Eigenvalues and eigenvectors

A. Eremenko

October 3, 2024

Let $F : V \to V$ be a linear operator on a vector space V. When V is finite-dimensional, we know how to represent F by a matrix. For this we need a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. We apply our operator to each vector of the basis, and expand the result in the same basis. The matrix A of F consists of the coefficients $(a_{i,j})$ of this expansion, namely the element $a_{i,j}$ is the coefficient at $\mathbf{v}_i$ of $F(v_j)$:

$$F(\mathbf{v}_j) = \sum_{i=1}^{n} a_{i,j}\mathbf{v}_i.$$

So A is a square matrix $n \times n$ where $n = \dim V$. For a given operator, the matrix A depends on the choice of basis, and the question we are going to address, is how to choose a basis for a given operator, so that the matrix has the simplest form. This will allow us to better understand the action of a linear operator.

Let us look for vectors, on which F acts in the simplest possible way:

$$F(\mathbf{v}) = \lambda\mathbf{v}, \quad v \neq 0, \tag{1}$$

where λ is a number. Such vectors are called *eigenvectors*, and corresponding λ is an *eigenvalue*. We will discuss the *eigenvalue-eigenvector problem*: how to find all eigenvalues and eigenvectors of a given operator. When $V = \mathbf{R}^n$ we can choose the standard basis and assume that $F(\mathbf{x}) = A\mathbf{x}$, so our problem is for a given square matrix to find all numbers λ and vectors $\mathbf{v}$ which satisfy

$$A\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \tag{2}$$

definition is always satisfied by $\mathbf{v} = \mathbf{0}$ with any λ. So eigenvector cannot be $\mathbf{0}$ by definition, but $\lambda = 0$ is fine, as any other number.

To approach this problem, we rewrite our equation as

$$(A - \lambda I)\mathbf{v} = \mathbf{0} \tag{3}$$

This means that the homogeneous equation with matrix $A - \lambda I$ has a non-trivial solution. We know a criterion for this:

$$\det(A - \lambda I) = 0. \tag{4}$$

Notice, that by using the determinant we split the problem: in (4) there is no $\mathbf{v}$, this is an equation in λ which allows us in principle to find all eigenvalues of A. Once an eigenvalue λ is known, the rest of the task is solving (3) for $\mathbf{v}$, and this is simple, and we know how to do this. Notice that the set of all eigenvectors corresponding to a fixed λ is a vector space $N(A - \lambda I)$; it is called the *eigenspace corresponding to* λ.

For example, the unit matrix has one eigenvalue, namely 1, and the eigenspace corresponding to this eigenvalue is the whole space.

Now we address the equation (4), which is the hard part of the business. It has a name: the *characteristic equation* of A.

First we notice that this is a *polynomial equation* with respect to λ, since the determinant of a matrix is a polynomial of its entries. Let us write it in the extended form:

$$\begin{vmatrix} a_{1,1} - \lambda & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} - \lambda & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} - \lambda \end{vmatrix} = 0.$$

Now recall the formula for the determinant: it is an alternating sum of products, each product contains n multiples: one element from each row and one from each column. It follows that only one of these products can contain λ^n and λ^{n-1}, namely the product

$$(a_{1,1} - \lambda)(a_{2,2} - \lambda) \dots (a_{n,n} - \lambda)$$

Multiplication gives

$$\det(A - \lambda I) = (-1)^n \lambda^n + (-1)^{n-1}(a_{1,1} + \dots + a_{n,n} + \dots.$$

Therefore:

1. $\det(A - \lambda I)$ is a polynomial of degree exactly n, with the coefficient at the top degree $(-1)^n$.

2. the coefficient at λ^{n-1} is $(-1)^{n-1}$ times the sum of the elements of the main diagonal of A.

It is also easy to see what the constant term of this polynomial is; for this we just plug $\lambda = 0$ to (4). So

3. The constant term of the characteristic polynomial is $\det(A)$.

The coefficient mentioned in statement 2 has a name, it is called the *trace* of the matrix,

$$\operatorname{tr}(A) = a_{1,1} + \ldots + a_{n,n}.$$

Like the determinant, the trace has some nice and useful properties, which will be discussed later. Especially simple expression we obtain for 2×2 matrices:

$$\det(A - \lambda I) = \lambda^2 - \operatorname{tr}(A)\lambda + \det(A).$$

So we have the problem of solving a polynomial equation. This is easy for $n = 2$ case, since there is a simple explicit formula. For $n = 3, 4$ such formulas also exist but unfortunately they are complicated, and really almost useless. For $n \geq 5$ no general algebraic formula exists.

Therefore, it remains to do the following:

a) Use numerical methods to obtain approximations of eigenvalues,

b) study their properties qualitatively(we will see how their qualitative properties are reflected in the meaningful and important properties of the linear operator).

c) Look for special important matrices which arise in applications and whose eigenvalues can be explicitly found.

All three directions are actually big areas of research in mathematics.

Let us recall the main facts about polynomial equations. Let $P(\lambda)$ be a polynomial of degree exactly n. If we know one root λ_1 then the polynomial is divisible by $(\lambda - \lambda_1)$ that is

$$P(\lambda) = (\lambda - \lambda_1)P_1(\lambda), \quad \deg P_1 = n - 1.$$

(In general, when we multiply polynomials, their degrees are added). The devision is performed by a simple algorithm which you probably studied in high school. Now we have the

Fundamental Theorem of Algebra. *Every polynomial (with real or complex coefficients) of degree ≥ 1 has a complex root.*

This means that when we are using complex numbers, every polynomial of degree n factors into factors of degree 1:

$$P(\lambda) = c(\lambda - \lambda_1)\dots(\lambda - \lambda_n),$$

where c is a constant (the coefficient at λ^n). Even when the polynomial has real coefficients, the roots can be complex (non-real), this explains why we need to use complex numbers, even when solving problems about real matrices.

The roots λ_j are not necessarily distinct; grouping the equal multiples we obtain

$$P(\lambda) = c(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2}\dots(\lambda - \lambda_k)^{m_k} \quad m_1 + \dots + m_k = n.$$

These integers m_k are called *multiplicities* of roots. Once you know a root, it is easy to find its multiplicity: if λ_1 is a root of multiplicity m, then

$$f(\lambda_1) = f'(\lambda_1) = \dots = f^{(m-1)}(\lambda_1) = 0 \quad \text{but} \quad f^{(m)}(\lambda_1) \neq 0.$$

We are frequently interested in real matrices; their characteristic polynomials are real (have real coefficients), and for such polynomials, the following observation is useful: *If λ_1 is a root of multiplicity m of a real polynomial P then the complex conjugate $\overline{\lambda_1}$ is also a root of multiplicity m.* So non-real roots of real polynomial come in complex conjugate pairs.

Exercise. Show that the product of eigenvalues (taken with multiplicities that is an eigenvalue of multiplicity m participates m times in this product) equals to the determinant, and the sum of eigenvalues (also with multiplicities) is $(-1)^{n-1}$ times the trace.

Examples.

1. For the matrix

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

the characteristic polynomial is

$$\lambda^2 - \mathrm{tr}(A)\lambda + \det(A) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1).$$

so the eigenvalues are $\lambda_1 = 3$ and $\lambda_1 = -1$. Eigenvectors $\mathbf{v}$ for λ_1 satisfy

$$0 = (A - \lambda_1 I)\mathbf{v} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \mathbf{v},$$

whose general solution is $(t, t)^T$, so the eigenspace is one-dimensional, and its basis can be taken as $(1, 1)^T$.

Eigenvector $\mathbf{v}$ for λ_2 satisfy

$$0 = (A - \lambda_2 I)\mathbf{v} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \mathbf{v},$$

so the eigenspace is also one-dimensional and a basis is $(-1, 1)^T$. Notice that $(1, 1)$ and $(-1, 1)$ are linearly independent, so they form a basis of $\mathbf{R}^2$.

2. Suppose that $A = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ and all d_j are distinct. Then the eigenvalues are $d_1, \ldots, d_n$, each eigenspace is one-dimensional and a basis consists of $e_1, \ldots, e_n$, the standard basis in $\mathbf{R}^n$. If there is a group of m equal d_j, then the corresponding eigenspace will have dimension m.

3. For the unit matrix, the only eigenvalue is 1, it is the root of the characteristic polynomial of multiplicity n. The eigenspace is n-dimensional and it is equal to $\mathbf{R}^n$.

Control question: what are the eigenvalues and eigenvectors of the matrix 5?

The following general observations show that Example 1 is typical in some sense: for a generic $n \times n$ matrix we expect n distinct eigenvalues, and each eigenspace is one-dimensional. Indeed, for a generic polynomial one can expect n distinct roots (of multiplicity 1), so we have n distinct eigenvalues. Now we have

Theorem 1. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Let $\mathbf{v}_1, \ldots,$ be eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_k$, respectively, and suppose that all λ_j are distinct. Suppose that v_j are linearly dependent, and consider the shortest non-trivial linear combination of them which is equal to $\mathbf{0}$ Suppose that it is

$$c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = \mathbf{0}, \tag{5}$$

then all $c_j \neq 0$. Now apply A to (3). We obtain

$$c_1\lambda_1\mathbf{v}_1 + \ldots + c_k\lambda_k\mathbf{v}_k = 0. \tag{6}$$

Now multiply (5) by λ_1 and subtract from (6):

$$c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + \ldots + c_k(\lambda_k - \lambda_1)\mathbf{v}_k = \mathbf{0}$$

This is a non-trivial linear combination since $\lambda_1 \neq \lambda_j$ for $j \geq 2$. But it is shorter than the original linear combination (3). This contradiction proves the Theorem.

So in the case of $n = \dim \mathbf{R}^n$ *distinct* eigenvalues, all eigenspaces are of dimension 1, and we can construct a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ consisting of eigenvectors. Suppose that the corresponding eigenvalues are $\lambda_1, \ldots, \lambda_n$. Consider the matrix $B = [\mathbf{v}_1, v_2, \ldots, v_n]$ consisting of eigenvectors as columns. Then

$$AB = [A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n] = [\lambda_1\mathbf{v}_1, \lambda_2\mathbf{v}_2, \ldots, \lambda_n\mathbf{v}_n] = B\Lambda,$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Since B is non-singular (its columns are linearly independent), this can be rewritten as

$$A = B\Lambda B^{-1}, \tag{7}$$

or

$$\Lambda = B^{-1}AB.$$

Now recall that A represents the linear operator, $F(\mathbf{x}) = A\mathbf{x}$, and Λ is *the matrix of this linear operator in the basis* $\mathbf{v}_1, \ldots, \mathbf{v}_n$. So we completed our task: *we found a basis in which the matrix of the given operator is especially simple, namely diagonal.* This is possible when there exists a basis of the whole space composed of eigenvectors of our operator.

For this it is sufficient (but not necessary!) that the operator F has n distinct eigenvalues.

Compare with the content of the lecture on Linear transformation, where we derived a rule of the change of the basis.

Operators (and matrices) for which (7) holds are called diagonalizable. These are exactly those operators for which there exists a basis of the whole space consisting of eigenvectors. In general, two square matrices A and C are called *similar* if there exists a non-singular matrix B such that $A = B^{-1}CB$. Such two matrices can be thought as representing the same operator in two different bases. Notice that the characteristic polynomial $\det(A-\lambda I)$ is invariant under similarity: similar matrices have the same characteristic polynomial. This follows from the properties of determinants:

$$\det(B^{-1}AB - \lambda I) = \det(B^{-1}(A - \lambda I)B) = \det(A - \lambda I).$$

As a consequence we obtain that trace, determinant and eigenvalues are invariant under similarity: similar matrices have the same characteristic polynomial, same trace, same determinant and same eigenvalues (including multiplicity).

Unfortunately, not all operators (matrices) are diagonalizable. Here is the simplest example

$$\begin{pmatrix} c & 1 \\ 0 & c \end{pmatrix}.$$

the characteristic equation is $(c-\lambda)^2 = 0$, so the only eigenvalue is $\lambda = c$, so the equation for eigenvectors is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mathbf{v} = 0,$$

and the eigenspace is one-dimensional, with basis $(1, 0)$. So this matrix is not diagonalizable.

One important application of diagonalization is finding all powers of a matrix explicitly. If we know the representation (7) then evidently

$$A^k = B^{-1}\Lambda^k B, \quad \text{where} \quad \Lambda^k = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k).$$

Even if we do not know the diagonalization explicitly, we can sometimes make important conclusions. For example, if we know somehow that all eigenvalues of a diagonalizable matrix satisfy $|\lambda_j| < 1$, then we can deduce that $A^k \to 0$ as $k \to \infty$.

Applications.

Linear difference equations with constant coefficients.

The sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

is the well known Fibonacci sequence. It is defined by a recurrent relation

$$a_{n+2} = a_{n+1} + a_n, \quad \mathbf{a}_0 = 0, \quad a_1 = 1.$$

How to produce a closed form formula for a_n?

Consider the vectors $\mathbf{a}_n = (a_{n+1}, a_n)^T$. Then

$$\mathbf{a}_{n+1} = A\mathbf{a}_n, \quad \mathbf{a}_0 = (1, 0)^T,$$

where

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

And $\mathbf{a}_n = A^n \mathbf{a}_0$. So one only has to find A^n explicitly. The characteristic equation is

$$\lambda^2 - \lambda - 1 = 0,$$

so the eigenvalues are

$$\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

Notice that

$$\lambda_1 + \lambda_2 = 1, \quad \lambda_1 - \lambda_2 = \sqrt{5}, \quad \lambda_1\lambda_2 = -1.$$

Solving for eigenvectors, we obtain two linearly independent eigenvectors:

$$\mathbf{v}_1 = (\lambda_1, 1)^T, \quad \mathbf{v}_2 = (\lambda_2, 1)^T.$$

So the diagonalizing matrix is

$$B = \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix}, \quad \text{and} \quad B^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{pmatrix}.$$