1 Last time: vector spaces

A *(real) vector space* V is a set containing a *zero vector*, denoted 0, with *vector addition* and *scalar multiplication* operations that let us produce new vectors $u + v \in V$ and $cv \in V$ from given elements $u, v \in V$ and $c \in \mathbb{R}$. Several conditions must be satisfied so that these operations behave exactly like vector addition and scalar multiplication for $\mathbb{R}^n$. Most importantly, we require that

1. $u + v = v + u$.
2. $v - v = 0$ where we define $u - v = u + (-1)v$.
3. $v + 0 = v$
4. $cv = v$ if $c = 1$.

There are a few other more technical conditions to give the full definition (see the notes from last time).

$\mathbb{R}^n$ and any subspace of $\mathbb{R}^n$ are vector spaces.

The definitions of a *subspace* of a vector space and of *linear transformations* between vector spaces are identical to the ones we have already seen for subspaces of $\mathbb{R}^n$ and linear transformations $\mathbb{R}^n \to \mathbb{R}^m$.

Most vector spaces that do not arise as subspaces of $\mathbb{R}^n$ are subspaces of the following general construction. Let X be a set and let V be a vector space. Then the set $\text{Map}(X, V)$ of all functions $f : X \to V$ is a vector space once we define $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ and $0(x) = 0 \in V$ whenever $f, g : X \to V$ and $c \in \mathbb{R}$ and $x \in X$.

Example. The set of linear functions $\mathbb{R}^n \to \mathbb{R}^m$ is a subspace of $\text{Map}(\mathbb{R}^n, \mathbb{R}^m)$.

Such things as the *span*, *linear combination*, and *linear independence* of vectors in a general vector space also have essentially the same definitions as their counterparts for vectors in $\mathbb{R}^n$.

A *basis* of a vector space V is, again, a linearly independent set of vectors whose span is V. The *dimension* of a vector space is the number of elements in any of its bases (which all have the same size).

Example. Let n be a positive integer and let P_n be the set of polynomials in a variable x with coefficients in $\mathbb{R}$ of degree at most n. Recall that a polynomial is a function like 3 or x or $x^7 + 3x^2 + \sqrt{2}x - 1$.

The *degree* of a polynomial is the largest integer d such that x^d is a term with a nonzero coefficient. Constant polynomials are defined to have degree 0. Another way to the define the degree of a nonzero polynomial f is as the unique integer d such that $\lim_{x\to\infty} \frac{f(x)}{x^d}$ exists and is nonzero. For example, $x^7 + 3x^2 + \sqrt{2}x - 1$ has degree 7 since

$$\lim_{x\to\infty} \frac{x^7+3x^2+\sqrt{2}x-1}{x^d} = \begin{cases} 0 & \text{if } d > 7 \\ 1 & \text{if } d = 7 \\ \text{does not exist} & \text{if } d < 7. \end{cases}$$

The set P_n is a vector space: it is a subspace of $\text{Map}(\mathbb{R}^n, \mathbb{R}^n)$ A basis is given by the polynomials $1 = x^0, x, x^2, x^3, \dots, x^n$, and so $\dim P_n = n + 1$.

One natural way that we encounter vector spaces of functions is as the sets of solutions to (linear) differential equations, like $f'' + f = 0$. If you study these things in a math or physics course, abstract vector spaces will come up again.

2 Eigenvectors and eigenvalues

We return to the concrete setting of $\mathbb{R}^n$ and its subspaces.

Let A be a square $n \times n$ matrix.

Definition. An *eigenvector* of A is a nonzero vector $v \in \mathbb{R}^n$ such that $Av = \lambda v$ for a number $\lambda \in \mathbb{R}$. (λ is the Greek letter "lambda.") The number λ is called the *eigenvalue* of A for the eigenvector v.

The etymology is German: "eigen" means "own" in the sense of "belonging to" or "possessed by." So the eigenvectors of a matrix A are the vectors which the matrix can most convincingly claim it possesses.

Example. If we are given A and v, it is easy to check whether v is an eigenvector: just compute Av and inspect whether this vector is a scalar multiple of v.

For example, if $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ and $v = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$ then

$$Av = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4v$$

so v is an eigenvector of A with eigenvalue -4.

Caution: Remember that only *nonzero* vectors can be eigenvectors. This is because the fact that $A0 = \lambda 0$ for some number λ is not interesting, as this always holds since $A0 = \lambda 0 = 0$. However, the number 0 is allowed to be an eigenvalue of A.

Example. What are the eigenvectors of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}?$$

If $v \in \mathbb{R}^4$ were an eigenvector with eigenvalue λ then

$$Av = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_2 \\ v_3 \\ v_4 \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}.$$

The last equation implies that $0 = \lambda v_4$ and $\lambda_4 = \lambda v_3$ and $v_3 = \lambda v_2$ and $v_2 = \lambda v_1$. In other words,

$$0 = \lambda v_4 = \lambda^2 v_3 = \lambda^3 v_2 = \lambda^4 v_1.$$

If $\lambda \neq 0$ then this would mean that $v_1 = v_2 = v_3 = v_4 = 0$, but remember that v should be nonzero. Therefore the only possible eigenvalue of A is $\lambda = 0$. The eigenvectors of A with eigenvalue 0 are

$$v = \begin{bmatrix} v_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where v_1 is any nonzero real number.

To say that λ is an eigenvalue of A means that there exists a vector $x \in \mathbb{R}^n$ such that $Ax = \lambda x$.

Recall that I_n denotes the $n \times n$ identity matrix. Since n is usually a fixed number in this lecture, we abbreviate by setting $I = I_n$.

Proposition. A number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $A - \lambda I$ is not invertible.

Proof. The equation $Ax = \lambda x$ has a nonzero solution $x \in \mathbb{R}^n$ if and only if $(A - \lambda I)x = 0$ has a nonzero solution, which occurs if and only if $\mathrm{Nul}(A - \lambda I) \neq \{0\}$, which is equivalent to $A - \lambda I$ being not invertible. $\square$

Example. If $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ then

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \texttt{RREF}(A - 7I).$$

Since $\texttt{RREF}(A - 7I) \neq I$, the matrix $A - 7I$ is not invertible so 7 is an eigenvalue of A.

Corollary. A number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Proof. Remember that $A - \lambda I$ is not invertible if and only if $\det(A - \lambda I) = 0$. $\square$

The proof of the proposition suggests another way of defining an eigenvector: the eigenvectors of A with eigenvalue λ are precisely the nonzero elements of the nullspace $\mathrm{Nul}(A - \lambda I)$. Since we know how to construct a basis for the nullspace of any matrix, we also know how to find all eigenvectors of a matrix for any given eigenvalue.

Example. In the previous example, $\texttt{RREF}(A - 7I) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ so $Ax = 7x$ if and only if $(A - 7I)x = 0$ if and only if $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ where $x_1 - x_2 = 0$. In this linear system, x_2 is a free variable, and we can rewrite x as $x = \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. This means $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a basis for $\mathrm{Nul}(A - 7I)$, so the set of all nonzero multiples of this vector give all the eigenvectors of A with eigenvalue 7.

One calls the set of all $v \in \mathbb{R}^n$ with $Av = \lambda v$ the *eigenspace* of A for λ. We also call this the λ-*eigenspace* of A. Note that this is just the nullspace of $A - \lambda I$. A number is an eigenvalue of A if and only if the corresponding eigenspace is nonzero (that is, contains a nonzero vector).

Example. Suppose we heard that $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$ had 2 as an eigenvalue.

To find a basis for the 2-eigenspace of A, we row reduce

$$A - 2I = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1/2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \texttt{RREF}(A - 2I).$$

Thus $Ax = 2x$ if and only if $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ where $x_1 - \frac{1}{2}x_2 + 3x_3 = 0$, i.e., if and only if

$$x = \begin{bmatrix} \frac{1}{2}x_2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

The vectors $\begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$ are then a basis for the 2-eigenspance of A.

The *(main) diagonal* of an $n \times n$ matrix is the set of positions $(1,1),(2,2),\ldots,(n,n)$. The diagonal entries are the entries in these positions. Recall that a matrix is *triangular* if its nonzero entries all appear on or above the diagonal, or all appear on or below the diagonal.

Theorem. The eigenvalues of a triangular square matrix A are its diagonal entries.

Proof. If A has diagonal entries $d_1, d_2, \ldots, d_n$ then $A - \lambda I$ is triangular with diagonal entries $d_1 - \lambda$, $d_2 - \lambda$, ..., $d_n - \lambda$. This means that $\det(A - \lambda I) = (d_1 - \lambda)(d_2 - \lambda)\cdots(d_n - \lambda)$ which is zero if and only if $\lambda = d_i$ for some i. □

Example. The eigenvalues of the matrix $\begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ are 3, 0, and 2.

The eigenvalues of $\begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$ are 4 and 1.

Here is our second main result of today.

Theorem. If $\lambda_1, \lambda_2, \ldots, \lambda_r$ are distinct eigenvalues for A and $v_1, v_2, \ldots, v_r \in \mathbb{R}^n$ are the corresponding eigenvectors, so that $Av_i = \lambda_i v_i$ for $i = 1, 2, \ldots, r$, then the vectors $v_1, v_2, \ldots v_r$ are linearly independent.

Proof. Suppose that the vectors $v_1, v_2, \ldots, v_r$ instead are linearly dependent. We will argue that this leads to a logical contradiction, so is impossible.

Under this hypothesis, there must exist an index $p > 0$ such that $v_1, v_2, \ldots, v_p$ are linearly independent and v_{p+1} is a linearly combination of $v_1, v_2, \ldots, v_p$. (If no such index existed then it would mean that each of the sets $\{v_1\}$, $\{v_1, v_2\}$, $\{v_1, v_2, v_3\}$, ..., $\{v_1, v_2, \ldots, v_r\}$ were linearly independent. But we have assume the contrary.)

Let $c_1, c_2, \ldots, c_p \in \mathbb{R}$ be scalars such that

$$v_{p+1} = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p.$$

If we multiply both sides by A and use the fact that each vector is an eigenvector, it follows that

$$\lambda_{p+1} v_{p+1} = A v_{p+1} = A(c_1 v_1 + c_2 v_2 + \cdots + c_p v_p) = c_1 A v_1 + c_2 A v_2 + \cdots + c_p A v_p = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_p \lambda_p v_p.$$

On the other hand, multiplying both sides by λ_{p+1} gives

$$\lambda_{p+1} v_{p+1} = c_1 \lambda_{p+1} v_1 + c_2 \lambda_{p+1} v_2 + \cdots + c_p \lambda_{p+1} v_p.$$

Subtracting the two equations gives

$$0 = \lambda_{p+1} v_{p+1} - \lambda_{p+1} v_{p+1} = c_1(\lambda_1 - \lambda_{p+1}) v_1 + c_2(\lambda_2 - \lambda_{p+1}) v_2 + \cdots + c_p(\lambda_p - \lambda_{p+1}) v_p.$$

Since the vectors $v_1, v_2, \ldots, v_p$ are linearly independent, we must have

$$c_1(\lambda_1 - \lambda_{p+1}) = c_2(\lambda_2 - \lambda_{p+1}) = \cdots = c_p(\lambda_p - \lambda_{p+1}) = 0.$$

Remember that $\lambda_1, \lambda_2, \ldots, \lambda_r$ are all distinct so the differences $\lambda_i - \lambda_{p+1}$ for $i = 1, 2, \ldots, p$ are all nonzero. Therefore we must actually have $c_1 = c_2 = \cdots c_p = 0$. But this implies that $v_{p+1} = 0$, contradicting our assumption that v_{p+1} is an eigenvector and therefore nonzero.

We conclude from the contradiction that actually the vectors $v_1, v_2, \ldots, v_r$ are linearly independent. □

The *characteristic equation* of a square matrix A is the equation $\det(A - xI) = 0$ where x is a variable. The expression $\det(A - xI)$ is a polynomial in x, which we call the *characteristic polynomial* of A.

Corollary. A number λ is an eigenvalue of A if and only if $p(\lambda) = 0$ where $p(x) = \det(A - xI)$ is the characteristic polynomial.

Example. The matrix

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

has characteristic polynomial $\det(A - xI) = (5 - x)(3 - x)(5 - x)(1 - x) = (5 - x)^2(x - 3)(1 - x)$.

Since this polynomial has two linear factors given by a constant multiple of $5 - x$, i.e., since $(5 - x)^2$ divides $\det(A - xI)$, we say that 5 is an eigenvalue of A with *(algebraic) multiplicity* 2.

The other eigenvalues 1 and 3 have multiplicity 1.