1 Last time: complex eigenvalues

Write $\mathbb{C}$ for the set of complex numbers $\{a + bi : a, b \in \mathbb{R}\}$.

Each complex number is a formal linear combination of two real numbers $a + bi$.

The symbol i is defined as the square root of -1, so $i^2 = -1$.

We add complex numbers like this:

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

We multiply complex numbers just like polynomials, but substituting -1 for i^2:

$$(a + bi)(c + di) = ac + (ad + bc)i + bd(i^2) = (ac - bd) + (ad + bc)i.$$

The order of multiplication doesn't matter since $(a + bi)(c + di) = (c + di)(a + bi)$.

Draw $a + bi \in \mathbb{C}$ as the vector $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$.

Theorem (Fundamental theorem of algebra)**.** Suppose

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \dots a_1 x + a_0$$

is a polynomial of degree n (meaning $a_n \neq 0$) with coefficients $a_0, a_1, \dots, a_n \in \mathbb{C}$. Then there are n (not necessarily distinct) numbers $r_1, r_2, \dots, r_n \in \mathbb{C}$ such that

$$p(x) = (-1)^n a_n (r_1 - x)(r_2 - x) \cdots (r_n - x).$$

One calls the numbers $r_1, r_2, \dots, r_n$ the *roots* of $p(x)$.

A root r has *multiplicity* m if exactly m of the numbers $r_1, r_2, \dots, r_n$ are equal to r.

Define $\mathbb{C}^n$ as the set of vectors with n-rows and entries in $\mathbb{C}$, that is:

$$\mathbb{C}^n = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} : v_1, v_2, \dots, v_n \in \mathbb{C} \right\}.$$

We have $\mathbb{R}^n \subset \mathbb{C}^n$. The sum $u + v$ and scalar multiple cv for $u, v \in \mathbb{C}^n$ and $c \in \mathbb{C}$ are defined exactly as for vectors in $\mathbb{R}^n$, except we use the addition and multiplication operations for $\mathbb{C}$ instead of $\mathbb{R}$. If A is an $n \times n$ matrix and $v \in \mathbb{C}^n$ then we define Av in the same way as when $v \in \mathbb{R}^n$.

Definition. Let A be an $n \times n$ matrix. A number $\lambda \in \mathbb{C}$ a *(complex) eigenvalue* of A if there exists a nonzero vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$, or equivalently if $\det(A - \lambda I) = 0$.

If A is an $n \times n$ matrix then A has n possibly complex and not necessarily distinct eigenvalues, counting repeated eigenvalues with their respective multiplicities.

Given $a, b \in \mathbb{R}$, we define the *complex conjugate* of the complex number $a + bi \in \mathbb{C}$ to be

$$\overline{a + bi} = a - bi \in \mathbb{C}.$$

If A is a matrix and $v \in \mathbb{C}^n$ then we define $\overline{A}$ and $\overline{v}$ as the matrix and vector given by replacing all entries of A and v by their complex conjugates.

Lemma. Let $z \in \mathbb{C}$. Then $\overline{z} = z$ if and only if $z \in \mathbb{R}$.

Lemma. If $y, z \in \mathbb{C}$ then $\overline{y+z} = \overline{y} + \overline{z}$ and $\overline{yz} = \overline{y} \cdot \overline{z}$.

Hence if A is an $m \times n$ matrix (with real or complex entries) and $v \in \mathbb{C}^n$ then $\overline{Av} = \overline{A}\overline{v}$.

If $z = a + bi \in \mathbb{C}$ then $z\overline{z} = (a+bi)(a-bi) = a^2 + b^2 \in \mathbb{R}$.

This indicates how to divide complex numbers:

$$\frac{1}{a+bi} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$$

and more generally

$$\frac{c+di}{a+bi} = \tfrac{1}{a+bi} \cdot (c+di).$$

Proposition. Suppose A is an $n \times n$ matrix with real entries. If A has a complex eigenvalue $\lambda \in \mathbb{C}$ with eigenvector $v \in \mathbb{C}^n$ then $\overline{v} \in \mathbb{C}^n$ is an eigenvector for A with eigenvalue $\overline{\lambda}$.

The *real part* of a complex number $a + bi \in \mathbb{C}$ is $\Re(a+bi) = a \in \mathbb{R}$.

The *imaginary part* of a $a + bi \in \mathbb{C}$ is $\Im(a+bi) = b \in \mathbb{R}$.

Define $\Re(v)$ and $\Im(v)$ for $v \in \mathbb{C}^n$ by applying $\Re(\cdot)$ and $\Im(\cdot)$ to each entry in v.

Theorem. Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and associated eigenvector $v \in \mathbb{C}^2$. Then $A = PCP^{-1}$ where

$$P = \left[\begin{array}{cc} \Re(v) & \Im(v) \end{array}\right] \qquad \text{and} \qquad C = \left[\begin{array}{cc} a & -b \\ b & a \end{array}\right].$$

2 Some final properties of eigenvalues and eigenvectors

Before moving on to to inner products and orthogonality, we prove a few remaining properties of the (complex) eigenvalues and eigenvectors of a matrix which are worth remembering.

Lemma. Suppose we can write a polynomial in x in two ways as

$$(\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

for some complex numbers $\lambda_1, \lambda_2, \dots, \lambda_n, a_0, a_1, \dots, a_n \in \mathbb{C}$. Then

$$a_n = (-1)^n \qquad \text{and} \qquad a_{n-1} = (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_n) \qquad \text{and} \qquad a_0 = \lambda_1\lambda_2\cdots\lambda_n.$$

Proof. The product $(\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$ is a sum of 2^n monomials corresponding to a choice of either λ_i or $-x$ for each of the n factors, multiplied together.

The only such monomial of degree n is $(-x)^n = (-1)^n x^n = a_n x^n$ so $a_n = (-1)^n$.

The only such monomial of degree 0 is $\lambda_1\lambda_2\cdots\lambda_n = a_0$.

Finally, there are n monomial which arise of degree $n-1$:

$$\lambda_1(-x)^{n-1} + (-x)\lambda_2(-x)^{n-2} + (-x)^2\lambda_3(-x)^{n-3} + \cdots + (-x)^{n-1}\lambda_n = (-1)^{n-1}(\lambda_1 + \cdots + \lambda_n)x^{n-1}$$

This sum must be equal to $a_{n-1}x^{n-1}$ so $a_{n-1} = (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$. □

Let A be an $n \times n$ matrix.

Define $\mathrm{tr}(A)$ as the sum of the diagonal entries of A. Call $\mathrm{tr}(A)$ the *trace* of A.

Example. $\mathrm{tr}\left(\begin{bmatrix} 1 & 0 & 7 \\ -1 & 2 & 8 \\ 2 & 4 & 3 \end{bmatrix}\right) = 1 + 2 + 3 = 6.$

Proposition. If A, B are $n \times n$ matrices then $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ and $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

However, usually $\mathrm{tr}(AB) \neq \mathrm{tr}(A)\mathrm{tr}(B)$, unlike for the determinant.

Proof. The diagonal entries of $A + B$ are given by adding together the diagonal entries of A with those of B in corresponding positions, so it follows that $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.

Let A_{ij} and B_{ij} be the entries of A and B in positions (i, j). Then

$$(AB)_{jj} = \sum_{i=1}^{n} A_{ij} B_{ji} \qquad \text{and} \qquad (BA)_{jj} = \sum_{i=1}^{n} B_{ij} A_{ji} = \sum_{i=1}^{n} A_{ji} B_{ij}$$

so

$$\mathrm{tr}(AB) = \sum_{j=1}^{n}\sum_{i=1}^{n} A_{ij} B_{ji} \qquad \text{and} \qquad \mathrm{tr}(BA) = \sum_{j=1}^{n}\sum_{i=1}^{n} A_{ji} B_{ij}.$$

These sums are equal, since if we swap the roles of i and j in one expression we get the other. □

Theorem. Let A be an $n \times n$ matrix (with entries in $\mathbb{R}$ or $\mathbb{C}$).

Suppose the characteristic polynomial of A factors as

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x) \cdots (\lambda_n - x).$$

Then $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$ and $\mathrm{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. In other words:

(a) The product of the (complex) eigenvalues of A, counted with multiplicity, is $\det(A)$.

(b) The sum of the (complex) eigenvalues of A, counted with multiplicity if $\mathrm{tr}(A)$.

Remark. The noteworthy thing about this theorem is that it is true for all matrices.

For a diagonalisable matrix the result is much easier to prove.

If $A = PDP{-}1$ where D is a diagonal matrix, then

$$\det(A) = \det(PDP^{-1}) = \det(P)\det(D)\det(P)^{-1} = \det(D) = \lambda_1 \lambda_2 \cdots \lambda_n$$

since $\det(P)\det(P^{-1}) = \det(PP^{-1}) = \det(I) = 1$. Also

$$\mathrm{tr}(A) = \mathrm{tr}(PDP^{-1}) = \mathrm{tr}(DP^{-1}P) = \mathrm{tr}(D) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

Before proving the theorem let's see an example.

Example. If we have

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

then $\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix}$ are eigenvectors of A with eigenvalues i, i, and $-i$. One can check that

$$\det(A - xI) = -x^3 + ix^2 - x + i = (i - x)^2(-i - x),$$

so the theorem asserts that $(i)(i)(-i) = -i^3 = i = \det(A)$ and $i + i + (-i) = i = \mathrm{tr}(A)$.

Proof of the theorem. We can write $\det(A - xI) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ for some numbers $a_0, a_1, \ldots, a_n \in \mathbb{C}$. By the lemma it suffices to show that $a_0 = \det(A)$ and $a_{n-1} = (-1)^{n-1}\mathrm{tr}(A)$.

The first claim is easy. The value of a_0 is given by setting $x = 0$ in $\det(A - xI)$, so $a_0 = \det(A)$.

Showing that $a_{n-1} = (-1)^{n-1}\mathrm{tr}(A)$ takes a little more work. Consider the coefficient a_{n-1} of x^{n-1} in the characteristic polynomial $\det(A - xI)$. Remember our formula

$$\det(A - xI) = \sum_{Z \in S_n} (-1)^{\mathrm{inv}(Z)} \Pi(Z, A - xI) \qquad (*)$$

where $\Pi(Z, A - xI)$ is the product of the entries of $A - xI$ in the nonzero positions of the permutation matrix Z. The key observation to make is that if $Z \in S_n$ is not the identity matrix then Z has at most $n - 2$ nonzero entries on the diagonal, so $\Pi(Z, A - xI)$ is a polynomial in x degree at most $n - 2$.

Therefore the formula (*) implies that

$$\det(A - xI) = \Pi(I, A - xI) + \text{polynomial terms of degree} \leq n - 2.$$

Let d_i be the diagonal entry of A in position (i, i). Then $\Pi(I, A - xI) = (d_1 - x)(d_2 - x)\cdots(d_n - x)$ and the coefficient of x^{n-1} in this polynomial must be equal to the coefficient of x^{n-1} in $\det(A - xI)$.

By the lemma, the coefficient of x^{n-1} in $(d_1 - x)(d_2 - x)\cdots(d_n - x)$ is $(-1)^{n-1}(d_1 + d_2 + \cdots + d_n) = (-1)^{n-1}\mathrm{tr}(A)$, and so $a_{n-1} = (-1)^{n-1}\mathrm{tr}(A)$. □

Here's one way we might use the preceding theorem.

Corollary. Suppose A is a 2×2 matrix. Let $p = \det A$ and $q = \mathrm{tr} A$.

Then A has distinct eigenvalues if and only if $q^2 \neq 4p$.

Proof. Suppose $a, b \in \mathbb{C}$ are the eigenvalues of A (repeated with multiplicity).

Then $ab = p$ and $a + b = q$ so $a(q - a) = qa - a^2 = p$ and therefore $a^2 - qa + p = 0$.

The quadratic formula implies that

$$a = \frac{q \pm \sqrt{q^2 - 4p}}{2} \qquad \text{and} \qquad b = \frac{q \mp \sqrt{q^2 - 4p}}{2}$$

so we have $a \neq b$ if and only if $q^2 - 4p \neq 0$. □

Some other useful properties:

Proposition. If A is a square matrix then A and A^T have the same eigenvalues.

Proof. In fact, A and A^T has the same characteristic polynomial since

$$\det(A - xI) = \det((A - xI)^T) = \det(A^T - xI^T) = \det(A^T - xI).$$

□

Proposition. Let A be a square matrix. Then A is invertible if and only if 0 is not one of its eigenvalues. Assume A is invertible. Then A and A^{-1} have the same eigenvectors, but v is an eigenvector of A with eigenvalue λ if and only if v is an eigenvector of A^{-1} with eigenvalue $1/\lambda$.

Proof. 0 is an eigenvalue of A if and only if $\det A = 0$ which occurs precisely when A is not invertible.

If A is invertible and $Av = \lambda v$ then $v = A^{-1}Av = A^{-1}\lambda v = \lambda A^{-1}v$ so $A^{-1}v = \lambda^{-1}v$. □

Corollary. If A is invertible and diagonalisable then A^{-1} is diagonalisable.

Proof. If A is invertible and diagonalisable, then $\mathbb{R}^n$ has a basis consisting of eigenvectors of A, but this basis is then also made up of eigenvectors of A^{-1}, so A^{-1} is diagonalisable. □

Corollary. If A is diagonalisable then A^T is diagonalisable.

Proof. If $A = PDP^{-1}$ then $A^T = (PDP^{-1})^T = (P^{-1})^T D^T P^T = QEQ^{-1}$ for the invertible matrix $Q = (P^{-1})^T = (P^T)^{-1}$ and the diagonal matrix $E = D^T$. □

3 Inner products and orthogonality

We return (for the most part) to the setting of vectors in $\mathbb{R}^n$ and matrices with real entries.

Definition. The *inner product* (also called the *dot product*) of two vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

in $\mathbb{R}^n$ is the scalar

$$u \bullet v = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Note that $u \bullet v = u^T v = v^T u = v \bullet u$.

For example, $\begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} \bullet \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = 6 - 10 + 3 = -1.$

In the textbook, the inner product is printed more like "$u \cdot v$."

This means the same thing as what I'm writing here as "$u \bullet v$."

Some easy properties of the inner product.

Let $u, v, w \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a) $u \bullet v = v \bullet u$.

(b) $(u + v) \bullet w = u \bullet w + v \bullet w$.

(c) $(cu) \bullet v = c(u \bullet v)$.

(d) $v \bullet v = v_1^2 + v_2^2 + \cdots + v_n^2 \geq 0$.

(e) if $v \bullet v = 0$ ten $v_1 = v_2 = \cdots = v_n = 0 \in \mathbb{R}$ so $v = 0 \in \mathbb{R}^n$.

Definition. The *length* of a vector $v \in \mathbb{R}^n$ is the nonnegative real number

$$\|v\| = \sqrt{v \bullet v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Although $u \bullet v$ can be any real number, we always have $\|v\| \geq 0$.

For any $v \in \mathbb{R}^n$ it holds that $\|v\|^2 = v \bullet v$ and $\|cv\| = |c|\|v\|$ for $c \in \mathbb{R}$.

Lemma. If $v \in \mathbb{R}^n$ then $\|v\| = 0$ if and only if $v = 0$.

Proof. The only way we can have $\|v\| = 0$ is if $v_1 = v_2 = \cdots = v_n = 0$. □

A vector $v \in \mathbb{R}^n$ is a *unit vector* if $\|v\| = 1$.

Proposition. If $v \in \mathbb{R}^n$ is a nonzero vector then $u = \frac{1}{\|v\|} v \in \mathbb{R}^n$ is a unit vector.

Proof. We have $\left\| \frac{1}{\|v\|} v \right\| = \frac{1}{\|v\|} \|v\| = 1$. □

We refer to $u = \frac{1}{\|v\|} v$ as the unit vector *in the same direction as* the nonzero vector $v \in \mathbb{R}^n$.

Example. If

$$v = \begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix}$$

then $\|v\| = \sqrt{1 + 4 + 4 + 0} = \sqrt{9} = 3$ so

$$u = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

is the unit vector in the same direction as v.

The *distance* between two vectors $u, v \in \mathbb{R}^n$ is the length of the their difference $\|u - v\|$.

Definition. Two vectors $u, v \in \mathbb{R}^n$ are *orthogonal* if $u \bullet v = 0$.

To motivate this definition we consider what it means in 2 dimensions.

Suppose $u = \begin{bmatrix} a \\ b \end{bmatrix}$ and $v = \begin{bmatrix} x \\ y \end{bmatrix}$ are orthogonal vectors in $\mathbb{R}^2$, so that $ax + by = 0$. Assume both u and v are nonzero (since a zero vector is orthogonal to any vector and so is not very interesting to consider).

If $a = 0$ then we must have $b \neq 0 = by$, so $y = 0$ and $u = \begin{bmatrix} 0 \\ b \end{bmatrix}$ and $v = \begin{bmatrix} x \\ 0 \end{bmatrix} = -\frac{x}{b} \begin{bmatrix} -b \\ 0 \end{bmatrix}$.

If $a \neq 0$ then $x = \frac{-b}{a} y$ so $v = \begin{bmatrix} -\frac{b}{a} y \\ y \end{bmatrix} = \frac{y}{a} \begin{bmatrix} -b \\ a \end{bmatrix}$.

We conclude the following from these cases:

Proposition. If $u, v \in \mathbb{R}^2$ are orthogonal and $u = \begin{bmatrix} a \\ b \end{bmatrix}$ then v is a scalar multiple $\begin{bmatrix} -b \\ a \end{bmatrix}$, which is the vector obtained by rotating u counterclockwise by 90 degrees. Thus orthogonal vectors in $\mathbb{R}^2$ are *perpendicular/orthogonal* in the usual sense of lines in planar geometry.

Suppose $V \subset \mathbb{R}^n$ is a subspace. The *orthogonal complement* of V is the set

$$V^\perp = \{w \in \mathbb{R}^n : v \bullet w = 0 \text{ for all } v \in V\}.$$

Pronounce "$V^\perp$" as "vee perp."

Proposition. If $V \subset \mathbb{R}^n$ is a subspace then its orthogonal complement $V^\perp \subset \mathbb{R}^n$ is also a subspace.

Proof. Since $v \bullet 0 = 0$ for all $v \in \mathbb{R}^n$ it holds that $0 \in V^\perp$.

If $x, y \in V^\perp$ and $c \in \mathbb{R}$ then $v \bullet cx = c(v \bullet x) = 0$ and $v \bullet (x + y) = v \bullet x + v \bullet y = 0 + 0 = 0$ for all $v \in V$ so cx and $x + y$ both belong to $V^\perp$. Hence $V^\perp$ is a subspace. □

The operation $(\cdot)^\perp$ relates the column space, null space, and transpose of a matrix in the following way:

Theorem. Suppose A is an $m \times n$ matrix. Then $(\mathrm{Col}\, A)^\perp = \mathrm{Nul}(A^T)$.

Proof. Write $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$ where $a_i \in \mathbb{R}^m$. Let $v \in \mathbb{R}^n$.

Then $v \in (\mathrm{Col}\, A)^\perp$ if and only if $v \bullet a_i = a_i^T v = 0$ for all i. This holds if and and only if

$$A^T v = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} v = 0 \in \mathbb{R}^m,$$

i.e., if and only if $v \in \mathrm{Nul}(A^T)$. □

Here is our last result for today:

Lemma. Let $V \subset \mathbb{R}^n$ be a subspace . Suppose $v_1, v_2, \dots, v_k$ is a basis for V and $w_1, w_2, \dots, w_l$ is a basis for $V^\perp$. Then the concatenated list of vectors $v_1, v_2, \dots, v_k, w_1, w_2, \dots, w_l$ is linearly independent.

Later, we will show that actually $k + l = n$ so these linearly independent vectors are a basis for $\mathbb{R}^n$.

Proof. The only way that the vectors $v_1, v_2, \dots, v_k, w_1, w_2, \dots, w_l$ can be linearly dependent is if we could write

$$a_1 v_1 + \dots + a_k v_k = b_1 w_1 + \dots + b_l w_l$$

for some coefficients $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_l \in \mathbb{R}$ which are not all zero. But then there would exist a nonzero vector in both V and $V^\perp$, which is impossible since any $u \in V \cap V^\perp$ has $u \bullet u = 0$ so $u = 0$. □