Fall 2019 Math 201-003 at Drexel: blackboard notes of 2019-11-11

Darij Grinberg

November 12, 2019

1. Eigenvalues and eigenvectors ("eigenstuff")

1.1. Definition and examples (recall)

Recall some material from last time and from the lecture before:

Definition 1.1.1. Let A be an $n \times n$-matrix. Let λ be a scalar (i.e., a real number).

(a) A **λ-eigenvector of** A means a nonzero vector $v \in \mathbb{R}^n$ such that $Av = \lambda v$.

(b) We say that λ is an **eigenvalue** of A if and only if there exists a λ-eigenvector of A.

Definition 1.1.2. Let A be an $n \times n$-matrix. We define

$$\chi_A(t) = \det(A - tI_n).$$

This is a polynomial in t, and is called the **characteristic polynomial** of A.

Proposition 1.1.3. Let A be an $n \times n$-matrix. Then, the eigenvalues of A are the roots of the characteristic polynomial $\chi_A(t)$.

Method for finding eigenvalues and eigenvectors of a matrix:
Given an $n \times n$-matrix A, find all eigenvalues and eigenvectors of A as follows:

- Calculate the characteristic polynomial $\chi_A(t) = \det(A - tI_n)$ of A.
- Find all roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ of $\chi_A(t)$. These are the eigenvalues of A.
- For each eigenvalue λ_i, compute the nonzero solutions to $(A - \lambda_i I_n) v = 0$ (for example, using Gaussian elimination). These are the λ_i-eigenvectors of A.

Proposition 1.1.4. Let A be a $n \times n$-matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be k distinct eigenvalues of A. For each $i \in \{1, 2, \ldots, k\}$, let $v_{i,1}, v_{i,2}, \ldots, v_{i,h_i}$ be some linearly independent λ_i-eigenvectors of A. Then, the list

$$\underbrace{v_{1,1}, v_{1,2}, \ldots, v_{1,h_1}}_{\text{eigenvectors for } \lambda_1},$$
$$\underbrace{v_{2,1}, v_{2,2}, \ldots, v_{2,h_2}}_{\text{eigenvectors for } \lambda_2},$$
$$\ldots,$$
$$\underbrace{v_{k,1}, v_{k,2}, \ldots, v_{k,h_k}}_{\text{eigenvectors for } \lambda_k}$$

(obtained by throwing all our eigenvectors for different eigenvalues together) is a list of linearly independent vectors.

Proposition 1.1.5. Let A be an $n \times n$-matrix that has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For each $i \in \{1, 2, \ldots, n\}$, let v_i be a λ_i-eigenvector of A. Then, $v_1, v_2, \ldots, v_n$ form a basis of $\mathbb{R}^n$.

1.2. Diagonalization

1.2.1. Motivation (recall)

The **Fibonacci numbers** are a sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers. They are defined recursively by

$$f_0 = 0, \qquad f_1 = 1, \qquad f_n = f_{n-1} + f_{n-2} \qquad \text{for all } n \geq 2.$$

We have showed that

$$\begin{pmatrix} f_m \\ f_{m+1} \end{pmatrix} = A^m \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \qquad \text{for each } m \geq 0, \tag{1}$$

where $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Thus, in order to compute f_m, it suffices to compute A^m.

Proposition 1.2.1. Let U be an invertible $n \times n$-matrix, and let D be any $n \times n$-matrix. Then,

$$\left(UDU^{-1}\right)^m = UD^mU^{-1}.$$

Thus, in particular, when $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$, then $D^m = \operatorname{diag}\left(d_1^m, d_2^m, \ldots, d_n^m\right)$, so this becomes

$$\left(UDU^{-1}\right)^m = U \operatorname{diag}\left(d_1^m, d_2^m, \ldots, d_n^m\right) U^{-1}. \tag{2}$$

1.2.2. The method

Recall the following definition:

Definition 1.2.2. Let A be an $n \times n$-matrix. A **diagonalization** of A means a pair of an invertible $n \times n$-matrix U and a diagonal $n \times n$-matrix D such that $A = UDU^{-1}$.

We stated (but did not prove) the following proposition:

Proposition 1.2.3. Let A be an $n \times n$-matrix.

(a) Suppose that $u_1, u_2, \ldots, u_n$ is a basis of $\mathbb{R}^n$ that consists of eigenvectors of A. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigenvalues (so each u_i is a λ_i-eigenvector). Note that some λ_i may be equal.

Set $U = [u_1 \mid u_2 \mid \cdots \mid u_n]$ and $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then, U, D is a diagonalization of A (that is, U is invertible, D is diagonal, and $A = UDU^{-1}$).

(b) Conversely, each diagonalization of A has this form. (In other words, if U, D is a diagonalization of A, then the columns of U form a basis of $\mathbb{R}^n$ that consists of eigenvectors of A, and the diagonal entries of D are the corresponding eigenvalues.)

We still need to prove this. First, let us apply this to the Fibonacci sequence:

Example 1.2.4. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ be the matrix that occurred in the computation of the Fibonacci numbers. We want to obtain a diagonalization of A (because this will help us compute A^m and thus f_m for all $m \geq 0$). Following Proposition 1.2.3 **(a)**, we try to achieve this by finding a basis of $\mathbb{R}^2$ consisting of eigenvectors of A. So let us find eigenvectors of A.

The characteristic polynomial of A is $\chi_A(t) = \det(A - tI_n) = \det \begin{pmatrix} 0-t & 1 \\ 1 & 1-t \end{pmatrix} = t^2 - t - 1$. So the eigenvalues of A are the roots of $t^2 - t - 1$, which are

$$\frac{1+\sqrt{5}}{2} \approx 1.618 \qquad \text{and} \qquad \frac{1-\sqrt{5}}{2} \approx -0.618 .$$

The first of these two roots is called the **golden ratio**. It is often denoted ϕ; it has the property that $\phi^2 = \phi + 1$, that is, $\phi - 1 = 1/\phi$. Let me denote the second root by ψ.

So

$$\phi = \frac{1+\sqrt{5}}{2} \approx 1.618 \qquad \text{and} \qquad \psi = \frac{1-\sqrt{5}}{2} \approx -0.618$$

are the two eigenvalues of A. Now, let's look for eigenvectors:

- The ϕ-eigenvectors of A are the nonzero $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ such that $A \begin{pmatrix} x \\ y \end{pmatrix} = \phi \begin{pmatrix} x \\ y \end{pmatrix}$. They are the nonzero multiples of $\begin{pmatrix} -\psi \\ 1 \end{pmatrix}$. (You can find this by Gaussian elimination.)
- The ψ-eigenvectors of A are the nonzero $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ such that $A \begin{pmatrix} x \\ y \end{pmatrix} = \psi \begin{pmatrix} x \\ y \end{pmatrix}$. They are the nonzero multiples of $\begin{pmatrix} -\phi \\ 1 \end{pmatrix}$. (You can find this by Gaussian elimination.)

So we have found two linearly independent eigenvectors of A, namely $\begin{pmatrix} -\psi \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -\phi \\ 1 \end{pmatrix}$. Thus, they form a basis of $\mathbb{R}^2$ consisting of eigenvectors of A.

Now, Proposition 1.2.3 **(a)** tells us that we can find a diagonalization (U, D) of A by setting

$$U = [u_1 \mid u_2 \mid \cdots \mid u_n] \text{ and } D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where the $u_1, u_2, \ldots, u_n$ are the eigenvectors of A (so $u_1 = \begin{pmatrix} -\psi \\ 1 \end{pmatrix}$ and $u_2 = \begin{pmatrix} -\phi \\ 1 \end{pmatrix}$) and where the $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues (so $\lambda_1 = \phi$ and $\lambda_2 = \psi$). Thus,

$$U = \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad D = \operatorname{diag}(\phi, \psi) = \begin{pmatrix} \phi & 0 \\ 0 & \psi \end{pmatrix}.$$

(You can easily double-check that these indeed satisfy $A = UDU^{-1}$.)

Now, let $m \geq 0$. Then, (2) (applied to $n = 2$, $d_1 = \phi$ and $d_2 = \psi$) yields

$$\left(UDU^{-1}\right)^m = U \operatorname{diag}(\phi^m, \psi^m) U^{-1} \qquad (\text{since } D = \operatorname{diag}(\phi, \psi)).$$

In view of $UDU^{-1} = A$, this rewrites as

$$\begin{aligned} A^m &= U \operatorname{diag}(\phi^m, \psi^m) U^{-1} = \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \operatorname{diag}(\phi^m, \psi^m) \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix}^{-1} \\ &\qquad \left(\text{since } U = \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \text{ and } n = 2 \text{ and } d_1 = \phi \text{ and } d_2 = \psi\right) \\ &= \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \underbrace{\begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix}^{-1}}_{\substack{= \frac{1}{(-\psi)-(-\phi)} \begin{pmatrix} 1 & \phi \\ -1 & -\psi \end{pmatrix} \\ \text{(by the formula } \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ \text{for the inverse of a } 2\times 2\text{-matrix)}}} \end{aligned}$$

$$= \underbrace{\frac{1}{(-\psi)-(-\phi)}}_{\substack{=\frac{1}{\sqrt{5}} \\ \text{(since a quick computation} \\ \text{shows that } (-\psi)-(-\phi)=\sqrt{5})}} \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \begin{pmatrix} 1 & \phi \\ -1 & -\psi \end{pmatrix}$$

$$= \frac{1}{\sqrt{5}} \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \begin{pmatrix} 1 & \phi \\ -1 & -\psi \end{pmatrix}.$$

We could multiply this out. But we want f_m, not A^m. There is a faster way to get f_m: From (1), we obtain

$$\begin{pmatrix} f_m \\ f_{m+1} \end{pmatrix} = \underbrace{A^m}_{=\frac{1}{\sqrt{5}} \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \begin{pmatrix} 1 & \phi \\ -1 & -\psi \end{pmatrix}} \underbrace{\begin{pmatrix} f_0 \\ f_1 \end{pmatrix}}_{\substack{=\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \text{(since } f_0=0 \text{ and } f_1=1)}}$$

$$= \frac{1}{\sqrt{5}} \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \underbrace{\begin{pmatrix} 1 & \phi \\ -1 & -\psi \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{=\begin{pmatrix} \phi \\ -\psi \end{pmatrix}}$$

$$= \frac{1}{\sqrt{5}} \begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} \phi^m & 0 \\ 0 & \psi^m \end{pmatrix} \begin{pmatrix} \phi \\ -\psi \end{pmatrix}}_{=\begin{pmatrix} \phi^m \phi \\ \psi^m (-\psi) \end{pmatrix}}$$

$$= \frac{1}{\sqrt{5}} \underbrace{\begin{pmatrix} -\psi & -\phi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m \phi \\ \psi^m (-\psi) \end{pmatrix}}_{\substack{=\begin{pmatrix} (-\psi)\phi^m\phi + (-\phi)\psi^m(-\psi) \\ * \end{pmatrix} \\ \text{(where the asterisk "}*\text{" stands for} \\ \text{an entry that we are not interested in)}}}$$

$$= \frac{1}{\sqrt{5}} \begin{pmatrix} (-\psi)\phi^m\phi + (-\phi)\psi^m(-\psi) \\ * \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \left((-\psi)\phi^m\phi + (-\phi)\psi^m(-\psi)\right) \\ * \end{pmatrix}.$$

Thus, by comparing the $(1,1)$-entries on both sides, we obtain

$$f_m = \frac{1}{\sqrt{5}} \left((-\psi)\phi^m\phi + (-\phi)\psi^m(-\psi)\right) = \frac{-1}{\sqrt{5}} \left(\underbrace{\phi\psi}_{=-1} \phi^m - \underbrace{\phi\psi}_{=-1} \psi^m \right)$$

$$= \frac{1}{\sqrt{5}} \left(\phi^m - \psi^m\right) = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^m - \left(\frac{1-\sqrt{5}}{2} \right)^m \right).$$

This is an **explicit** formula for the Fibonacci number f_m. It is known as the **Binet formula**. It shows that f_m grows exponentially with m, with growth rate $\phi \approx 1.618\ldots$.

In order to prove Proposition 1.2.3, we will need a lemma that follows quickly from the definition of matrix multiplication:

Lemma 1.2.5. Let $n \in \mathbb{N}$. Let $u_1, u_2, \ldots, u_n$ be n vectors in $\mathbb{R}^n$. Let $U = [u_1 \mid u_2 \mid \cdots \mid u_n]$ (this is an $n \times n$-matrix).

(a) If A is any $n \times n$-matrix, then

$$AU = [Au_1 \mid Au_2 \mid \cdots \mid Au_n].$$

(b) If $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal $n \times n$-matrix, then

$$UD = [\lambda_1 u_1 \mid \lambda_2 u_2 \mid \cdots \mid \lambda_n u_n].$$

Proof of Lemma 1.2.5. **(a)** Recall that any two $n \times n$-matrices A and B and any $j \in \{1, 2, \ldots, n\}$ satisfy

$$\operatorname{col}_j(AB) = A \cdot \operatorname{col}_j B. \tag{3}$$

(This is a particular case of Proposition 2.6.2 **(d)** in the notes from 2019-09-23.)

Now, let A be any $n \times n$-matrix. Then, for each $j \in \{1, 2, \ldots, n\}$, we have

$$\operatorname{col}_j(AU) = A \cdot \underbrace{\operatorname{col}_j U}_{\substack{=u_j \\ (\text{since } U = [u_1 \mid u_2 \mid \cdots \mid u_n])}} \qquad (\text{by (3), applied to } B = U)$$
$$= Au_j.$$

Thus, the n columns of the matrix AU are $Au_1, Au_2, \ldots, Au_n$. In other words, $AU = [Au_1 \mid Au_2 \mid \cdots \mid Au_n]$. This proves Lemma 1.2.5 **(a)**.

(b) Let $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be a diagonal $n \times n$-matrix.

We already know what happens to a matrix when we multiply it on the right by a diagonal matrix: Namely, in the solution of Exercise 1 on homework set #2 (applied to $A = U$ and $d_i = \lambda_i$), we have observed that the matrix $U \cdot \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is obtained from U by scaling the j-th column by λ_j for each $j \in \{1, 2, \ldots, n\}$. Thus, the columns of the matrix $U \cdot \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are the columns of U, scaled by $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. In other words, the columns of the matrix $U \cdot \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are $\lambda_1 u_1, \lambda_2 u_2, \ldots, \lambda_n u_n$ (since the columns of U are $u_1, u_2, \ldots, u_n$). In other words,

$$U \cdot \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = [\lambda_1 u_1 \mid \lambda_2 u_2 \mid \cdots \mid \lambda_n u_n].$$

In view of $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, this rewrites as $UD = [\lambda_1 u_1 \mid \lambda_2 u_2 \mid \cdots \mid \lambda_n u_n]$. This proves Lemma 1.2.5 **(b)**. $\square$

Let us now prove Proposition 1.2.3:

Proof of Proposition 1.2.3. **(a)** The Inverse Matrix Theorem (more precisely, the implication **(d)** $\Longrightarrow$ **(k)** from Theorem 1.2.1 in the notes from 2019-10-16) shows that the matrix U is invertible (since its n columns $u_1, u_2, \ldots, u_n$ form a basis of $\mathbb{R}^n$). The matrix D is diagonal. It remains to prove that $A = UDU^{-1}$. Equivalently, we need to prove that $AU = UD$.

Lemma 1.2.5 **(a)** yields

$$AU = [Au_1 \mid Au_2 \mid \cdots \mid Au_n] = [\lambda_1 u_1 \mid \lambda_2 u_2 \mid \cdots \mid \lambda_n u_n]$$
$$(\text{since } Au_i = \lambda_i u_i \text{ for each } i \text{ (because each } u_i \text{ is a } \lambda_i\text{-eigenvector)}).$$

On the other hand, Lemma 1.2.5 **(b)** yields

$$UD = [\lambda_1 u_1 \mid \lambda_2 u_2 \mid \cdots \mid \lambda_n u_n].$$

So we got the same expression for AU as for UD. Thus, $AU = UD$. Multiplying both sides of this equality by U^{-1} on the right, we obtain $A = UDU^{-1}$. This proves Proposition 1.2.3 **(a)**.

(b) The same argument, done in reverse, proves part **(b)**. (See [Strickland, proof of Proposition 14.4] for details.) $\square$

Some more examples of diagonalization of matrices:

Example 1.2.6. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. In Example 1.3.4 in the notes from 2019-11-06, we showed that the eigenvectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}$$

of A (for eigenvalues $1, 2, 3$) form a basis of $\mathbb{R}^3$. We can now use this basis to diagonalize A: Namely, Proposition 1.2.3 **(a)** yields that

$$U = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix} \qquad \text{and} \qquad D = \operatorname{diag}(1, 2, 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

form a diagonalization of A.

Thus, (2) yields that

$$A^m = U \operatorname{diag}(1^m, 2^m, 3^m)\, U^{-1} \qquad \text{for each } m \geq 0.$$

Example 1.2.7. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Does this A have a diagonalization?

In Example 1.3.5 in the notes from 2019-11-06, we found that there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of A. Hence, Proposition 1.2.3 **(b)** shows that there is no diagonalization of A.

So some matrices have a diagonalization (and usually many), while others don't.

In the case of the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ (which was used to compute the Fibonacci numbers), we were in luck, since our matrix A had a diagonalization. What would have happened if it didn't? Let's see:

Example 1.2.8. Define a sequence $(g_0, g_1, g_2, \ldots)$ of integers by

$$g_0 = 0, \qquad g_1 = 1, \qquad g_n = 2g_{n-1} - g_{n-2}.$$

This time, we get $\begin{pmatrix} g_m \\ g_{m+1} \end{pmatrix} = A \begin{pmatrix} g_0 \\ g_1 \end{pmatrix}$ for $A = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$.

The matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ has characteristic polynomial $t^2 - 2t + 1 = (t-1)^2$. Thus, its only eigenvalue is 1. The corresponding eigenvectors are nonzero multiples of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Thus, there is no basis of $\mathbb{R}^2$ consisting of eigenvectors of A. Hence, there is no diagonalization of A. So we cannot compute g_m by the same method that we used to find f_m.

However, there is a very simple formula for g_m: namely,

$$g_m = m \qquad \text{for each } m \geq 0.$$

You can prove this by induction on m.

There is a more general concept than diagonalization – the so-called **Jordan normal form** – that exists for **any** $n \times n$-matrix. (See [Hefferon, §Five.IV] or a typical course on Linear Algebra 2 or Abstract Algebra 1 or 2.)

1.3. Complex numbers

1.3.1. Motivation

Now, let us return to an example of a matrix that we have not managed to find any eigenvectors of:

Example 1.3.1. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The eigenvalues of A in $\mathbb{R}$ don't exist at all (as we have seen in Example 2.1.13 in classwork from 2019-11-04), since the characteristic polynomial $\chi_A(t) = t^2 + 1$ has no real roots (its plot stays strictly above the x-axis). But let us **pretend** that there is a new "mythical number" i, which is a square root of -1, in the sense that $i^2 = -1$. Then, of course, $(-i)^2 = -1$ as well, since $(-i)^2 = i^2$. (Here we are assuming that this "mythical number" i and its whole retinue of numbers derived from it satisfy the same rules as the familiar real numbers, such as the identity $(-x)^2 = x^2$.) Thus, $-i$ and i are roots of the polynomial $t^2 + 1$. Since a degree-2 polynomial should have only 2 roots (at least if these "mythical numbers" behave as nicely as our familiar real numbers), we thus have found all the roots of this polynomial.

Thus, the eigenvalues of A are the "mythical numbers" i and $-i$. We can find the corresponding eigenvectors using Gaussian elimination:

- The i-eigenvectors are the nonzero scalar multiples of $\begin{pmatrix} i \\ 1 \end{pmatrix}$.
- The $(-i)$-eigenvectors are the nonzero scalar multiples of $\begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Of course, these vectors $\begin{pmatrix} i \\ 1 \end{pmatrix}, \begin{pmatrix} -i \\ 1 \end{pmatrix}$ don't belong to $\mathbb{R}^2$, and thus cannot form a basis of $\mathbb{R}^2$. But if we redo all the linear algebra that we did using our new "mythical numbers" instead of real numbers as scalars, then they form a basis of the analogue of $\mathbb{R}^2$.

This should give a diagonalization $A = UDU^{-1}$ with

$$U = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad D = \operatorname{diag}(i, -i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

This looks neat, but you should have a lingering doubt: Why can we just introduce a new "number" i satisfying $i^2 = -1$?

After all, this isn't much different from introducing a new "number" ∞ satisfying $0 \cdot \infty = 1$. But ∞ quickly leads to contradictions, at least if you pretend that it behaves like a normal number (for example, $0 \cdot \underbrace{(0 \cdot \infty)}_{=1} = 0 \cdot 1 = 0$, but $\underbrace{(0 \cdot 0)}_{=0} \cdot \infty = 0 \cdot \infty = 1$, so the associative law no longer holds). Why does our mythical i not lead to contradictions?

1.3.2. Informal introduction

Before we address this existential doubt, let us experiment a bit with our new toy, leaving aside the question whether it has a rigorous meaning.

We have introduced a new "number" i that satisfies $i^2 = -1$. We call it the **imaginary unit**[1]. Using familiar operations such as addition, subtraction and multiplication, we can combine this new "number" i with our familiar real numbers, and obtain new "numbers" – which are called **complex numbers** (a formal definition will be given below). For example, $2 + 3i - 7i^3$ is a complex number; so is $i^5 - \pi i + \frac{1}{3}$. Let us calculate a bit with complex numbers (assuming that they satisfy the same laws as real numbers, such as associativity, commutativity and distributivity):

$$(1+i)(1-i) = 1 + i - i + i(-i) = 1 + i(-i) = 1 - \underbrace{i^2}_{=-1} = 1 - (-1) = 2;$$

$$i^4 = \Big(\underbrace{i^2}_{=-1}\Big)^2 = (-1)^2 = 1;$$

$$i^5 = \underbrace{i^4}_{=1} i = i;$$

$$(-i)^2 = (-i)(-i) = -(-ii) = ii = i^2 = -1;$$

$$(1+i) + (1-i) = 2;$$

$$\frac{1}{1+i} = \frac{1-i}{(1+i)(1-i)} = \frac{1-i}{2} = \frac{1}{2} - \frac{1}{2}i.$$

The number i is often called $\sqrt{-1}$ because its square is -1; but it is only one of two complex numbers whose square is -1 (namely, i and $-i$).

The above computations suggest that every complex number has the form $a + bi$ for some real numbers a and b. Let us see why: If we have two numbers of the form $a + bi$, then their sum, difference, product and quotient is also of this form. Namely,

$$(a+bi) + (c+di) = (a+c) + (b+d)i; \tag{4}$$

$$(a+bi) - (c+di) = (a-c) + (b-d)i; \tag{5}$$

$$\begin{aligned}(a+bi)(c+di) &= ac + bic + adi + bidi \\ &= ac + bci + adi + bd\underbrace{i^2}_{=-1} \\ &= ac + bci + adi - bd \\ &= (ac-bd) + (ad+bc)i;\end{aligned} \tag{6}$$

$$\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2},$$

[1] The name is probably referencing the fact that we just invented it. Arguably, all numbers are just as imaginary to some extent...

where the last equation requires $(c,d) \neq (0,0)$.

These formulas show that if we start with real numbers and the new "number" i, then the standard operations ($+$, $-$, $\cdot$ and $/$) do not take us out of the set of numbers of the form $a+bi$. Thus, all complex numbers have the form $a+bi$ for $a,b \in \mathbb{R}$.

We have assumed that all the standard rules for real numbers (associativity, commutativity, distributivity) still hold for these new numbers. What about inequalities?

Here we are in for a disappointment. If inequalities ($\geq$, $>$, $\leq$, $<$) would still make sense for complex numbers (and behave anything like they do for real numbers), then we would get the following contradiction:

- If $i \geq 0$, then $i^2 \geq 0$, contradicting $i^2 = -1$.
- If $i < 0$, then $-i > 0$ and thus $i^2 = (-i)^2 > 0$ (since $-i > 0$), contradicting $i^2 = -1$.

So i is neither ≥ 0 nor < 0. Thus, inequalities do not work for our new numbers.

So how can we be sure that the other things (addition, subtraction, multiplication) work and don't lead to contradictions?

1.3.3. Rigorous definition of complex numbers

The trick is to define complex numbers rigorously (using existing objects such as real numbers), instead of obtaining them by introducing a mythical "number" i whose existence has to be taken for granted. How can we do this? We forget about the mythical i, and instead define complex numbers as pairs of real numbers. Don't worry – we will gain our i back in a moment, and it will be a rigorously defined i rather than some mythical object of unclear existence status.

Here is the promised rigorous definition of complex numbers and the most basic operations (addition, subtraction and multiplication[2]) on them:[3]

Definition 1.3.2. (a) A **complex number** is defined as a pair (a,b) of two real numbers.

(b) We let $\mathbb{C}$ be the set of all complex numbers.

(c) For each real number r, we denote the complex number $(r,0)$ as $r_{\mathbb{C}}$ (and we will later equate it with r).

(d) We let i be the complex number $(0,1)$.

(e) We define three binary operations $+$, $-$ and $\cdot$ on $\mathbb{C}$ by setting

$$(a,b) + (c,d) = (a+c, b+d), \tag{7}$$

$$(a,b) - (c,d) = (a-c, b-d), \tag{8}$$

[2]We will define division and exponentiation (with integer exponents) later.

[3]See [19s, Section 4.1] for more details on this.

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc). \tag{9}$$

(Thus, the operations $+$ and $-$ are just entrywise addition and subtraction, just as for row vectors. The operation $\cdot$ is more complicated, and we will soon see why we have defined it in this particular way.)

(f) As usual, we write $\alpha\beta$ for $\alpha \cdot \beta$ if α and β are complex numbers.

As usual, we write $-\alpha$ for $0_{\mathbb{C}} - \alpha$ if α is a complex number.

For example, this definition yields

$$\underbrace{i}_{=(0,1)} \underbrace{i}_{=(0,1)} = (0,1) \cdot (0,1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0) = (-1)_{\mathbb{C}}.$$

So we get $i^2 = -1$, at least if we abbreviate ii as i^2.

If you compare this rigorous definition of complex numbers with the informal introduction we gave before it, you will now find that they lead to the same notion of complex numbers: The pair (a, b) from the rigorous definition corresponds to the $a + bi$ from the informal introduction. The operations $+$, $-$ and $\cdot$ on the former pairs correspond precisely to the operations $+$, $-$ and $\cdot$ on the latter "numbers", because the equalities (7), (8) and (9) are precisely the equalities (4), (5) and (6) rewritten in terms of the rigorous definition. (This explains why we have chosen these exact definitions of $+$, $-$ and $\cdot$ and no others – we chose them in order to recover the equalities (4), (5) and (6).)

So we know what complex numbers are and how to add, subtract and multiply them. Do these operations behave well? Yes, as the following theorem shows:

Theorem 1.3.3. All standard rules for addition, subtraction and multiplication (e.g., commutativity, associativity, distributivity) hold for complex numbers.

Proof. This is all straightforward. See [19s, Theorem 4.1.2]. □

The complex numbers include copies of our old real numbers, because of the following theorem:

Theorem 1.3.4. The complex numbers $r_{\mathbb{C}}$ (for $r \in \mathbb{R}$) behave exactly as the real numbers r: For example,

$$(r + s)_{\mathbb{C}} = r_{\mathbb{C}} + s_{\mathbb{C}};$$
$$(r - s)_{\mathbb{C}} = r_{\mathbb{C}} - s_{\mathbb{C}};$$
$$(rs)_{\mathbb{C}} = r_{\mathbb{C}} s_{\mathbb{C}}$$

for any two real numbers r and s.

Thus, we can identify each real number r with the corresponding complex number $r_{\mathbb{C}} = (r, 0)$. Let us do so; thus, the complex numbers become an extension of the real numbers. [4]

Now that we have agreed to identify each $r \in \mathbb{R}$ with $r_{\mathbb{C}}$, we have

$$(a, b) = a + bi \qquad \text{for every } (a, b) \in \mathbb{C}.$$

This is because

$$\begin{aligned}
&\underbrace{a}_{\substack{=a_{\mathbb{C}} \\ =(a,0)}} + \underbrace{b}_{\substack{=b_{\mathbb{C}} \\ =(b,0)}} \underbrace{i}_{=(0,1)} \\
&= (a, 0) + (b, 0)\,(0, 1) = (a, 0) + (b \cdot 0 - 0 \cdot 1, b \cdot 1 + 0 \cdot 0) \\
&= (a, 0) + (0, b) = (a + 0, 0 + b) = (a, b)\,.
\end{aligned}$$

We can define quotients of complex numbers (i.e., we can divide a complex number by any other, as long as the latter is nonzero): Given two complex numbers α and $\beta \neq 0$, we can define $\dfrac{\alpha}{\beta}$ to be the unique complex number γ such that $\alpha = \beta\gamma$. This exists, because

$$\frac{a + bi}{c + di} = \frac{(a + bi)\,(c - di)}{(c + di)\,(c - di)} = \frac{(ac + bd) + (bc - ad)\,i}{c^2 + d^2}.$$

We can define integer powers of complex numbers: If $n \in \mathbb{Z}$ and $\alpha \in \mathbb{C}$, then

$$\begin{aligned}
\alpha^n &= \underbrace{\alpha\alpha \cdots \alpha}_{n \text{ times}} && \text{if } n > 0; \\
\alpha^0 &= 1; \\
\alpha^n &= \left(\frac{1}{\alpha}\right)^{-n} && \text{if } n < 0.
\end{aligned}$$

The latter equality requires α to be nonzero. In particular, $i^2 = ii = -1$.

The main disadvantage of complex numbers compared to real numbers is the lack of inequalities. There is no reasonable way to define what it means for a complex number to be $\leq$ to another complex number. (In fact, we have already seen how that would lead to contradictions.)

Also, there is no good way to define non-integer powers of complex numbers (for example, $i^{\sqrt{2}}$ does not have a well-defined meaning). (Some people would give you a definition, but it would depend on whom you ask, and it would probably not satisfy standard rules like $\alpha^{\beta\gamma} = \left(\alpha^{\beta}\right)^{\gamma}$.)

[4] Theorem 1.3.4 shows that this identification does not mess up our algebraic operations! If this theorem was false, then equating r with $r_{\mathbb{C}}$ would make expressions like $r + s$, $r - s$ and rs ambiguous, as their meaning would depend on whether we are adding/subtracting/multiplying the real numbers r and s or the corresponding complex numbers $r_{\mathbb{C}}$ and $s_{\mathbb{C}}$.

1.3.4. The Argand diagram

There is a visual way to represent complex numbers. I will introduce it briefly; see [19s, Subsection 4.1.8 and further] for more details (and pictures!).

As you know, the real numbers are typically viewed as points on a line. We similarly represent complex numbers as points on the plane. Namely, the complex number $(a,b) = a + bi$ (with $a, b \in \mathbb{R}$) will be represented by the point (a, b) (in Cartesian coordinates). This way of representing complex numbers called the **Argand diagram**.

Note that addition and subtraction of complex numbers (a, b) are defined in the same way as for vectors: entrywise. Therefore, the sum of two complex numbers can be found on the Argand diagram by the parallelogram rule, just as the sum of two vectors. What about multiplication?

Multiplying a complex number by a real number is easy: If $a, b, c \in \mathbb{R}$, then

$$\underbrace{a}_{\substack{=a_{\mathbb{C}} \\ =(a,0)}} (b,c) = (a,0)(b,c) = (ab - 0c, ac + 0b) = (ab, ac).$$

Thus, multiplying a complex number by a real number a is simply multiplying its both entries by a. On the Argand diagram, this corresponds to a homothety with center at the origin (i.e., scaling).

How do we multiply two complex numbers?

To give a description, we introduce some notations:

Definition 1.3.5. Let z be a complex number. Write z in the form $z = (a, b) = a + bi$ for two real numbers a, b.

Let P_z be the point corresponding to the complex number z on the Argand diagram. (As you remember, this is the point with coordinates (a, b).)

(a) The real numbers a and b are called the **real part** and the **imaginary part** of z. They are the Cartesian coordinates of the point P_z.

(b) The **absolute value** of z is defined to be the real number $\sqrt{a^2 + b^2}$. This is the distance between the origin and P_z. This is the first (radial) **polar coordinate** of P_z. It is denoted by $|z|$.

(c) Assume that $z \neq 0$. Then, consider the angle φ (with $-\pi < \varphi \leq \pi$) at which the ray from 0 to P_z stands to the ray from 0 to 1 (i.e., the positive half-axis). This is the second (angular) **polar coordinate** of P_z. It is denoted by $\arg z$, and is called the **argument** of z.

Example 1.3.6. We have

$$|1 + i| = |(1,1)| = \sqrt{1^2 + 1^2} = \sqrt{2} \qquad \text{and} \qquad \arg(1 + i) = \pi/4 = 45^\circ.$$

Similarly,

$$|1 - i| = |(1,-1)| = \sqrt{1^2 + (-1)^2} = \sqrt{2} \qquad \text{and} \qquad \arg(1 - i) = -\pi/4 = -45^\circ.$$

References

[19s] Darij Grinberg, *Introduction to Modern Algebra (UMN Spring 2019 Math 4281 notes)*, 29 June 2019.
http://www.cip.ifi.lmu.de/~grinberg/t/19s/notes.pdf

[AndAnd14] Titu Andreescu, Dorin Andrica, *Complex Numbers from A to... Z*, 2nd edition, Springer 2014.

[Hefferon] Jim Hefferon, *Linear Algebra*, 3rd edition, 1 January 2017, http://joshua.smcvt.edu/linearalgebra/ .

[Strickland] Neil Strickland, *MAS201 Linear Mathematics for Applications*, lecture notes, 28 September 2013.
http://neil-strickland.staff.shef.ac.uk/courses/MAS201/

[Swanso18] Irena Swanson, *Introduction to Analysis*, with construction of the number systems, 19 June 2019.
https://people.reed.edu/~iswanson/analysisconstructR.pdf