Linear Algebra (not only) for computer scientists

Translation Luděk Kučera

Milan Hladík

24. května 2017

Chapter 10 Eigenvalues

Eigenvalues (also called "characteristic numbers") are, similarly as determinants, certain characterization of a matrix. Unlike determinants, their significance is very important.

Definition 10.1 (Eigenvalues and eigenvectors) Let $A \in \mathbb{C}^{n\times n}$. Then $\lambda \in \mathbb{C}$ is an *eigenvalue* of A and $x \in \mathbb{C}^n$ is the corresponding *eigenvector*, if $Ax = \lambda x$, $x \neq o$.

Let us note that $x \neq o$ is necessary condition, otherwise every $\lambda \in \mathbb{C}$ would be an eigenvalue. On the other hand, $\lambda = 0$ is possible and allowed.

Let us also note that an eigenvector is not unique - any multiple by a nonzero number is an eigenvector as well. This is why eigenvectors are sometimes normalized so that $\|x\| = 1$.

Of course, eigenvalues can be defined over any field, but we will stay with $\mathbb{R}$ and $\mathbb{C}$. We will se that complex numbers can not be avoided even if the matrix A is real.

Eigenvalues can be defined in a more general way. Let V be a vector space and $f : V \mapsto V$ is a linear mapping. Then λ is an eigenvalue and $x \neq o$ is an eigenvector, if $f(x) = \lambda x$. However we will mostly deal with eigenvalues of matrices, because, in view of matrix representation of linear mappings, the problem of eigenvalues of linear mappings can be reduced to matrices. [1]

Example 10.2 (Geometric interpretation of eigenvalues and eigenvectors) An eigenvector repreesents an invariant direction of the mapping $x \mapsto Ax$. In other words, if v is an eigenvector, then the line $\text{span}\{v\}$ maps into itself. An eigenvalue represents scaling in this direction.

- Flipping along the line $y = -x$, the matrix of the mapping is $A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$

  Figure, eigenvector 1, eigenvector $(-1, 1)^T$

[1] Remark of L.K.: However, it is my strong feeling that, in order to *understand* what we are doing with eigenvalues and eigenvectors, we have to view them as eigenvalues and eigenvectors of *linear mappings*.

Figure, eigenvector -1, eigenvector $(1,1)^T$

- (0^o rotation, the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$

  Figure, no real eigenvalues

Theorem 10.3 (Characterization of eigenvalues and eigenvectors) Let $A \in \mathbb{C}^{n\times n}$. Then
(1) $\lambda \in \mathbb{C}$ is an eigenvalue if and only if $\det(A - \lambda I_n) = 0$.
(2) $x \in \mathbb{C}^n$ is an eigenvector corresponding to an eigenvalue $\lambda \in \mathbb{C}$ if and only if $o \neq x \in \mathrm{Ker}(A - \lambda I_n)$.
Proof.

(1) $\lambda \in \mathbb{C}$ is an eigenvalue of A if and only if $(A - \lambda I_n)x = o$ $x \neq o$, which is equivalent to singularity ot the matrix $A - \lambda I_n$, which in turn is equivalent to $\det(A - \lambda I_n) = 0$.

(2) Analogically, $x \in \mathbb{C}^n$ is an eigenvector corresponding to an eigenvalue $\lambda \in \mathbb{C}$ if and only if $(A - \lambda I_n)x = o$, $x \neq o$, i.e., x is in the kernel of $A - \lambda I_n$.

A consequence of the theorem is that an eigenvalue λ is associated with $\dim \mathrm{Ker}(A - \lambda I_n) = n - \mathrm{rank}(A - \lambda I_n)$ linearly independent eigenvectors.

10.1 Characteristic polynomial

Definition 10.4 (Characteristic polynomial) A *characteristic polynomial* of a matrix $a \in \mathbb{C}^{n\times n}$ with respect to a variable λ is $p_A(\lambda) = \det(A - \lambda I_n)$.

It follows from the definition of a determinant that the characteristic polynomial can be expressed as

$$p_A(\lambda) = \det(A - \lambda I_n) = (-1)^n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

Hence it is really a polynomial and its degree is n. It can be seen that $a_{n-1} = (-1)^{n-1}(a_{11} + \cdots + a_{nn})$ and after a substitution $\lambda = 0$ we get $a_0 = \det(A)$.

Acording to the Fundamental Theorem of Algebra (Theorem 1.1) the polynomial has n complex roots (including multiplicities).

Theorem 10.5 Eigenvalues of a matrix $A \in \mathbb{C}^{n\times n}$ are exactly the roots of its characteristic polynomial $p_A(\lambda)$ and their number is n (including multiplicities).

Definition 10.6 (Algebraic and geometrical multiplicity of an eigenvalue). Let $\lambda \in \mathbb{C}$ be an eigenvalue of a matrix $A \in \mathbb{C}^{n\times n}$. *Algebraic multiplicity* of λ is equal to the multiplicity of λ as a root pf $p_A(\lambda)$. *Geometric multiplicity* of λ is equal to $n - \mathrm{rank}(A - \lambda I_n)$, i.e., to the number of linearly independent eigenvectors sorresponding to λ.

Algebraic multiplicity is always greater or equal to the geometric multiplicity, see section 10.4. In the sequal, the tem multiplicity will denote the algebraic multiplicity.

Definition 10.7 (Spectrum and spectral radius). Let $A \in \mathbb{C}^{n\times n}$has eigenvalues $\lambda_1, \ldots, \lambda_n$. Then the *spectrum* of the matrix A is the set of its eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ and the spectral radius is $\rho(A) = \max_{i=1,\ldots,n} |\lambda_i|$.

Computing eigenvalues as roots of the characteristic polynomial is not very effective. Moreover, there is neither formula nor an algorithm for roots of a polynomial and the roots are computed by iterative methods. The same is true for eigenvalues (see Theorem 10.16). Nevertheless, for certain special matrices eigenvalues could be determined easily.

Example 10.8 (Eigenvalues of a triangular matrix)

- Let $A \in \mathbb{C}^{n\times n}$ be a triangular matrix. Then its eigenvalues are elements of the diagonal, because $\det(A - \lambda I_n) = (a_{11} - \lambda)\cdots(a_{nn} - \lambda)$.
- Especially, I_n has the eigenvalue 1, which has multiplicity n. The set of eigenvectors is the set of all non-zero vectors.
- Especially, $)_n$ has the eigenvalue 0, which has multiplicity n. The set of eigenvectors is the set of all non-zero vectors.
- Especially, the matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has an eigenvalue 1, which has the algebraical multiplicity 2. The only eigenvector (normalized to the length 1) is $(1,0)^T$, and hence the geometrical multiplicity of the eigenvalue 1 is one.[2]

Example 10.9 Consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ from Example 10.2. Then

$$p_A(\lambda) = \det(A - \lambda I_n) = \det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1.$$

The roots of a polynomial, end hence the eigenvalues of A, are $+i$ and $-i$ with the corresponding eigenvectors $(1,-i)^T$ and $(1,i)^T$.

Example 10.10 Consider the matrix $A = \begin{pmatrix} 1.5 & 0.75 \\ 0 & 1 \end{pmatrix}$. The corresponding linear transformation $x \mapsto Ax$ geometrically represents skewing and elongation in the x direction by 50%. See the figure.

Figure of a bridge

The eigenvalues of the matrix A are 1.5 and 1, and the corresponding eigenvectors are $(1,0)^T$ and $(0,1)^T$. The first eigenvalue and eigenvector say that the picture is elongated by 50 % in the x direction, the second ones say that the figure is not deformed in the direction y.

Recall that trace(A) is the trace of the metrix A, i.e., the sum of its diagonal elements.

Theorem 10.11 (The sum and product of eigenvalues) Let $A \in \mathbb{C}^{n\times n}$ be a matrix with eigenalues $\lambda_1, \ldots, \lambda_n$. Then
(1) $\det(A) = \lambda_1 \cdots \lambda_n$,

[2]Remark L.K. This matrix is the distillation of the worst possible matrix behavior from the point of view of the spectral theory, as we will see later.

(2) $\text{trace}(A) = \lambda_1 + \cdots + \lambda_n$,

Proof.
(1) We know that $\det(A - \lambda I_n) = (-1)^n(\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$. Put $\lambda = 0$ to get

$$\det(A) = (-1)^n(-\lambda_1)\cdots(-\lambda_n) = \lambda_1 \cdots \lambda_n.$$

(2) Let us compare coefficients of λ^{n-1} in the different representations of the characteristic polynomial of A. In the formula for $\det(A - \lambda I_n$ we get that the coefficient is obtained from the product $(a_{11} - \lambda)\cdots(a_{nn} - \lambda)$ and its value is $(-1)^{n-1}(a_{11} + \cdots + a_{nn}$. The coefficient of λ^{n-1} in the expression $(-1)^n(\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$ is obviously $(-1)^n(-\lambda_1 - \cdots - \lambda_n$. Comparing two expressions we get $\det(A) = (-1)^n(-\lambda_1)\cdots(-\lambda_n) = \lambda_1 \cdots \lambda_n$. ♣

Let us note that comparing other coefficients of the characteristic polynomial we can get similar, but more complicated relations of elements and eigenvalues of the matrix A.

Theorem 10.12 (Properties of eigenvalues) Let $A \in \mathbb{C}^{n\times n}$ be a matrix with eigenalues $\lambda_1, \ldots, \lambda_n$ and the corresponding eigenvectors $x_1, \ldots, x_n$. Then
(1) A is regular if and only it 0 is not its eigenvalue,
(2) if a is regular, then A^{-1} has eigenalues $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ and the corresponding eigenvectors $x_1, \ldots, x_n$.
(3) A^2 has eigenalues $\lambda_1^2, \ldots, \lambda_n^2$ and the corresponding eigenvectors $x_1, \ldots, x_n$,
(4) αA has eigenvalues $\alpha\lambda_1, \ldots, \alpha\lambda_n$ and the corresponding eigenvectors $x_1, \ldots, x_n$,
(5) $A+\alpha I_n$ has eigenalues $\lambda_1+\alpha, \ldots, \lambda_n+\alpha$ and the corresponding eigenvectors $x_1, \ldots, x_n$,
(6) A^T has eigenalues $\lambda_1, \ldots, \lambda_n$ but, in general, different eigenvectors.

Proof. We will prove the first two statements, others are left to the reader.
(1) A has the eigenvector 0 if and only if $0 = \det(A - 0I_n) = \det(A)$, i.e., if it is singular.
(2) It is $Ax_i = \lambda_i x_i$ for each $i = 1, \ldots, n$.Multiplying by A^{-1} we get $x_i = \lambda_i A^{-1}x_i$ and by dividing by $\lambda_i \neq 0$ we get $\lambda^{-1}x_i = A^{-1}x_i$.

The assumtion that the metrix A is real is substantial in the following theorem, that is not generally valid for a complex matrix.

Theorem 10.13 If $\lambda \in \mathbb{C}$ is an eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$, than the complex conjugate $\overline{\lambda}$ is an eigenvalue of A as well.

Proof. We know that λ is the root of $p_A(\lambda) = (-1)^n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots = a_1\lambda + a_0 = 0$. The complex conjugate of the equation is $p_A(\lambda) = (-1)^n\overline{\lambda}^n + a_{n-1}\overline{\lambda}^{n-1} + \cdots = a_1\overline{\lambda} + a_0 = 0$, and therefore $\overline{\lambda}$ is also a root of $p_A(\lambda)$.

Example 10.14 The spectrum of a real matrix is symmetric by the real axis. A complex matrix can have an arbitrary set of n complex numbers as its spectrum.

Figure Spectrum of a real matrix.

Definition 10.15 (Companion matrix) Let $p(x) = x^n + a_{n-1} + \cdots + a_1x + a_0$ be a polynomial. The companion matrix of the polynomial $p(x)$ is an $n \times n$ matrix defined as follows (see the Czech text):

$$C(p) = \begin{pmatrix} 0 & \dots & \dots & 0 & -a_0 \\ 1 & \dots & \dots & \dots & -a_1 \\ 0 & \dots & \dots & \dots & -a_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 & -a_{n-1} \end{pmatrix}$$

Theorem 10.16 (On companion matrix) The characteristic polynomial of $C(p)$ verifies $p_{C(p)}(\lambda) = (-1)^n p(\lambda)$, and hence the eigenvalues of $C(p)$ correspond to the roots of the polynomial $p(\lambda)$.

Proof.

$$p_{C(p)}(\lambda) = \det(C(p) - \lambda I_n) = \det \begin{pmatrix} -\lambda & \dots & \dots & 0 & -a_0 \\ 1 & \dots & \dots & \dots & -a_1 \\ 0 & \dots & \dots & \dots & -a_2 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & -\lambda & -a_{n-2} \\ 0 & \dots & \dots & 1 & -a_{n-1} \end{pmatrix}$$

Add λ-times the last row to the previous one, then λ-times the second last row to the previous one, etc. We ger

$$p_{C(p)}(\lambda) = \det \begin{pmatrix} -\lambda & \dots & \dots & 0 & -p(\lambda) \\ 1 & \dots & \dots & \dots & -a_1 \\ 0 & \dots & \dots & \dots & -a_2 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & -\lambda & -a_{n-2} - a_{n-1}\lambda - \lambda^2 \\ 0 & \dots & \dots & 1 & -a_{n-1} \end{pmatrix}$$

The Laplacian evaluation by the first row gives
$p_{C(p)}(\lambda) = (-1)^{n+1}(-p(\lambda)) \det(I_{n-1}) = (-1)^n p(\lambda)$. ♣

The consequence of the theorem is that computing roots of real polynomials and eigenvalues of matrices are mutually reducible problems. Theorem 10.5 reduces eigenvalue search to computing roots of a polynomial, and Theorem 10.16 is the converse. This means that there is no finite algorithm to compute eigenvalues that can be evaluated only numerically.

10.2 Cayley-Hamilton theorem

Skipped

10.3 Diagonalization

Solving of systems of linear equations using Gauss-Jordan elimination used elementary transformations. They do not change the set of solutions, and can be used to modify a matrix to a form that makes it easy to get the solution. It is therefore natural to look for similar transformations of a matrix that do not change the spectrum. Elementary transformations do change the spectrum. A

convenient transformation is so called *silimarity*, because it does not chnge the spectrum. And if, using similarity, a matrix is transformed to a diagonal form, we won / the diagonal is composed of eigenvalues.

Definition 10.20 (Similarity) Matrices $A, B \in \mathbb{C}^{n\times n}$ are *similar*, if there is a regular matrix $S \in \mathbb{C}^{n\times n}$ such that $A = SBS^{-1}$.

Insert

The following text is an insert, which is not contained in the original text, where similarity is introduced without explaining *why* similarity preserves the spectrum and how it was found. I'll try to explain the underlying idea that naturally leads to similarity.

The text explains eigenvalues and eigenvectors as properties of matrices. I am strongly convinced that the true understanding of the spectral theory requires viewing eigenvalues and eigenvectors as properties of *linear mappings*.

Imagine we have a linear mapping $f : V \mapsto V$ of a vector space V into itself. In order to work with an abstract linear mapping f, we will represent it by a matrix. To do this, we have to choose some basis $\mathbf{B} = \mathbf{b}_1, \ldots, \mathbf{b}_n$ of V. Each vector $\mathbf{v} \in V$ can be uniquely represented as a linear combination $\mathbf{v} = \alpha_1\mathbf{b}_1 + \cdots + \alpha_n\mathbf{b}_n$. The n-tuple $(\alpha_1, \ldots, \alpha_n)^T$ will be denoted $\mathbf{v}|_\mathbf{B}$, the coordinates of $\mathbf{v}$ with respect to $\mathbf{B}$. The symbol ^{T} denotes a "transposition" and means that we will consider the n-tuple $\mathbf{v}|_\mathbf{B}$ as a *column* vector.

For each $i = 1, \ldots, n$, the vector $f(\mathbf{b}_i)$ of the basis $\mathbf{B}$ can be uniquely written as a linear combination of the vectors of the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, i.e.,

$$f(\mathbf{b}_i) = \alpha_{i,1}\mathbf{b}_1 + \cdots + \alpha_{i,n}\mathbf{b}_n = \sum_{j=1}^{n} \alpha_{i,j}\mathbf{b}_j. \qquad (A)$$

Let us define a matrix A by $A_{i,j} = \alpha_{i,j}$, where $\alpha_{i,j}$ are coefficients from the previous paragraph.

I will prove that $\boxed{f(\mathbf{v})|_\mathbf{B} = A\ \mathbf{v}|_\mathbf{B}}$, i.e., in order to get coordinates of $f(\mathbf{v})$ with respect to the basis $\mathbf{B}$ (considered as a column vector), we take $\mathbf{v}|_\mathbf{B}$, the coordinates of $\mathbf{v}$ with respect to the basis $\mathbf{B}$ and multiply the vector by the matrix A.

In fact, $\mathbf{v}|_\mathbf{B} = (\beta_1, \ldots, \beta_n)^T$ means that $\mathbf{v} = \beta_1\mathbf{b}_1 + \cdots + \beta_n\mathbf{b}_n$. Now, in view of linearity of f,

$$f(\mathbf{v}) = f\left(\sum_{i=1}^{n} \beta_i\mathbf{b}_i\right) = \sum_{i=1}^{n} \beta_i f(\mathbf{b}_i) = \sum_{i=1}^{n} \beta_i \sum_{j=1}^{n} \alpha_{i,j}\mathbf{b}_j =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i\alpha_{i,j}\mathbf{b}_j = \sum_{j=1}^{n}\sum_{i=1}^{n} \beta_i\alpha_{i,j}\mathbf{b}_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \alpha_{i,j}\beta_i\right)\mathbf{b}_j, \qquad (B)$$

which means that the coordinates of $f(\mathbf{v})$ with respect to the base $\mathbf{B}$ are exactly those that we get when we multiply the (column) coordinates or $\mathbf{v}$ with respect to $\mathbf{B}$ by the matrix A.

However, the representation of the mapping f by a matrix *depends* on a choice of the basis $\mathbf{B}$. How the matrix representing f changes when we choose another basis of the space V, that we will denote by $\mathbf{C}$ and that is composed of vectors $\mathbf{c}_1, \ldots, \mathbf{c}_n$.

Each vector $\mathbf{c}_i$ is a linear combination of the vector of the basis $\mathbf{B}$. If we denote j-th element of the coordinates of $\mathbf{c}_i$ by $\sigma_{i,j}$, then

$$\mathbf{c}_i = \sum_{j=1}^{n} \sigma_{i,j} \mathbf{b}_j \quad \text{for } i = 1, \ldots, n. \tag{C}$$

The numbers $\sigma_{i,j}$ determine a matrix S such that $S_{i,j} = \sigma_{i,j}$. The matrix S is very useful: suppose that, for a given vector $\mathbf{v} \in V$ we know its coordinates with respect to the basis $\mathbf{C}$, but we want to know its coordinates with respect to the basis $\mathbf{B}$. Assume that $\mathbf{v}|_{\mathbf{C}} = (\alpha_1, \ldots, \alpha_n)^T$. This implies that

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{c}_i. \tag{B}$$

Now,

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{c}_i = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \sigma_{i,j} \mathbf{b}_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \sigma_{i,j} \mathbf{b}_j =$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_i \sigma_{i,j} \mathbf{b}_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \sigma_{i,j} \alpha_i \right) \mathbf{b}_j, \tag{D}$$

which means that the coordinates of $\mathbf{v}$ with respect to the basis $\mathbf{B}$ we obtain by multiplying the coordinates of $\mathbf{v}$ with respect to $\mathbf{C}$ by the matrix S, i.e. $\boxed{\mathbf{v}|_{\mathbf{B}} = S\ \mathbf{v}|_{\mathbf{C}}}$. We immediately get also that $\mathbf{v}|_{\mathbf{C}} = S^{-1}\ \mathbf{v}|_{\mathbf{B}}$.

And now, we arrive to the main observation of this insert: Assume that we know the matrix representation of a linear mapping $f : V \mapsto V$ with respect to the basis $\mathbf{B}$ and we want to know the matrix representation of f with respect to $\mathbf{C}$. Let us do the following: $\mathbf{v}|_{\mathbf{C}}$, the coordinates of $\mathbf{V}$ with respect to $\mathbf{C}$, are transformed by thier multiplying by the matrix S to the coordinates of $\mathbf{v}$ with respect to $\mathbf{B}$, i.e., $\mathbf{v}|_{\mathbf{B}} = S\ \mathbf{v}|_{\mathbf{C}}$. Then, by multiplying $\mathbf{v}|_{\mathbf{B}}$ by the matrix A, representing the mapping f with respect to $\mathbf{B}$, we get the coorinates of $f(\mathbf{v})$ with respect to $\mathbf{B}$, which are subsequently transformed to the coordinates of $f(\mathbf{v})$ with respect to $\mathbf{C}$ by multiplying by the matrix S^{-1}. In other words,

$$f(\mathbf{v})|_{\mathbf{B}} = A\ \mathbf{v}|_{\mathbf{B}} \tag{E}$$

implies that

$$f(\mathbf{v})|_{\mathbf{C}} = S^{-1}\ f(\mathbf{v})|_{\mathbf{B}} = S^{-1}\ A\ \mathbf{v}|_{\mathbf{B}} = S^{-1}\ A\ S\ \mathbf{v}|_{\mathbf{C}} = (S^{-1}AS)\ \mathbf{v}|_{\mathbf{C}}. \tag{F}$$

On the ther hand, given a basis $\mathbf{B}$ represented by vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$ and a matrix S, define the basis $\mathbf{C}$ represented by vectors $\mathbf{c}_1, \ldots, \mathbf{c}_n$ as by the formula

(C). Then, in the same way as above, it can be proved that the linear mapping, represented by the matrix A with respect to the basis $\mathbf{B}$ is the same as the linear mapping, represented by the matrix $S^{-1}AS$ with respect to the basis $\mathbf{C}$.

What was said above shows that two matrices are similar if and only if they represent the same linear mapping $f : V \mapsto V$ with respect to two basis of a vector space V.

If $\mathbf{B}$ is a basis of a vector space V, $f : V \mapsto V$ is a linear mapping, and a matrix A represents f with respect to $\mathbf{B}$, then λ is an eigenvalue of the mapping f if and only if λ is an eigenvalue of A and the corresponding eigenvector of A is $\mathbf{v}|_{\mathbf{B}}$, where $\mathbf{v}$ is the corresponding eigenvector of f:

$$f(\mathbf{v}) = \lambda\mathbf{v} \quad \text{if and only if} \quad A\ \mathbf{v}|_{\mathbf{B}} = f(\mathbf{v})|_{\mathbf{B}} = \lambda\mathbf{v}|_{\mathbf{B}}$$

It is therefore not surprising that similar matrices have the same eigenvalues and there is a simple correspondence of their eigenvectors - they are representations of the same abstract linear mapping (with respect to differen basis).

End of insert

Example 10.21 Matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

are similar using $S = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Theorem 10.22 (Eigenvalues of similar matrices) Similar matrices have the same eigenvalues.

Proof. In view of similarity, there is a regular matrix S such that $A = SBS^{-1}$. Then

$$p_A(\lambda) = \det(A - \lambda I_n) = \det(SBS^{-1} - \lambda S I_n S^{-1}) = \det(S(B - \lambda I_n)S^{-1}) =$$

$$= \det(S)\det(B - \lambda I_n)\det(S^{-1}) = \det(B - \lambda I_n) = p_B(\lambda). \quad \clubsuit$$

Let us note that the theorem says nothing about eigenvectors, they can be different. What is unchanged is their number.

Insert

The relationship between eigenvectors of two similar matrices is in fact very simple.

Theorem 10.22bis (Eigenvectors of similar matrices) If $A = SBS^{-1}$, where S is a regular matrix, and v is and eigenvector of A corresponding to an eiganvalue λ, then $S^{-1}v$ is an eigenvector of B, corresponding to an eiganvalue λ.

Proof. If $Av = \lambda v$, then $SBS^{-1}v = \lambda v$, i.e. $BS^{-1}v = \lambda S^{-1}v$. $\clubsuit$

End of insert

Exercise 10.23 Show that similarity is a reflexive, symmetric and transitive binary relation. This implies that the similarity is an equivalence.

Theorem 10.24 Let matrices $A, B \in \mathbb{C}^{n\times n}$ are similar, and λ is their eigenvalue. Then the number of eigenvectors corresponding to λ is the same for both matrices.

Proof. (Different than in the Czech text) If $A = SBS^{-1}$, then S represents an isomorphism (one-to-one linear mapping) of the vector space into itself that maps eigenvectors of B corresponding to λ onto eigenvectors of A corresponding to λ, see Theorem 10.22bis. ♣

Definition 10.25 (Diagonalizability) A matrix A is *diagonalizable*, if it is similar to some diagonal matrix.

Example 10.26 There are matrices that are not diagonalizable, e.g.

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

is not diagonalizable. It has an eigenvalue 0, which has the multiplicity 2. If A were diagonalizable, it would have been similar to the null matrix, i.e., $A = S0S^{-1} = 0$, a contradiction. ♣

Theorem 10.27 (Characterization of diagonalizability). A matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors.

Proof. Implication "⇒": A diagonal matrix has n eigenvectors $e_1, \ldots, e_n$ [3] that are linearly independent. A regular matrix S that brings similarity of A and a diagonal matrix give an isomorphism of eigenvectors of the diagonal matrix and eigenvectors of A, and hence A has n linearly independent eigenvectors as well.

Implication "⇐": Let A has eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding linearly independent eigenvectors $x_1, \ldots, x_n$ that are regarded as column vectors. Define a regular matrix $S = (x_1 \mid \cdots \mid x_n)$ and a diagonal matrix $\Lambda = \mathrm{diag})\lambda_1, \ldots, \lambda_n)$. Then

$$AS_{*j} = AS_{*j} = Ax_j = \lambda_j x_j = \Lambda_{jj} S_{*j} = S(\Lambda_{jj} e_j) = S\Lambda_{*j} = (S\Lambda)_{*j}.$$

This implies that $AS = S\Lambda$, since the vectors $x_1, \ldots, x_n$ (that are columns of S) are linearly independent, and hence S is regular, and therefore $A = S\Lambda S^{-1}$. ♣

The proof of the theorem is constructive, it shows how to diagonalize a matrix if we know its eigenvalues and eigenvectors. Similarly the knowledge of the diagonalization formula $A = S\Lambda S^{-1}$ it is easy to determine eigenvectors

[3] e_j is a column vector that has the j-th element equal to 1 and the remainimg elements equal to 0.

(the eigenvectors are colums of S in the order of eigenvectors at the diagonal of Λ).

Another view of diagonalization is geometric: we know that an eigenvector represents an invariant direction of a linear mapping $x \mapsto Ax$. Imagine that A is a matrix of a linear mapping $f : \mathbb{C}^n \mapsto \mathbb{C}^n$ with respect to a basis B. Diagonalization is a search of a basis of the vector space such that the matrix representing f with respect to that basis is diagonal.

Example 10.28 (Geometric interpretation of diagonalization). Let

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Eigenvalues and eigenvectors of the matrix A are:

$$\lambda_1 = 4 \quad x_1 = (1,1)^T, \qquad \lambda_2 = 2, \quad x_2 = (-1,1)^T.$$

The diagonalization has form:

$$A = S\Lambda S^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{pmatrix}.$$

Geometric interpretation: in the coordination system of eigenvectors of the matrix the matrix of the mapping is diagonal and the mapping is just scaling along axes. I the picture belowyou can see the unit square and its image by the mapping $x \mapsto Ax$. On the left, the situation, when we work in the coordination system of the canonical basis, on the right, the situation in the soordination system of the basis composed of the eigenvectors.

Figure at the page 126 of the Czech text

Now, we will show that eigenvectors corresponding to different eigenvalues are linearly independent.

Theorem 10.29 (Eigenvectors of different eigenvalues) Let $\lambda_1, \ldots, \lambda_k$ be mutually different eigenvalues (not necessarily all eigenvalues) of a matrix $A \in \mathbb{C}^{n \times n}$. Then the corresponding eigenvectors $x_1, \ldots, x_k$ are linearly independent.

Proof. By mathematical induction on k. The theore is obvious for $k = 1$, because an eigenvector is not a zero vector.

Induction step $k \leftarrow k - 1$. Consider a linear combination

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = o. \tag{10.1}$$

Multiplying by the matrix A gives

$$A(\alpha_1 x_1 + \cdots + \alpha_k x_k) = \alpha_1 A(x_1) + \cdots + \alpha_k A(x_k) = \alpha_1 \lambda_1 x_1 + \cdots + \alpha_k \lambda_k x_k = o. \tag{10.2}$$

Subtracting the λ_k-times (10.1) from (10.2) gives

$$\alpha_1(\lambda_1 - \lambda_k)x_1 + \cdots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = o.$$

The induction hypothesis says that $x_1, \ldots, x_{k-1}$ are linearly independent, which implies $\alpha_1 = \cdots = \alpha_{k-1} = 0$. A substitution into (10.1) gives $\alpha_k x_k = o$, which implies $\alpha_k = 0$. ♣

Consequence 10.30 If a matrix $A \in \mathbb{C}^{n\times n}$ has n different eigenvalues, then it is diagonalizable.

Now, we will show several theoretical and practical applications of diagonalization.

Example 10.31 (The power of a matrix). Let $A = S\Lambda S^{-1}$ be a diagonalization of the matrix $A \in \mathbb{C}^{n\times n}$. Then

$$A^2 = S\Lambda S^{-1} S\Lambda S^{-1} = S\Lambda^2 S^{-1}.$$

In a more general way we get

$$A^k = S\Lambda^k S^{-1} = S \begin{pmatrix} \lambda_1^k & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n^k \end{pmatrix} S^{-1}.$$

We can even study asymptotic behavior. Simplified:

$$\lim_{k\to\infty} A^k = S \begin{pmatrix} \lim_{k\to\infty} \lambda_1^k & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lim_{k\to\infty} \lambda_n^k \end{pmatrix} S^{-1} =$$

$$= \begin{cases} 0, & \text{if} \rho(A) < 1, \\ \text{diverges}, & \text{if} \rho(A) > 1, \\ & \\ & \text{converges/diverges}, \text{if} \rho(A) = 1, \end{cases}$$

In general, the case $\rho(A) = 1$ can not be resolved: for $A = I_n$ the powers of A converge to I_n, while for $A = -I_n$ the powers oscillate between I_n and $-I_n$.

It is again useful to follow the geometric view of the mapping $x \mapsto Ax$. The powers of the matrix A correspond to the composition of the matrix A with itself. If absolute values of all eigenvalues are smaller than 1 (i.e., $\rho(A) < 1$), then the linear mapping contracts distances and then it converges to zero for $k \to \infty$. If at least one eigenvalue is larger than 1, then the linear mapping extends distances in the direction of the corresponding eigenvector, and hence it diverges for $k \to \infty$. The case $\rho(A) = 1$ is the most interesting. This case occurs, e.g., for orthogonal matrices or for matrices of Markov chains, that will be dealt with in Example 10.51.

Theorem 10.32 Let $A, B \in \mathbb{C}^{n\times n}$. Then the matrices AB and BA have the same eigenvalues, including their multiplicities.

Proof. The matrices

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \quad \text{resp.} \quad \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

are block triangular, and therefore thay have the same eigenvalues as AB, resp. BA, plus an eigenvalue 0 with multiplicity n. Now, it is sufficient to show that the above block matrices have the same spectrum. The matrices are similar by the matrix

$$S = \begin{pmatrix} I & A \\ 0 & I \end{pmatrix},$$

because

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix} = \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} \qquad \clubsuit$$

The preceding theorem is valid for rectangular matrices $A, B^T \in \mathbb{R}^{m\times n}$, but only for non-zero eigenvalues; the multiplicity of zero eigenvectors can be (and typically are) different.

Example 10.33 (Recurrent formulae and Fibonacci). Let us consider a sequence $a_1, a_2, \ldots$, given by a recurrent relation

$$a_n = pa_{n-1} + qa_{n-2},$$

where a_1, a_2 are given first two values of the sequence, and p, q are constants. Let us show, how to "break" the recursion and to give an explicit form of the n-th element of the sequence. The same method works even for more complicated recursions, where a_n depends on more than two previous elements.

Let us express the recursion in a matrix form:

$$\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix}.$$

If we denote

$$x_n = \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}, \qquad A = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix},$$

the recursion has the form

$$x_n = Ax)n - 1 = A^2 x_{n-2} = \cdots = A^{n-2} x_2.$$

Therefore we need to determine higher powers of the matrix A. We can use the method of Example 10.31. Find a diagonal form $A = S\Lambda S^{-1}$, and then $x_n = S\lambda^{n-1}S^{-1}x_2$. Now, we are finished, the explicit form of $A - N$ is hidden in the first element of the vector $x_n = S\Lambda^{n-1}S^{-1}x_2$.

In particular, let us show this method for Fibonacci sequence, which is given by $a_n = a_{n-1}+a_{n-2}$, and the first two elements of the sequence are $a_1 = a_2 = 1$. Denote $\varphi = \frac{1}{2}(1+\sqrt{5})$, which is the golden ratio. Now

$$x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = S\Lambda S^{-1},$$

where

$$S = \begin{pmatrix} -1 & \varphi \\ \varphi & 1 \end{pmatrix}, \quad S^{-1} = \frac{\sqrt{5}}{5}\begin{pmatrix} 1-\varphi & 1 \\ 1 & \varphi-1 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 1-\varphi & 0 \\ 0 & \varphi \end{pmatrix}.$$

Therefore

$$a_n = S_{1*}\Lambda^{n-2}S^{-1}x_2 = \frac{5-3\sqrt{5}}{10}(1-\varphi)^{n-2} + \frac{5+3\sqrt{5}}{10}\varphi^{n-2},$$

which can easily be rewritten to more usual form

$$a_n = -\frac{\sqrt{5}}{5}(1-\varphi)^n + \frac{\sqrt{5}}{5}\varphi^n. \qquad \clubsuit$$

10.4 Jordan normal form

The simplest form of a matrix that can be obtained by elementary row transformations, is a reduced echelon form. But what is the simplest form of a matrix that can be obtained by similarity? This form is not a diagonal form, because we already know that not all matrices are diagonalizable. Nevertheless, every matrix can be transform by similarity to a relatively simple form, which is called *Jordan normal form* [4].

Definition 10.34 (Jordan cell (or block)). Let $\lambda \in \mathbb{C}$, $k \in \mathbb{N}$. The *Jordan cell* $J_k(\lambda)$ is a square matrix of order k defined as follows:

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}.$$

The Jordan cell has an eigenvalue λ, which has multiplicity k, but it has only one corresponding eigenvector, because the matrix $J_k(\lambda) - \lambda I_k$ has the rank $k-1$.

[4] The author is a French mathematician Marie Ennemond Camille Jordan. However, a co-author of Gauss-Jordan elimination is someone else, a German geodet Wilhelm Jordan.

Definition 10.35 (Jordan normal form). A matrix $J \in \mathbb{C}^{n\times n}$ is in the *Jordan normal form*, if it is in a block-diagonal form

$$J = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{k_m}(\lambda_m) \end{pmatrix}.$$

and the blocks on the diagonal are Jordan cells $J_{k_1}(\lambda_1), \ldots, J_{k_m}(\lambda_m)$.

The values λ_i and k_i need not be different. Similarly certain Jordan cell can apperar many times. [5]

Theorem 10.36 (On Jordan normal form). Every matrix $J \in \mathbb{C}^{n\times n}$ is similar to a matrix in the Jordan normal form. Up to the order of the Jordan cells, this matrix is uniquely determined.

Proof.

For a full proof, see Bečvář 2005, Bican 2009, Horn and Johnson 1985. Here we will say at least some ideas of the proof and the construction of a Jordan normal form. We are looking for a basis, with respect to it the mapping $x \mapsto Ax$ has the Jordan normal form. We will show how Jordan cells and the corresponding part of the basis look for an eigenvalue λ. Without loss of generality, we suppose that $\lambda = 0$, otherwise we will use the matrix $A - \lambda I_n$.

Let us consider the following subspaces of $\mathbb{C}^n$:

$$\mathrm{Ker}(A) \subsetneqq \mathrm{Ker}(A^2) \subsetneqq \cdots \subsetneqq \mathrm{Ker}(A^p) = \mathrm{Ker}(A^{p+1}) = \cdots$$

Let $v \in \mathrm{Ker}(A^p) - \mathrm{Ker}(A^{p-1})$ and consider the basis B represented by vectors $A^{p-1}v, \ldots, Av, v$. In the subspace generated by such vectors and with respect to the basis B, the linear mapping $x \mapsto Ax$ has the matrix that is the Jordan cell $J_p(0)$.

Similarly we will get the remaining Jordan cells, but we have to generalize slightly the method . Instead of a vector v, we will take a system $v_1, \ldots, v_\ell$ of the size $\dim \mathrm{Ker}(A^p) - \dim \mathrm{Ker}(A^{p-1})$ that completes some basis of $\mathrm{Ker}(A^{p-1})$ to a basis of $\mathrm{Ker}(A^p)$. Every such vector v_i is a source of a chain of vectors $A^{p-1}v_i, \ldots, Av_i, v_i$ that corresponds to the Jordan cell $J_p(0)$; the number of such cells is ℓ. if$\ell' = \dim \mathrm{Ker}(A^{p-1}) - \dim \mathrm{Ker}(A^{p-2}) - \ell > 0$, we have to complete vectors $Av_1, \ldots, Av_\ell$ by new vectors $v_{\ell+1}, \ldots, v_{\ell+\ell'}$ to complete some basis of $\mathrm{Ker}(A^{p-2})$ to a basis of $\mathrm{Ker}(A^{p-1})$. Such new vectors v_j become sources of chains $A^{p-2}v_j, \ldots, Av_j, v_j$ that correspond to ℓ' Jordan cells $J_{p-1}(0)$. The procedure is repeated to the dimension 1.♣

[5] Remark L.K.: The identity matrix I_n is in the Jordan normal form, the diagonal consists of n identical blocks equal to the matrix (1) of the rank 1.

Example 10.37 A matrix

$$\begin{pmatrix} 5 & -2 & 2 & -2 & 0 \\ 0 & 6 & -1 & 3 & 2 \\ 2 & 2 & 7 & -2 & -2 \\ 2 & 3 & 1 & 2 & -4 \\ -2 & -2 & -2 & 6 & 11 \end{pmatrix}$$

has eigenvalues 5 (the multiplicity 2) and 7 (the multiplicity 3). Since $3 = \mathrm{rank}(A - 5I_5) = \mathrm{rank}(A - 5I_5)^2$, we will look for two chains of the length !. We will find two linearly independent vectors $x_1, x_2 \in \mathrm{rank}(A - 5I_5)$, e.g., $x_1 = (-2, 1, 1, 0, 0)^T$ and $x_2 = (-1, 1, 0, -1, 1)^T$ that form the first two vectors of the basis we are looking for.

Now, let us consider the eigenvector 7. We have $\mathrm{rank}(A - 7I_5) = 3$ and $3 = \mathrm{rank}(A - 7I_5)^2 = \mathrm{rank}(A - 7I_5)^3 = 2$. We will choose $x_4 \in \mathrm{Ker}(A - 7I_5)^2 - \mathrm{Ker}(A - 5I_5)$, e.g., $x_4 = (1, 0, 1, 0, 0)^T$ and the corresponding part of the basis is formed by the chain $x_3 = (A - 7I_5)x_4 = (0, -1, 2, 3, -4)^T$. The last vector of the basis is a vector from $\mathrm{Ker}(A - 7I_5)$ that is linearly independent with x_3, e.g., $x_5 = (0, 1, 1, 0, 1)^T$.

The basis we are looking for is represented by vectors $x_1, \ldots, x_5$. If the vectors are put to rows of a matrix S, we get

$$J = S^{-1}AS = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 7 & 1 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{pmatrix},$$

which is the Jordan normal form of the matrix A.

The Jordan form of a matrix is unstable in the sense that a very small change of elements of a matrix sometimes causes a complete change of the Jordan form / the Jordan form is not a continuous function of elements of a matrix. This is reason whe the computing system Maple did not include computation of the Jordan form to its library.

The proof of 10.36 suggest the method how to compute the Jordan normal form of a given matrix A, but we will now concentrate to certain properties of the Jordan form. From 10.24, we know that the number of eigenvalues and eigenvectors is not changed by similarity transformations. Since every Jordan cell corresponds to one eigenvector, we get

Theorem 10.38 The number of Jordan cells corresponding to λ is equal to the number of eigenvectors for λ.

As a consequence we get that the (algebraic) multiplicity of every eigenvalue λ is always greater or equal to the number of the corresponding eigenvetors (the geometric multiplicity).

Consequence 10.39 The multiplicity of an eigenvalue λ is greater or equal to the number of the eigenvetors that correspond to λ.

However, such information is not sufficient to determine the Jordan normal form. In general, we need to know more, e.g., as it follows from the next formula.

Remark 10.40 (Sizes and the number of cells) The number of cells $J_k(\lambda)$ of a matrix $\mathbb{C}^{n\times n}$ in the resulting Jordan form is

$$\operatorname{rank}\left((A-\lambda I_n)^{k-1}\right) - 2\operatorname{rank}\left((A-\lambda I_n)^{k}\right) + \operatorname{rank}\left((A-\lambda I_n)^{k+1}\right).$$

Proof. See, e.g., Horn and Johnson [1985], Meyer [2000]. The formula can be derived from the idea of the proof of 10.36 on the Jordan normal form.

Now, let us show several examples ofapplications of the Jordan normal form.

Example 10.41 (Powers of a matrix) In Example 10.31 we mentioned using of the diagonalization for computing of powers of a matrix. Now, using the Jordan normal form, we can generalize the statement of arbitrary $A \in \mathbb{C}^{n\times n}$. Let $A = SJS^{-1}$, then

$$A_k = SJ_kS^{-1} = S\begin{pmatrix} J_{k_1}(\lambda_1)^k & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{k_m}(\lambda_m)^k \end{pmatrix} S^{-1}.$$

Now, it is necessary to understand the behavior of Jordan cells $J_{k_i}(\lambda_i)^k$, $i = 1, \ldots, m$. Asymptitically we get similarly as for diagonalizable matrices

$$\lim_{k\to\infty} A^k = \begin{cases} 0, & \text{if} \rho(A) < 1, \\ \text{diverges}, & \text{if} \rho(A) > 1, \\ \text{converges/diverges}, & \text{if} \rho(A) = 1, \end{cases}$$

Example 10.42 (Matrix function) Let us ask: how to define a matrix function, e.g., $\cos(A)$, e^A, etc.? For a real function $f : \mathbb{R} \to \mathbb{R}$ and a matrix $A \in \mathbb{R}^{n\times n}$ it would be possible to define $f(A)$ by applying f to each element of A separately,

$$f(A) = \begin{pmatrix} f(a_{11}) & \cdots & f(a_{1n}) \\ \vdots & & \vdots \\ f(a_{n1}) & \cdots & f(a_{nn}) \end{pmatrix}, \tag{10.3}$$

but such a matrix will not have good properties. Let us do it in another way. Suppose that the function $f : \mathbb{R} \to \mathbb{R}$ can be expressed as an infinite series $f(x) = \sum i = 0^{\infty} a_i x^i$; real analytical functions like $\sin(x)$, $\exp(x)$, etc. fulfill this assumption. Then, it is natural to define $f(A) = \sum i = 0^{\infty} a_i A^i$, We can compute powers of the metrix, and hence if $A = SJS^{-1}$, then

$$f(A) = \sum_{i=0}^{\infty} a_i SJ^iS^{-1} = Sf(J)S^{-1}.$$

It is easy to see that

$$f(J) = \begin{pmatrix} f(J_{k_1}(\lambda_1)) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & f(J_{k_m}(\lambda_m)) \end{pmatrix}.$$

It remains to define the image of Jordan cells $f(J_{k_i}(\lambda_i)$. This is trivial for $k_i = 1$, the cell is a matrix of order 1. If $k_i > 1$, the image is more complicated [Meyer, 2000]:

$$f(J) = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{f^{(k_i-1)}(\lambda_i)}{(k_i-1)!} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & f'(\lambda_i) \\ 0 & \cdots & 0 & f(\lambda_i) \end{pmatrix}.$$

For example, the function $f(x) = x^2$ has the matrix extension $f(A) = A^2$, i.e., the classical matrix power. On the other hand, the formula 10.3 would compute separate powers of particular elements of the matrix, which is not what we want.

Example 10.43 (A system of linear differential equations). Let us consider so called system of linear differential equations

$$u(t)' = Au(t) \tag{10.4}$$

where $A \in \mathbb{R}^{n\times n}$. Our goal is to find an unknown function $u : \mathbb{R} \to \mathbb{R}^n$ that fulfills the system for certain boundary condition of the form $u(t_0) = u_0$. In the case $n = 1$ the solution of a differential equation $u(t)' = au(t)$ is a function $u(t) = v \cdot e^{at}$, where $v \in \mathbb{R}$ is an arbitrary number (to be selected so that the boundary condition is fulfilled). This suggests a motivation (supported by a deeper theory) to look for a solution of the general case in the form

$$u(t) = (u_1(t), \ldots, u_n(t)) = (v_1 e^{\lambda t}, \ldots, v_n e^{\lambda t}) = e^{\lambda t} v,$$

where v_i, λ are unknowns, $i = 1, \ldots, n$. A substitution of $u(t) = e^{\lambda t}v$ to 10.4 gives

$$\lambda e^{\lambda t} v = e^{\lambda t} Av, \quad \text{which gives} \quad \lambda v = Av.$$

This is immediately the problem of computing eigenvalues and eigenvectors. Suppose that the matrix A has eigenvectors $\lambda_1, \ldots, \lambda_n$ and the correesponding eigenvectors $x_1, \ldots, x_n$. Then the solution of 10.4 is $u(t) = \sum_{i=1}^{n} \alpha_i e^{\lambda_i t} x_i$, where $\alpha_i \in \mathbb{R}$ is obtained from the boundary conditions.

Let us consider a particular example:

$$u_1'(t) = 7u_1(t) - 4u_2(t)$$

$$u_2'(t) = 5u_1(t) - 2u_2(t)$$

The matrix $A = \begin{pmatrix} 7 & -1 \\ 5 & -2 \end{pmatrix}$ has eigenvalues 2 and 3 and the corresponding eigenvectors $(4,5)^T$ and $(1,1)^T$. The solution of the problem has form

$$\begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} = \alpha_1 e^{2t} \begin{pmatrix} 4 \\ 5 \end{pmatrix} + \alpha_2 e^{3t} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \alpha_1, \alpha_2 \in \mathbb{R}.$$

When solving linear differential equations, the knowledge of eigenvalues and eigenvectors is important. The Jordan normal form is used in the case when the matrix A is not diagonalizable and the defficiency of the number of eigenvalues implies that the α_i's in the formula for $u(t)$ are not scalars (i.e., numbers), but polynomials of t of the oreder corresponding to the multiplicity of λ_i.

10.5 Symmetric matrices

Real symmetric matrices have several remarkable properties concerning eigenvalues. Their main properties are that they are always diagonalizable and their eigenvalues are real. [6]

Let us first look at the generalization of transposition and symmetry for complex matrices.

Definition 10.44 (Hermitian matrix and transposition) A *hermitian transposition* [7] of a matrix $A \in \mathbb{C}^{n\times n}$ is the matrix $A^* = (\overline{A})^T$. A matrix $A \in \mathbb{C}^{n\times n}$ is *hermitian*, if $A^* = A$.

Let us recall that $\overline{A}$ is the matrix that results from replacing all elements of A by the coresponding complex conjugates.

The hermitian transposition has similar properties as the classical transposition, e.g., $(A^*)^* = A$, $(\alpha A)^* = \overline{\alpha} A^*$, $(A+B)^* = A^* + B^*$, $(AB)^* = B^* A^*$.

Using hermitian transpozition, we can define unitary matrices (that generalize the notion of an orthogonal matrix for complex matrices, see Definition 8.46) as matrices $Q \in \mathbb{C}^{n\times n}$ that fulfill $Q^*Q = I_n$.

Example 10.45 Given matrices

$$\begin{pmatrix} 2 & 1+i \\ 1+i & 5 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1+i \\ 1-i & 5 \end{pmatrix}$$

the first one is symmetric, but not hermitian, and the second one is hermitian, but not symmetric. For real matrices both notions are the same.

Theorem 10.46 (Eigenvalues of symmetric matrices) Eigenvalues of real symmetric matrices and more generally complex hermitian matrices are real numbers.

[6] Remark L.K. And their eigenvectors corresponding to different eigenvalues are orthogonal.

[7] Charles Hermite (1822-1901) was a French mathematician. He proved, among others, that e is transcenential, i.e., it is not a root of any polynomial with rational coefficients

Proof.

Let $A \in \mathbb{C}^{n\times n}$ be hermitian and let $\lambda \in \mathbb{C}$ is its eigenvalue and $x \in \mathbb{C}^n$ be the corresponding eigenvectors of the unit length, i.e., $\|x\|_2 = 1$. Then $Ax = \lambda x$, after multiplying by x^* we get $x^*Ax = \lambda x^*x = \lambda$. Now,

$$\lambda = x * Ax = x^*A^*x = (x * Ax)^* = \lambda^*.$$

This implies $\lambda = \lambda^*$, and hence $\lambda \in \mathbb{R}$. ♣

Let us remark that complex symmetric matrices can have proper complex eigenvalues.

The following theorem says that symmetric real matrices are diagonalizable. [8] Moreover, they are diagonalizable in a very specific way: it is possible to construct an orthonormal basis that consists of their eigenvectors, which means that the similarity matrix is orthogonal.

Theorem 10.47 (Spectral decomposition of symmetric real matrices) Given a symmetric real matrix $A \in \mathbb{R}^{n\times n}$, there is an orthogonal $Q \in \mathbb{R}^{n\times n}$ and a diagonal $\Lambda \in \mathbb{R}^{n\times n}$ such that $A = Q\Lambda Q^T$..

Proof.

By the methematical induction by n. The case $n = 1$ is trivial: $\Lambda = A$ and $Q = 1$.

The induction step $n \leftarrow n-1$. Let λ be an eigenvalue of A and x is the corresponding eigenvector of the unit length, i.e., $\|x\|_2 = 1$. Let us complete (see 8.23) the vector x, that represents an orthonormal system of one vector, to an orthogonal matrix $S = (x|\cdots)$. Because $(A-\lambda I_n)x = o$, we get $(A-\lambda I_n)S = (o|\cdots)$, which means that $S^T(A-\lambda I_n)S = S^T(o|\cdots) = (o|\cdots)$. And since this matrix is symmetric, we get

$$S^T(A - \lambda I_n)S = \begin{pmatrix} 0 & o^T \\ o & A' \end{pmatrix},$$

where A' is certain symmetric matrix of the order $n-1$. Using the induction hypothesis, the matrix A' has a spectral decomposition $A' = Q'\Lambda' Q'^T$, where λ' is diagonal, and Q' is orthogonal. The matrix and the equation can be extended to the order n as follows:

$$\begin{pmatrix} 0 & o^T \\ o & A' \end{pmatrix} = \begin{pmatrix} 1 & o^T \\ o & Q' \end{pmatrix} \begin{pmatrix} 0 & o^T \\ o & \Lambda' \end{pmatrix} \begin{pmatrix} 1 & o^T \\ o & Q'^T \end{pmatrix},$$

which can easily be verified by simple multiplication. Let us denots

$$R = \begin{pmatrix} 1 & o^T \\ o & Q' \end{pmatrix}, \quad \Lambda'' = \begin{pmatrix} 0 & o^T \\ o & \Lambda' \end{pmatrix}.$$

The matrix $\mathbb{R}$ is orthogonal (Theorem 8.50 (4)), the matrix Λ'' is diagonal. Now, we can write

[8] Augustin Louis Cauchy, 1829

$$S^T(A - \lambda I_n)S = R\Lambda'' R^T,$$

which implies

$$A = SR\Lambda'' R^T S^T + \lambda I_n = SR\Lambda'' R^T S^T + \lambda SRR^T S^T = SR(\Lambda'' + \lambda I_n)R^T S^T.$$

This gives the required decomposition $A = Q\Lambda Q^T$, where $Q - SR$ is an orthogonal matrix, and $\Lambda = \Lambda'' + \lambda I_n$ is diagonal. ♣

Similarly we can get a spectral decomposition of a hermitian matrix in the form $A = Q\Lambda Q^*$, where Q is unitary matrix.

Remark 10.48 (Another form of a spectral decomposition) Let a symmetric $A \in \mathbb{R}^{n\times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ and the corresponding orthonormal eigenvectors. This means that, in the spectral decomposition $A = Q\Lambda Q^T$ it is $\Lambda_{ii} = \lambda_i$ and $Q_{*i} = x_i$. Let us write

$$A = Q\Lambda Q^T = Q\left(\sum_{i=1}^n \lambda_i e_i e_i^T\right) Q^T = \sum_{i=1}^n \lambda_i Q e_i e_i^T Q^T = \sum_{i=1}^n \lambda_i Q_{*i} Q_{*i}^T = \sum_{i=1}^n \lambda_i x_i x_i^T.$$

The form $A = \sum_{i=1}^n \lambda_i x_i x_i^T$ is an alternative form of the spectral decomposition, where the matrix A is decomposed to a sum of n matrices of the rank 1 or 0. Moreover, $x_i x_i^T$ is the matrix of the projection to the line $\mathrm{span}\{x_i\}$, which means that, from geometric point of view, the mapping $x \mapsto Ax$ can be viewed as a sum of n mapping, each of which is a projection to a line (orthogonal to other lines) and scaling by the value of λ_i.

There is one nice (even though a bit theoretical) consequence of the spectral decomposition that gives a formula for computing the largest and the smallest eigenvalue. [9] The consequence says that the largest, the smallest, resp., eigenvalue is the largest, the smallest, resp., value of the quadratic function $f(x) = x^T Ax$ on the unit sphere.

Theorem 10.49 (Courant-Fischer). Let $\lambda_1 \geq \cdots \geq \lambda_n$ be eigenvalues of a symmetric matrix $A \in \mathbb{R}^{n\times n}$. Then

$$\lambda_1 = \max_{\|x\|_2=1} x^T Ax, \quad \lambda_n = \min_{\|x\|_2=1} x^T Ax.$$

Proof. The proof is given just for λ_1, the second part is analagous.

The inequality "$\leq$": Let x_1 be the eigenvector corresponding to λ with the norm $\|x_1\|_2 = 1$. Then $Ax_1 = \lambda_1 x_1$. Multiplying by x_1^T from the left we get

[9] Ernst Fischer proved the formula in 1905, Richard Courant gave a generalization to operators of infinite dimansions in 1920. Their formula involves the intermediate eigenvalues $\lambda_2, \ldots, \lambda_{n-1}$ as well, but this case is more complicated. The simple version is sometimes called Ralyleigh-Ritz theorem.

$$\lambda_1 = \lambda_1 x_1^t x_1 = x_1^T A x_1 \leq \max_{\|x\|_2=1} x^T AX.$$

The inequality "$\geq$": Let $x \in \mathbb{R}^n$ be an arbitrary vector such that $\|x\|_2 = 1$. Denote $y = Q^T x$, and then $\|y\|_2 = 1$ (Theorem 8.50 (2)). Using the spectral decomposition $A = Q\Lambda Q^T$ we get

$$x^T Ax = x^T Q\Lambda Q^T x = y^T \Lambda y = \sum_{i=0}^{n} \lambda_i y_i^2 \leq \sum_{i=1}^{n} \lambda_1 y_i^2 = \lambda_1 \|y\|_2^2 = \lambda_1.$$

10.6 Theory of non-negative matrices

Skipped

10.7 Computing eigenvalues

As it has already been mentioned (remarks to Theorem 10.16), eigenvalues are computed by purely numerical iterative methods, and looking for them using the characteristic polynomial of a matrix is not an effective procedure. In this section we will show an easy estimation of eigenvalues, and an easy method of computing the largest eigenvalue. Another method, the popular QR algorithm, will be presented in 13.3. For very precise computation of eigenvalues of symmetric (especially positive definite) matrices one can use Jacobi method (see, e.g., Rohn. 2004) and Lanczos method [10] (see Meyer, 2000) for very large sparse symmetric matrices.

Gerschgoring disks skipped.

Now we will show a simple method for computing the dominant eigenvalue[11]. Despite of its simplicity it became the basis of some numerical methods, e.g., inverse iteration method[12] or the Rayleigh quotient iteration[13] for symmetric matrices.

Theorem 10.55 (The power method). Let $A \in \mathbb{C}^{n\times n}$.
1. Choose $o \neq x_0 \in \mathbb{C}^n$, $i := 1$;
2. **while not** a halting condition is not satisfied **do**
3. $y_i; = Ax_{i-1}$,
4. $x_i := \frac{1}{\|y_i\|_2 y_i}$.
5. $i := i + 1$,
6. **end while**

The output $\lambda_1 := x_{i-1}^T y_i$ is an estimation of the eigenvalue, $v_1 := x_i$ is an estimation of the corresponding eigenvector.

[10] The algorithm was discovered by the Hungarian mathematician Cornelius Lanczos in 1950

[11] The power method was presented by the american mathematician and physicist Richard Edler von Mises in 1929

[12] the author is the German mathematician Helhut Wielandt, 1944

[13] the author is the English physicist John William Strutt, lord Rayleigh, the Nobel price winner in physics (1904)

Example 10.56 Let

$$A = \begin{pmatrix} 1 & 4 & 2 \\ 4 & 2 & 2 \\ 2 & 2 & -1 \end{pmatrix}, \quad x_0 = (1, 0, 1)^T.$$

The course of the computation and the source code for Matlab / Octave (see the Czech version of the lecture notes).

The method is interupted when the value of $x_{i-1}^T y_i$ or the vector x_i are stable; then $x_i \approx x_{i-1}$ is an estimation of the eigenvector, and $x_{i-1}^T y_i = x_{i-1}^T A x_{i-1} \approx x_i^T \lambda x_{i-1} \approx \lambda$ is an estimation of the corresponding eigenvector. The method could be slow, it is difficult to estimate an error and the rate of convergence, and, moreover, the behavior of the method strongly depends on the initial choice of x_0. On the other hand the method is robust (rounding error are not too important) and the mtrhod can easily be applied to very large sparse metrices. The method sometimes doesn't converge, but the convergence can be guaranteed under certain conditions.

Theorem 10.57 (The convergence of the power method). Let $A \in \mathbb{R}^{n\times n}$ be a matrix with eigenvalues $|\lambda_1| > |\lambda_2| \leq \cdots \leq |\lambda_n|$ and the corresponding eigenvectors $v_1, \ldots, v_n$ of the length 1. Let $x_0$has a non-zero coordinate in the direction of v_1. Then x_i converges (up to a signature) to the eigenvector v_1 and $x_{i-1}^T y_i$ converges to the eigenvalue λ_1.

Proof. Let $x_0 = \sum_{j=1}^n \alpha_j v_j$, where $\alpha_1 \neq 0$, see the assumptions. Then $x_i = \frac{1}{\|A^i x_0\|} A^i x_0$ and

$$A^i x_0 = A^i \left(x_0 = \sum_{j=1}^n \alpha_j v_j \right) = \sum_{j=1}^n \alpha_j A^i v_j = \sum_{j=1}^n \alpha_j \lambda_j^i v_j = \lambda_1^i \left(\alpha_1 v_1 + \sum_{j=2}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^i v_j \right).$$

since the vectors x_i are always normalized, the multiplicative factor λ_1^i is not interesting. The remaining vector converges to $\alpha_1 v_1$, because $\left|\frac{\lambda_j}{\lambda_1}\right| < 1$ and therefore $\left|\frac{\lambda_j}{\lambda_1}\right|^i \to 0$ for $i \to \infty$.

Suppose now that x_i approximates well the eigenvector v_1 Then $x_{i-1}^T y_i = x_{i-1}^T A x_{i-1} = x_{i-1}^T \lambda_1 x_{i-1} \lambda_1 \|x_{i-1}\|_2^2 = \lambda_1$. ♣

The power method computes just the dominant eigenvalue and eigenvector. However, the following technique makes it possible to eliminate one eigenvalue, e.g., the dominant one, and then we can use the power method recursively to get the remaining eigenvalues. We give first a simple version for symmetric matrices, and then, in Remark 10.59, the general method.

Theorem 10.58 (A deflation of an eigenvalue). Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, $\lambda_1, \ldots, \lambda_n$ its eigenvalues and $v_1, \ldots, v_n$ the corresponding eigenvectors. Then the matrix $A - \lambda_1 v_1 v_1^T$ has eigenvalues $0, \lambda_2, \ldots, \lambda_n$, and eigenvectors $v_1, \ldots, v_n$.

Proof. In view of Remark 10.48, we can write $A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$. Then $A - \lambda_1 v_1 v_1^T = 0 v_1 v_1^T + \sum_{i=2}^{n} \lambda_i v_i v_i^T$, which is the spectral decomposition of the matrix $A - \lambda_1 v_1 v_1^T$, which is the spectral decomposition of the matrix $A - \lambda_1 v_1 v_1^T$.

Remark 10.59 (A deflation of an eigenvalue of a general matrix). Let λ be an eigenvalue and x be the corresponding eigenvector of the matrix $A \in \mathbb{R}^{n \times n}$. Let x be completed to a regular matrix S so that $S_{*1} = x$. Then

$$S^{-1}AS = S^{-1}A(x|\cdots) = S^{-1}(\lambda x|\cdots) = (\lambda e_1|\cdots) = \begin{pmatrix} \lambda & \cdots & \\ o & & A' \end{pmatrix}.$$

In view of symmetry, A' has the same eigenvalues as A, just the multiplicity of λ is smaller by !. Thus, the remaining eigenvalues and eigenvectors of A can be found using A'.

Example 10.60 (Google^{TM} and PageRank[14]). Consider a web network with the following parameters:
N web pages,
$a_{ij} = 1$ if j-th page points to i-th one, $a_{ij} = 0$ otherwise,
b_j is the number of pointers from the j-th page,
x_i is the importance of the i-th page.

The gpoal is to determine the importances $x_1, \ldots, x_n$ of particular pages. The main idea of the Google PageRank is in determining the importance of the i-th page so that it is proportional to the sum og the importances of pages, where it is pointed. This means that we are solving the eguation $x_i = \sum_{j=1}^{n} \frac{a_{ij}}{b_j} x_j$, $i = 1, \ldots, N$. In the matrix setting $A'x = x$, where $a'_{ij} := \frac{a_{ij}}{b_j}$. This means that x is the eigenvector of A that corresponds to the eigenvalue 1. The eigenvalue 1 is dominant, which can easily be seen from Gerschgoring disks of the matrix A^T (the sum of columns of the metrix A' is 1, i.e., all Gerschgoring's disks have the rightmost end in the

[14] From 1997, the authors are Sergey Brin and Larry Page