

week1: Eigenvalues, Eigenvectors and Diagonalization of Matrices

Let us start by recalling the definitions of eigenvalues and eigenvectors of matrices. These concepts play an important role in many areas of the natural sciences, engineering and quantitative economics. They were defined as follows:

Definition 1.1 (eigenvector and eigenvalue): Let A be an arbitrary $n \times n$ matrix. A vector $\vec{v} \neq \vec{0}$ with

$$A\vec{v} = \lambda\vec{v} \quad (1.1)$$

is called an eigenvector of the matrix A with eigenvalue λ .

Remark: If $\vec{v}$ is an eigenvector with eigenvalue λ , then $\vec{w} := c\vec{v}$ with an arbitrary constant $c \neq 0$ is also an eigenvector with the same eigenvalue λ , since

$$A\vec{w} = A(c\vec{v}) = cA\vec{v} = c\lambda\vec{v} = \lambda c\vec{v} = \lambda\vec{w}$$

Eigenvectors are therefore never unique; at best they are unique up to normalization. Let us first look at an example:

Example: Let $a \in \mathbb{R}$ be a real parameter and let $\epsilon \in \{+1, -1\}$ be a sign. Determine all eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & 1 \\ \epsilon a^2 & 1 \end{pmatrix} \quad (1.2)$$

When does a basis of eigenvectors exist?

Solution: We have to find all solutions of the equation

$$\begin{aligned} A\vec{v} &= \lambda\vec{v} \\ \Leftrightarrow A\vec{v} - \lambda\vec{v} &= \vec{0} \\ \Leftrightarrow (A - \lambda Id)\vec{v} &= \vec{0} \end{aligned} \quad (1.3)$$

Here Id is the 2×2 identity matrix, for which we have $Id\vec{v} = \vec{v}$,

$$Id\vec{v} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \vec{v}$$

We can regard equation (1.3) as a system of linear equations for the coordinates v_1 and v_2 of $\vec{v} = (v_1, v_2)$. Suppose the matrix

$$M := A - \lambda Id$$

were invertible. Then we could multiply (1.3) by M^{-1} and would necessarily obtain

$$\vec{v} = M^{-1}M\vec{v} = M^{-1}(A - \lambda Id)\vec{v} \stackrel{(1.3)}{=} M^{-1}\vec{0} = \vec{0}$$

However, we want to find vectors $\vec{v} \neq \vec{0}$. This is only possible if the matrix M is not invertible. A necessary and sufficient condition for this is

$$\det M = \det[A - \lambda Id] \stackrel{!}{=} 0 \quad (1.4)$$

Specifically, in our case this gives

$$\det \left[\begin{pmatrix} 1 & 1 \\ \epsilon a^2 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] = \det \begin{pmatrix} 1-\lambda & 1 \\ \epsilon a^2 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - \epsilon a^2 \stackrel{!}{=} 0$$

or

$$(\lambda - 1)^2 \stackrel{!}{=} \epsilon a^2 \quad (1.5)$$

Let us first consider the

Case $\epsilon = +1$: We obtain the eigenvalues

$$\lambda_{\pm} = 1 \pm a \quad (1.6)$$

and the eigenvectors $\vec{v} = (v_1, v_2)$ are solutions of (1.3),

$$\begin{pmatrix} 1-\lambda_{\pm} & 1 \\ a^2 & 1-\lambda_{\pm} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \mp a & 1 \\ a^2 & \mp a \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (1.7)$$

or, written out explicitly,

$$\begin{aligned} \mp a v_1 + v_2 &= 0 \\ a^2 v_1 \mp a v_2 &= 0 \end{aligned} \quad (1.8)$$

If we multiply the first row in (1.8) by $\mp a$, we obviously obtain exactly the second row, so we can drop the second row; it is redundant. We therefore have to solve

$$\mp a v_1 + v_2 = 0 \quad (1.9)$$

and thus obtain the eigenvectors

$$\vec{v}_+ = \begin{pmatrix} 1 \\ +a \end{pmatrix}, \quad \vec{v}_- = \begin{pmatrix} 1 \\ -a \end{pmatrix} \quad (1.10)$$

which can also be confirmed by explicit calculation,

$$\begin{aligned} \begin{pmatrix} 1 & 1 \\ a^2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a \end{pmatrix} &= \begin{pmatrix} 1+a \\ a^2+a \end{pmatrix} = (1+a) \begin{pmatrix} 1 \\ a \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ a^2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -a \end{pmatrix} &= \begin{pmatrix} 1-a \\ a^2-a \end{pmatrix} = (1-a) \begin{pmatrix} 1 \\ -a \end{pmatrix} \end{aligned} \quad (1.11)$$

For $a \neq 0$, $\vec{v}_+$ and $\vec{v}_-$ are linearly independent and form a basis of $\mathbb{R}^2$. For $a = 0$ there is only one eigenvector $\vec{v} = \vec{e}_1 = (1, 0)$, so that no basis of eigenvectors can be given. Let us now consider the

Case $\epsilon = -1$: In this case equation (1.5) reads

$$(\lambda - 1)^2 \stackrel{!}{=} -a^2 \quad (1.12)$$

The solution set of equation (1.12) obviously depends on whether we allow only real numbers or also complex numbers. We have to make a further case distinction:

Complex case: The underlying space is $\mathbb{C}^2$ and complex eigenvalues $\lambda \in \mathbb{C}$ are allowed: In this case we obtain the complex eigenvalues

$$\lambda_{\pm} = 1 \pm i a \quad (1.13)$$

and the eigenvectors $\vec{v}_{\pm}$ follow from the system of equations

$$\begin{pmatrix} 1 - \lambda_{\pm} & 1 \\ -a^2 & 1 - \lambda_{\pm} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \mp i a & 1 \\ -a^2 & \mp i a \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (1.14)$$

or, written out explicitly,

$$\begin{aligned} \mp i a v_1 + v_2 &= 0 \\ -a^2 v_1 \mp i a v_2 &= 0 \end{aligned} \quad (1.15)$$

If we now multiply the first row in (1.15) by $\mp i a$, we again obtain the second row, so we can again drop it. We then have to solve

$$\mp i a v_1 + v_2 = 0 \quad (1.16)$$

and thus obtain the eigenvectors

$$\vec{v}_+ = \begin{pmatrix} 1 \\ +i a \end{pmatrix}, \quad \vec{v}_- = \begin{pmatrix} 1 \\ -i a \end{pmatrix} \quad (1.17)$$

For $a \neq 0$ these are again linearly independent and form a basis of $\mathbb{C}^2$.

Real case: The underlying space is $\mathbb{R}^2$ and we want to allow only real eigenvalues $\lambda \in \mathbb{R}$: In this case equation (1.12) has no solutions and there are no eigenvalues and no eigenvectors. ■

Let us record the following terminology:

Definition 1.2 (characteristic polynomial): Let A be an arbitrary $n \times n$ matrix. Then the polynomial

$$\chi_A(\lambda) := \det[A - \lambda Id] \quad (1.18)$$

is called the characteristic polynomial of the matrix A . The eigenvalues of a matrix A are then exactly the zeros of the characteristic polynomial $\chi_A(\lambda)$, and the eigenvectors are exactly the nontrivial solutions of the system of linear equations

$$(A - \lambda Id) \vec{v} = \vec{0} \quad (1.19)$$

The set of eigenvalues and eigenvectors depends on whether the underlying space is $\mathbb{R}^n$ or $\mathbb{C}^n$.

Definition 1.3 (diagonalizability): Let A be an arbitrary $n \times n$ matrix. Then A is called

- real diagonalizable if there is a basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors with real eigenvalues $\lambda_1, \dots, \lambda_n$.
- complex diagonalizable if there is a basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors with complex eigenvalues $\lambda_1, \dots, \lambda_n$.

Since every real number is also a complex number, we have: A real diagonalizable $\Rightarrow A$ complex diagonalizable. The converse obviously does not hold, as we have just seen in the example: there are complex diagonalizable matrices which are not real diagonalizable.

Theorem 1.1: Let A be a real or complex diagonalizable $n \times n$ matrix with a basis of eigenvectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ with eigenvalues $\lambda_1, \dots, \lambda_n$, that is, $A \vec{v}_k = \lambda_k \vec{v}_k$. Furthermore, let V be the matrix of eigenvectors, more precisely

$$V := \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{pmatrix} \quad (1.20)$$

and let V^{-1} be the inverse matrix of V . Then

$$A = V D V^{-1} \quad (1.21)$$

where D is the diagonal matrix with the entries $\lambda_1, \dots, \lambda_n$ on the main diagonal and zeros everywhere else,

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} =: \text{Diag}(\lambda_1, \dots, \lambda_n) \quad (1.22)$$

Proof: Let $\vec{e}_1, \dots, \vec{e}_n$ be the standard basis vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$. By definition of V we then have

$$V \vec{e}_k = \vec{v}_k \quad (1.23)$$

since the columns of a matrix contain the images of the basis vectors. We apply V^{-1} to both sides and obtain

$$\vec{e}_k = V^{-1} \vec{v}_k \quad (1.24)$$

Furthermore,

$$D \vec{e}_k = \lambda_k \vec{e}_k \quad (1.25)$$

Altogether, therefore,

$$V D V^{-1} \vec{v}_k \stackrel{(1.24)}{=} V D \vec{e}_k \stackrel{(1.25)}{=} V \lambda_k \vec{e}_k = \lambda_k V \vec{e}_k \stackrel{(1.23)}{=} \lambda_k \vec{v}_k = A \vec{v}_k \quad (1.26)$$

Since $\vec{v}_1, \dots, \vec{v}_n$ form a basis by assumption, every $\vec{x}$ can be written as a linear combination of the $\vec{v}_k$, hence $A\vec{x} = V D V^{-1}\vec{x}$ also holds for arbitrary $\vec{x}$, and the theorem is proven. ■

1.2 Diagonalizability of Matrices

In this section we want to give a simple criterion for the diagonalizability of matrices which is of great importance, in particular in physics. Let us first recall the standard scalar products on $\mathbb{R}^n$ and on $\mathbb{C}^n$; they are defined by

$$(\vec{x}, \vec{y})_{\mathbb{R}^n} := x_1 y_1 + \dots + x_n y_n = \sum_{j=1}^n x_j y_j \quad (1.27)$$

$$(\vec{w}, \vec{z})_{\mathbb{C}^n} := \bar{w}_1 z_1 + \dots + \bar{w}_n z_n = \sum_{j=1}^n \bar{w}_j z_j \quad (1.28)$$

where $\bar{w}_j$ denotes the complex conjugate of the complex number w_j . So if, for instance, $w_j = a_j + ib_j$ with real a_j, b_j , then $\bar{w}_j := a_j - ib_j$. The reason for defining it this way for complex vectors is that one then still has the property

$$(\vec{z}, \vec{z})_{\mathbb{C}^n} \geq 0 \quad \text{and} \quad (\vec{z}, \vec{z})_{\mathbb{C}^n} = 0 \Leftrightarrow \vec{z} = \vec{0} \quad (1.29)$$

which one always wants a scalar product to have. A basic property matrices do have in the context of a real or complex scalar product is the following identity:

$$(A\vec{x}, \vec{y})_{\mathbb{R}^n} = (\vec{x}, A^T \vec{y})_{\mathbb{R}^n} \quad (1.30)$$

$$(A\vec{w}, \vec{z})_{\mathbb{C}^n} = (\vec{w}, A^* \vec{z})_{\mathbb{C}^n} \quad (1.31)$$

In a scalar product, matrices can be brought to the other side by taking the transpose or, in the complex case, the adjoint matrix. Here the adjoint matrix A^* is given by

$$A^* := \bar{A}^T \quad (1.32)$$

with the bar $\bar{\cdot}$ denoting complex conjugation again.

Definition 1.4 (symmetric and self-adjoint matrices): An $n \times n$ matrix A is called

$$\begin{array}{ll} \text{symmetric} & \stackrel{\text{Def.}}{\Leftrightarrow} A^T = A \\ \text{self-adjoint} & \stackrel{\text{Def.}}{\Leftrightarrow} A^* = A \end{array}$$

A quick example: Let

$$A_1 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & i \\ i & 3 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & i \\ -i & 3 \end{pmatrix} \quad (1.33)$$

Then the matrices A_1 and A_2 are symmetric and the matrices A_1 and A_3 are self-adjoint. We have now made all the preparations needed to state the following important theorem:

Theorem 1.2 (diagonalizability of symmetric and self-adjoint matrices):

- a) Every real, symmetric $n \times n$ matrix A is diagonalizable. More precisely: There is an orthonormal basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors with

$$A \vec{v}_k = \lambda_k \vec{v}_k \quad \text{and} \quad (\vec{v}_j, \vec{v}_k)_{\mathbb{R}^n} = \delta_{j,k} \quad (1.34)$$

- b) Every complex, self-adjoint $n \times n$ matrix A is diagonalizable. More precisely: There is an orthonormal basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors with

$$A \vec{v}_k = \lambda_k \vec{v}_k \quad \text{and} \quad (\vec{v}_j, \vec{v}_k)_{\mathbb{C}^n} = \delta_{j,k} \quad (1.35)$$

Moreover, all eigenvalues are necessarily real, $\lambda_k \in \mathbb{R}$.

Proof: Without proof. A complete proof of the complex version can be found, for example, in the recent lecture notes *Lineare Algebra für Elektrotechnik* (Linear Algebra for Electrical Engineering) by Volker Bach at TU Braunschweig,

<https://hsrm-mathematik.de/WS2627/AdvMath-AP-BE/LinAlg-ET-WS2526-TU-Braunschweig.pdf>

You can find the theorem there in Chapter X.3 under the name ‘Spektralsatz’ (spectral theorem). ■

Let us recall equations (1.20) and (1.21) for diagonalizable matrices,

$$A = V D V^{-1}$$

where the columns of V contain the eigenvectors,

$$V = \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{pmatrix}$$

For orthonormal bases, ONBs, V^{-1} becomes particularly simple. To see this, consider the matrix products $V^T V$ or $V^* V$: If we have a real ONB, we can write

$$\begin{aligned} V^T V &= \begin{pmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_n & - \end{pmatrix} \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{pmatrix} = \left((\vec{v}_i, \vec{v}_j)_{\mathbb{R}^n} \right)_{i,j=1,\dots,n} \\ &\stackrel{\text{ONB}}{=} (\delta_{i,j})_{i,j=1,\dots,n} = Id \end{aligned} \quad (1.36)$$

so that

$$V^T = V^{-1} \quad (1.37)$$

and

$$A = V D V^T \quad (1.38)$$

If we have a complex orthonormal basis, we consider

$$\begin{aligned}
V^*V &= \bar{V}^T V = \begin{pmatrix} - & \overrightarrow{v_1} & - \\ & \vdots & \\ - & \overrightarrow{v_n} & - \end{pmatrix} \begin{pmatrix} | & & | \\ \overrightarrow{v_1} & \cdots & \overrightarrow{v_n} \\ | & & | \end{pmatrix} = \left((\overrightarrow{v_i}, \overrightarrow{v_j})_{\mathbb{C}^n} \right)_{i,j=1,\dots,n} \\
&\stackrel{\text{ONB}}{=} (\delta_{i,j})_{i,j=1,\dots,n} = Id
\end{aligned} \tag{1.39}$$

so that

$$V^* = V^{-1} \tag{1.40}$$

and

$$A = V D V^* \tag{1.41}$$

Definition 1.5 (orthogonal and unitary matrices): The following types of matrices are in one-to-one correspondence with the orthonormal bases of $\mathbb{R}^n$ or $\mathbb{C}^n$:

a) The set

$$O(n) := \{ V \in \mathbb{R}^{n \times n} \mid V^T = V^{-1} \} \tag{1.42}$$

is called the set of orthogonal matrices.

b) The set

$$U(n) := \{ U \in \mathbb{C}^{n \times n} \mid U^* := \bar{U}^T = U^{-1} \} \tag{1.43}$$

is called the set of unitary matrices.

Because $\det[A^{-1}] = 1/\det A$ and $\det[A^T] = \det A$, we find for orthogonal matrices

$$\det V = \det[V^T] = \det[V^{-1}] = 1/\det V$$

or

$$(\det V)^2 = 1 \tag{1.44}$$

Hence, since orthogonal matrices are real matrices by definition,

$$\det V = \pm 1 \tag{1.45}$$

For unitary matrices the calculation looks as follows:

$$\overline{\det U} = \det \bar{U} = \det[\bar{U}^T] = \det U^* = \det[U^{-1}] = 1/\det U$$

so that

$$\det U \overline{\det U} = |\det U|^2 = 1 \tag{1.46}$$

or

$$|\det U| = 1 \tag{1.47}$$

One then also defines the set of special orthogonal matrices $SO(n)$ and the set of special unitary matrices $SU(n)$ by

$$SO(n) := \{ V \in O(n) \mid \det V = +1 \} \tag{1.48}$$

$$SU(n) := \{ U \in U(n) \mid \det U = +1 \} \tag{1.49}$$

These notations are standard and are frequently used in the literature.