Fall 2019 Math 201-003 at Drexel: blackboard notes of 2019-11-06

Darij Grinberg

November 12, 2019

1. Eigenvalues and eigenvectors ("eigenstuff")

1.1. Definition and examples (recall)

Recall some material from last time:

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 $\chi_A(t)$.

Proposition 1.1.4. If a matrix A is triangular, then its eigenvalues are its diagonal entries.

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.

1.2. The characteristic polynomial, explicitly

The following theorem gives a more-or-less explicit formula for the coefficients of the characteristic polynomial of a matrix. It is **not** a great way of computing them (too much work), but it can help double-check them.

Theorem 1.2.1. Let A be an $n \times n$-matrix. Then, its characteristic polynomial $\chi_A(t)$ is a polynomial of degree n, and equals

$$\chi_A(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n,$$

where

- $c_0 = \det A$.
- $c_n = (-1)^n$.
- $c_{n-1} = (-1)^{n-1} \operatorname{Tr} A$. (Recall: $\operatorname{Tr} A$ is the trace of A, as defined on homework set #1.)
- For each $k \in \{0, 1, \ldots, n\}$, we have

$$c_k = (-1)^k \cdot \left(\begin{array}{c} \text{the sum of the determinants of all submatrices} \\ \text{of } A \text{ obtained from } A \text{ by removing } k \text{ rows} \\ \text{and the corresponding } k \text{ columns} \end{array} \right).$$

  (Here, "corresponding k columns" means "the k columns with the same indices as the removed k rows". So, for example, if we remove rows 2 and 5, then we must remove columns 2 and 5.)

Instead of a formal proof, let us confirm this theorem on an example:

Proof of Theorem 1.2.1 for $n = 3$. Assume that $n = 3$. Write A as $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$.

Then, the definition of $\chi_A(t)$ yields

$$\begin{aligned} \chi_A(t) &= \det(A - tI_n) = \det \begin{pmatrix} a-t & b & c \\ d & e-t & f \\ g & h & i-t \end{pmatrix} \\ &= (a-t)(e-t)(i-t) + bfg + cdh - (a-t)fh - c(e-t)g - bd(i-t) \\ &= \left(aei - (ae + ai + ei)t + (a+e+i)t^2 - t^3 \right) \\ &\qquad + bfg + cdh + (afh - fht) - (ceg - cgt) - (bdi - bdt) \end{aligned}$$

$$= \underbrace{\left(aei + bfg + cdh - afh - ceg - bdi\right)}_{\substack{=\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \\ =\det A}}$$
$$+ \underbrace{\left(ae + ai + ei - fh - cg - bd\right)}_{\substack{=(ae-bd)+(ai-cg)+(ei-fh) \\ =\left(\begin{array}{c}\text{the sum of the determinants of all} \\ \text{submatrices obtained from } A \text{ by removing} \\ \text{one row and the corresponding column}\end{array}\right)}} t$$
$$+ \underbrace{(a+e+i)}_{=\operatorname{Tr} A} t^2 - t^3.$$

This confirms the theorem for $n = 3$. □

Remark 1.2.2. Many authors define $\chi_A(t)$ not as $\det(A - tI_n)$ (as we did), but instead as $\det(tI_n - A)$. This does not change much (in particular, this does not change the roots of $\chi_A(t)$), because

$$\det(tI_n - A) = (-1)^n \cdot \det(A - tI_n).$$

Proof. Recall the fact (Corollary 1.5.3 in classwork from 2019-10-30) that says that if B is an $n \times n$-matrix, and if λ is a scalar, then $\det(\lambda B) = \lambda^n \det B$. Applying this to $\lambda = -1$ and $B = A - tI_n$, we obtain $\det((-1)(A - tI_n)) = (-1)^n \cdot \det(A - tI_n)$. Since $(-1)(A - tI_n) = tI_n - A$, this rewrites as $\det(tI_n - A) = (-1)^n \cdot \det(A - tI_n)$. □

1.3. Linear independence of eigenvectors

Eigenvectors become particularly useful when we have many of them and they are linearly independent. Let us take this for a given for now (we will later see how they are useful), and see some ways to quickly get lots of linearly independent eigenvectors.

We begin with the following fact ([Strickland, Proposition 13.19]):

Proposition 1.3.1. 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 be a λ_i-eigenvector of A. Then, $v_1, v_2, \ldots, v_k$ are linearly independent.

Proof. See [Strickland, Proposition 13.19]. □

Proposition 1.3.1 says that a list of eigenvectors of the same matrix is automatically linearly independent if they belong to distinct eigenvalues. There is an even stronger result ([Strickland, Remark 13.21], proved in his exercises), which shows linear independence even if some of the eigenvectors belong to the same eigenvalue, as long as the eigenvectors to each eigenvalue are linearly independent among themselves. Here is the precise statement:

Proposition 1.3.2. 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.

The next fact is [Strickland, Proposition 13.22]:

Proposition 1.3.3. 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 $\mathbb{R}^n$.

Proof. Proposition 1.3.1 yields that $v_1, v_2, \ldots, v_n$ are linearly independent. But they are n vectors in $\mathbb{R}^n$, so as we have learned, this means that they form a basis of $\mathbb{R}^n$. $\square$

How often is the assumption of the proposition satisfied? In other words, how likely is it for an $n \times n$-matrix A to have n distinct eigenvalues? Here are some informal answers:

- If we content ourselves with **complex** eigenvalues, then "almost every" $n \times n$-matrix A has n distinct eigenvalues. In other words, a "randomly chosen" $n \times n$-matrix A has n distinct eigenvalues. This is not a rigorous statement at this point.

- For a matrix A that comes from a mathematical problem (a matrix with "meaning"), it may very well happen that A does not have n distinct eigenvalues. Some of the most "interesting" matrices do not have n distinct eigenvalues.

- If we want real eigenvalues, it happens fairly often that an $n \times n$-matrix has less than n of them. It may even happen that it has none.

For example, here is a "randomly chosen" 3×3-matrix: $\begin{pmatrix} 5 & 1 & 3 \\ 9 & 7 & 6 \\ 2 & -3 & 0 \end{pmatrix}$. Its characteristic polynomial: $-t^3 + 12t^2 - 38t - 21$. Its roots (i.e., the eigenvalues of the matrix) are (approximately) $6.2389 - 2.2444i$, $6.2389 + 2.2444i$, -0.4777. The first two are non-real complex numbers; only the third is a real number. There are no good exact formulas for these roots, because $-t^3 + 12t^2 - 38t - 21$ is an unremarkable degree-3 polynomial. (There is a very messy formula that involves both square and cube roots, but you really don't want to use that one.)

Let us do a few examples where we compute the eigenvalues and eigenvectors of a matrix A and attempt to form a basis out of the eigenvectors.

Example 1.3.4. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. Then, the eigenvalues of A are $1, 2, 3$ (by Proposition 1.1.4). Thus, Proposition 1.3.3 says that if we pick one eigenvector for each eigenvalue, we obtain a basis of $\mathbb{R}^3$. Let us do this:

- Let us find the 1-eigenvectors. These are the nonzero vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ such that $A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1 \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Solving this, we find the eigenvector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

- Let us find the 2-eigenvectors. These are the nonzero vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ such that $A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2 \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Solving this, we find the eigenvector $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

- Let us find the 3-eigenvectors. These are the nonzero vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in$

$\mathbb{R}^3$ such that $A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 3\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Solving this, we find the eigenvector $\begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}$.

Now, we have found three eigenvectors:

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

Since these are eigenvectors for distinct eigenvalues, Proposition 1.3.3 says that they form a basis of $\mathbb{R}^3$.

Example 1.3.5. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Then, $\chi_A(t) = (1-t)^3$. Thus, the only eigenvalue is 1. Thus, we cannot apply Proposition 1.3.3. But if we can find 3 linearly independent 1-eigenvectors, then we can apply Proposition 1.3.2 (or simply recall that 3 linearly independent vectors in $\mathbb{R}^3$ must always form a basis of $\mathbb{R}^3$). But can we find them?

The 1-eigenvectors of A are the nonzero vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ such that $A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. These are the nonzero scalar multiples of $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Thus, any two of them are proportional. Hence, there are no 3 linearly independent 1-eigenvectors of A (or even 2 such). Thus, there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of A.

Example 1.3.6. Let $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. Then,

$$\chi_A(t) = \det(A - tI_n) = \det\begin{pmatrix} 0-t & 0 & 1 \\ 0 & 1-t & 0 \\ 1 & 0 & 0-t \end{pmatrix} = (1-t)\underbrace{\det\begin{pmatrix} 0-t & 1 \\ 1 & 0-t \end{pmatrix}}_{\substack{=(0-t)^2-1^2 \\ =(t+1)(t-1)}}$$

(by Laplace expansion along the 2-nd row)

$$= (1-t)(t+1)(t-1) = -(t-1)^2(t+1).$$

So the eigenvalues of A are 1 and -1. So Proposition 1.3.3 does not give us a basis of eigenvectors (since 1 and -1 are only 2 eigenvalues, not 3). However, we still have a chance of finding such a basis, if we find enough linearly independent eigenvectors.

- The 1-eigenvectors of A are the nonzero $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ such that $A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1 \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. These are all nonzero vectors of the form $\begin{pmatrix} z \\ y \\ z \end{pmatrix}$. We can thus find two linearly independent 1-eigenvectors: namely, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

- The (-1)-eigenvectors of A are the nonzero $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ such that $A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (-1) \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. These are all nonzero vectors of the form $\begin{pmatrix} -z \\ 0 \\ z \end{pmatrix}$. Thus, we can pick $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ as one of them.

Altogether, we can now conclude from Proposition 1.3.2 that

$$\underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}}_{\text{independent 1-eigenvectors}}, \underbrace{\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}}_{(-1)\text{-eigenvector}}$$

form a basis of $\mathbb{R}^3$ consisting of eigenvectors. (Of course, you can easily check this without invoking Proposition 1.3.2.)

Example 1.3.7. 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). Thus, there is no basis of $\mathbb{R}^2$ that consists of eigenvectors of A. (But we can fix this by extending our number system to the so-called **complex numbers**, which we will see next time.)

Remark 1.3.8. Let n be odd. Then, every $n \times n$-matrix with real entries has **at least one** real eigenvalue.

Proof. Let A be an $n \times n$-matrix. Then, its characteristic polynomial $\chi_A(t)$ has odd degree (since it has degree n, which is odd). But any real polynomial of odd degree has at least one real root. Thus, A has at least one real eigenvalue. □

1.4. Diagonalization

1.4.1. Motivation

Example 1.4.1. The **Fibonacci numbers** are a sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers. They are defined recursively by

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

Here is a table of the first 9 Fibonacci numbers:

m	0	1	2	3	4	5	6	7	8
f_m	0	1	1	2	3	5	8	13	21

.

You might wonder: Is there an explicit formula for f_m ? Is there a faster way to compute f_m than working one's way up recursively? How fast does f_m grow?

One way to approach this is using matrices: Instead of computing f_m, let us look at the vectors $\begin{pmatrix} f_m \\ f_{m+1} \end{pmatrix}$ for $m \geq 0$. The nice thing about these vectors is that each of them determines the next one:

$$\begin{pmatrix} f_{m+1} \\ f_{m+2} \end{pmatrix} = \begin{pmatrix} f_{m+1} \\ f_{m+1} + f_m \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_m \\ f_{m+1} \end{pmatrix} = A \begin{pmatrix} f_m \\ f_{m+1} \end{pmatrix},$$

where $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Thus, for example,

$$\begin{aligned} \begin{pmatrix} f_5 \\ f_6 \end{pmatrix} &= A \begin{pmatrix} f_4 \\ f_5 \end{pmatrix} = AA \begin{pmatrix} f_3 \\ f_4 \end{pmatrix} = AAA \begin{pmatrix} f_2 \\ f_3 \end{pmatrix} = AAAA \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \\ &= AAAAA \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} = A^5 \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}. \end{aligned}$$

More generally,

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

(Formally speaking, you can prove this by induction on m.) Thus, in order to compute f_m, it suffices to compute A^m.

Now, how can we easily compute the m-th power of a matrix?

Let's look at diagonal matrices first. The product of two diagonal matrices is given by simply multiplying their diagonal entries:

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix} = \begin{pmatrix} ac & 0 \\ 0 & bd \end{pmatrix}.$$

Thus, by induction, we see that a similar rule holds for taking a diagonal matrix to a power:

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^m = \begin{pmatrix} a^m & 0 \\ 0 & b^m \end{pmatrix} \qquad \text{for each } m \geq 0.$$

But our matrix A is not diagonal! So this doesn't directly help us find A^m.

But let's extend this trick to matrices that are not diagonal in themselves, but have the form

$$UDU^{-1} \qquad \text{where } D \text{ is diagonal and } U \text{ is invertible.}$$

For example:

$$\left(UDU^{-1}\right)^2 = UD\underbrace{U^{-1}U}_{=I}DU^{-1} = U\underbrace{DD}_{=D^2}U^{-1} = UD^2U^{-1};$$

$$\left(UDU^{-1}\right)^3 = UD\underbrace{U^{-1}U}_{=I}D\underbrace{U^{-1}U}_{=I}DU^{-1} = U\underbrace{DDD}_{=D^3}U^{-1} = UD^3U^{-1};$$

$$\left(UDU^{-1}\right)^4 = UD\underbrace{U^{-1}U}_{=I}D\underbrace{U^{-1}U}_{=I}D\underbrace{U^{-1}U}_{=I}DU^{-1} = U\underbrace{DDDD}_{=D^4}U^{-1} = UD^4U^{-1}.$$

You can probably see how this sequence of equalities goes on; the result is the following:[1]

Proposition 1.4.2. 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

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

Proof. The first equality is easily proved by induction on m. $\square$

[1]The notation $\operatorname{diag}(p_1, p_2, \ldots, p_n)$ stands for the diagonal $n \times n$-matrix whose diagonal entries are $p_1, p_2, \ldots, p_n$ (from top-left to bottom-right). For example, $\operatorname{diag}(2, 9, 4) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

Therefore, if we can write our matrix A in the form $A = UDU^{-1}$ for an invertible U and a diagonal D, then we can easily compute any power of A. In particular, if we can do this for $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, then we can find an explicit formula for A^m and thus for the Fibonacci number f_m.

1.4.2. The method

So how do we write a square matrix A as UDU^{-1} ? Is it possible at all?

This is known as **diagonalization**:

Definition 1.4.3. 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 $A = UDU^{-1}$.

Note that $A = UDU^{-1}$ can be rewritten equivalently as $U^{-1}AU = D$.

Later, we will see what diagonalizing a matrix "really means": A diagonalization of A is an "alternative coordinate system" in which A "becomes a diagonal matrix". For now, think of this as a vague idea.

How do we find a diagonalization of a matrix A ? The following fact ([Strickland, Proposition 14.4]) is crucial:

Proposition 1.4.4. 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 will prove and apply this next time.

References

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