbf{e}_1 \end{aligned}$$

therefore $\psi(F)\mathbf{e}_1$ would be the vector with the following first entries

$$\alpha_{k-1}, \ldots, \alpha_0, 1,$$

and this is not possible since $\psi(F) = 0$.

The 1rst Gershgorin theorem, when applied to F and F^T, allows to find for the zeros λ_i of the polynomial

$$\lambda^n + \sum_{i=0}^{n-1} a_i \, \lambda^i$$

the following bounds:

$$\begin{aligned} |\lambda_i| &\le \max \{ \, |a_0|, 1 + |a_1|, \ldots, 1 + |a_{n-1}| \, \} \\ |\lambda_i| &\le \max \{ \, 1, \ \sum_{i=0}^{n-1} |a_i| \, \}. \end{aligned}$$

■

8. Diagonal dominance

The matrices with diagonal dominance are an important instance of matrices often involved in the numerical resolution of differential problems.

2.40 Definitions. A matrix $A \in \mathbf{C}^{n\times n}$ is called *diagonally dominant* if, for each $i = 1, \ldots, n$,

$$|a_{ii}| \geq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|,$$

and at least an index s exists such that

$$|a_{ss}| > \sum_{\substack{j=1\\ j\neq s}}^{n} |a_{sj}|. \tag{28}$$

A matrix $A \in \mathbf{C}^{n\times n}$ is called *strictly diagonally dominant* if for each $i = 1, \ldots, n$,

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}} |a_{ij}|.$$

The two definitions of diagonal dominance and of strict diagonal dominance can be given with regard to the columns, when sums are made along columns instead rows, and in this case one says that the *diagonal dominance (strict diagonal dominance) is by columns.* ∎

2.41 Theorem. *If $A \in \mathbf{C}^{n\times n}$ is a strictly diagonally dominant, or a diagonally dominant and irreducible matrix, then A is nonsingular. Moreover, if A has all real and positive principal entries, then the eigenvalues of A have positive real parts; if A is also hermitian, then A is positive definite.*

Proof. If A strictly diagonally dominant, from theorem 2.35 one has that the Gershgorin circles have radii smaller than the distances of their centers from the origin of the complex plane, thus none of them can include the origin, and therefore A cannot have 0 as eigenvalue.

If A is diagonal dominant (not stricly) and irreducible, then the origin lies on the boundary of some Gershgorin circle. If 0 were eigenvalue of A, then from theorem 2.38 the origin should lie on the boundaries of all the Gershgorin circles, but this is not possible for one circle at least, as a consequence of (28).

Moreover, if A is strictly diagonally dominant or diagonally dominant and irreducible, and if all the principal entries of A are positive, no circle can contain complex numbers with negative real part. Therefore if A is also hermitian, its eigenvalues are real and positive, and from theorem 2.31 the matrix turns out to be positive definite. ∎

Since the eigenvalues of the matrix A are the same of A^T, the statements of theorem 2.41 hold also when diagonal dominance (strict diagonal dominance) is by columns.