1 Last time: properties of eigenvalues

The *trace* of a square matrix A is the sum of its diagonal entries.

We denote that by the symbol $\mathrm{tr}(A)$.

For 2×2 matrices we have $\mathrm{tr}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + d$.

Although it does not hold that $\mathrm{tr}(AB) = \mathrm{tr}(A)\mathrm{tr}(B)$, we have both

$$\mathrm{tr}(AB) = \mathrm{tr}(BA) \qquad \text{and} \qquad \det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA)$$

if A and B are $n \times n$ matrices.

Suppose A is an $n \times n$ matrix whose (complex) eigenvalues, repeated with multiplicities, are $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$. This is the same thing as saying that the characteristic polynomial of A factors as

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

Theorem. In this setup, it holds that $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$ and $\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

In words, one says that the product of the eigenvalues of A, repeated with multiplicity, is the determinant of A, while the sum of the eigenvalues of A, repeated with multiplicity, is the trace of A.

Here's one application.

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 $q^2 - 4p \neq 0$. □

We also noted a few other properties:

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

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 $1/\lambda$.

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

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

2 Inner products and orthogonality

Definition. The *inner product* or *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

$$u \bullet v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = u^T v = v^T u = v \bullet u.$$

For example, $\begin{bmatrix} a \\ b \end{bmatrix} \bullet \begin{bmatrix} -b \\ a \end{bmatrix} = -ab + ab = 0$ for any $a, b \in \mathbb{R}$.

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

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

Essential properties of length and inner product.

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

(a) $u \bullet v = v \bullet u$ and $(u + v) \bullet w = u \bullet w + v \bullet w$ and $(cv) \bullet w = c(v \bullet w)$, while $\|cv\| = |c|\|v\|$.

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

(c) $v \bullet v = 0$ if and only if $\|v\| = 0$ if and only if $v = 0 \in \mathbb{R}^n$.

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

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

If $v \in \mathbb{R}^n$ is any nonzero vector, then the *unit vector in the direction of* v is $u = \frac{1}{\|v\|} v \in \mathbb{R}^n$.

Example. The unit vector is the direction of

$$v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{is} \quad u = \frac{1}{\sqrt{1^2+1^2+1^2+1^2}} v = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

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

When u and v are orthogonal we also say that "u is orthogonal to v."'

"Orthogonal" is a synonym and generalisation of "perpendicular" to higher dimensions.

Proposition. If $u, v \in \mathbb{R}^2$ are nonzero vectors that are orthogonal to each other, so that $u \bullet v = 0$. Then u and v, drawn as arrows in the xy-plane, belong to perpendicular lines through the origin. In other words, these vectors are perpendicular/orthogonal in the usual sense of planar geometry.

Concretely, $u, v \in \mathbb{R}^2$ 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.

Proof. Write $u = \begin{bmatrix} a \\ b \end{bmatrix}$ and $v = \begin{bmatrix} x \\ y \end{bmatrix}$. Then $u \bullet v = ax + by = 0$.

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

Thus v is a scalar multiple of $\begin{bmatrix} -b \\ a \end{bmatrix}$.

To see that $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} -b \\ a \end{bmatrix}$ are perpendicular, draw a picture. Consider the triangles with vertices $(0,0), (a,0), (a,b)$ and $(0,0), (-b,0), (-b,a)$. These triangles are congruent, and the angle between $\begin{bmatrix} -b \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -b \\ a \end{bmatrix}$ plus the angle between $\begin{bmatrix} a \\ 0 \end{bmatrix}$ and $\begin{bmatrix} a \\ b \end{bmatrix}$ must be 90 degrees. □

3 Orthogonal complements

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 $(\operatorname{Col} A)^\perp = \operatorname{Nul}(A^T)$.

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

If $v \in (\operatorname{Col} A)^\perp$ then we must have $v \bullet a_i = a_i^T v = 0$ for all i.

Conversely, if $v \bullet a_i = a_i^T v = 0$ for all i then

$$(c_1 a_1 + c_2 a_2 + \dots + c_n a_n) \bullet v = c_1 \underbrace{(a_1 \bullet v)}_{=0} + c_2 \underbrace{(a_2 \bullet v)}_{=0} + \dots + c_n \underbrace{(a_n \bullet v)}_{=0} = 0$$

for any scalars $c_1, c_2, \dots, c_n \in \mathbb{R}$ so $v \in (\operatorname{Col} A)^\perp$.

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

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

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

Lemma. Let $V \subset \mathbb{R}^n$ be a subspace.

If $w \in V \cap V^\perp$ then $w = 0$.

Proof. If $w \in V$ and $w \in V^\perp$ then $w \bullet w = 0$ so $w = 0$. □

Proposition. Let $V \subset \mathbb{R}^n$ be a subspace. If $S \subset V$ and $T \subset V^\perp$ are two sets of linearly independent vectors, then $S \cup T$ is also linearly independent.

Proof. Suppose there was a nontrivial linear dependence among the elements of $S \cup T$ equal to zero. Rewrite this linear dependence so that the terms from S are on the left side of $=$ and the terms from T are on the other side. Then we would have an equation of the form

$$\underbrace{a_1 v_1 + \cdots + a_k v_k}_{\in V} = \underbrace{b_1 w_1 + \cdots + b_l w_l}_{\in V^\perp}$$

where $v_1, \dots, v_k \in S$ and $w_1, \dots, w_l \in T$, 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 such an equation would imply that a nonzero element of V is equal to a nonzero element of $V^\perp$, which is impossible by the lemma. □

Corollary. If $V \subset \mathbb{R}^n$ is a subspace then $\dim V^\perp \leq n - \dim V$.

Proof. If S is a basis for V and T is a basis for $V^\perp$ then $\dim V + \dim V^\perp = |S| + |T| = |S \cup T|$. Since $S \cup T$ is a set of linearly independent vectors in $\mathbb{R}^n$, its size must be at most n. □

4 Orthogonal bases and orthogonal projections

Proposition (Generalised Pythagorean theorem)**.** Two vectors $u, v \in \mathbb{R}^n$ are orthogonal if and only if

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. The proof is just a little algebra:

$$\|u+v\|^2 = (u+v) \bullet (u+v) = u \bullet (u+v) + v \bullet (u+v) = u \bullet u + u \bullet v + v \bullet u + v \bullet v = \|u\|^2 + \|v\|^2 + 2(u \bullet v).$$

Then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ if and only if $u \bullet v = 0$.

The equivalence of this proposition to the classical Pythagorean theorem boils down to our observation earlier that orthogonal vectors in $\mathbb{R}^2$ form the sides of a right triangle. □

A collection of vectors $u_1, u_2, \dots, u_p \in \mathbb{R}^n$ is *orthogonal* if $u_i \bullet u_j = 0$ whenever $1 \leq i < j \leq p$.

In particular, an *orthogonal basis* of $\mathbb{R}^n$ is a basis in which any two vectors are orthogonal.

Theorem. Suppose the vectors $u_1, u_2, \dots, u_p \in \mathbb{R}^n$ are orthogonal and all nonzero. Then $u_1, u_2, \dots, u_p$ are linearly independent.

Proof. Suppose $c_1u_1 + c_2u_2 + \cdots + c_pu_p = 0$ for some coefficients $c_1, c_2, \ldots, c_p \in \mathbb{R}$.

For each $i = 1, 2, \ldots, p$, we then have

$$0 = (c_1u_1 + c_2u_2 + \cdots + c_pu_p) \bullet u_i = c_1(u_1 \bullet u_i) + c_2(u_2 \bullet_i) + \cdots + c_p(u_p \bullet u_i) = c_i\|u_i\|^2$$

since $u_j \bullet u_i = 0$ if $i \neq j$. But since u_i is nonzero, $\|u_i\|^2 \neq 0$, so it must hold that $c_i = 0$. As this argument applies to each index i, we deduce that $c_1 = c_2 = \cdots = c_p = 0$.

In other words, the only way we can have $c_1u_1 + c_2u_2 + \cdots + c_pu_p = 0$ is if all of the coefficients are zero, which is the definition of linear independence. □

Corollary. Any set of nonzero, orthogonal vectors is an orthogonal basis for the subspace they span.

Any set of n nonzero, orthogonal vectors in $\mathbb{R}^n$ is an orthogonal basis for $\mathbb{R}^n$.

Proposition. Suppose $u_1, u_2, \ldots, u_p$ is an orthogonal basis for a subspace $V \subset \mathbb{R}^n$.

Let $y \in V$. Then we can write $y = c_1u_2 + c_2u_2 + \cdots + c_pu_p$ where

$$c_i = \frac{y \bullet u_i}{u_i \bullet u_i} = \frac{y \bullet u_i}{\|u_i\|^2}.$$

Proof. A basis must span V, so $y = c_1u_2 + c_2u_2 + \cdots + c_pu_p$ for some coefficients $c_1, c_2, \ldots, c_p \in \mathbb{R}$.

Since $y \bullet u_i = c_i(u_i \bullet u_i)$ for each $i = 1, 2, \ldots, p$, the result follows. □

Example. Suppose $u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ and $u_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$.

You can check that these three vectors form as orthogonal subset of $\mathbb{R}^3$.

For example, $u_1 \bullet u_3 = -3/2 - 2 + 7/2 = 0$.

The vectors are therefore linearly independent, so are an orthogonal basis for $\mathbb{R}^3$.

For $y = \begin{bmatrix} 6 \\ 1 \\ 8 \end{bmatrix}$ we have

$$y \bullet u_1 = 11 \quad \text{and} \quad y \bullet u_2 = -12 \quad \text{and} \quad y \bullet u_3 = -33$$

while

$$u_1 \bullet u_1 = 11 \quad \text{and} \quad u_2 \bullet u_2 = 6 \quad \text{and} \quad u_3 \bullet u_3 = 33/2$$

so

$$y = u_1 - 2u_2 - 2u_3.$$

Let $u \in \mathbb{R}^n$ be a nonzero vector. Suppose $y \in \mathbb{R}^n$ is any vector.

Definition. The *orthogonal projection* of y onto u is the vector

$$\boxed{\widehat{y} = \frac{y \bullet u}{u \bullet u}u.}$$

Note that this vector is scalar multiple of y, and can be zero.

The *component of* y *orthogonal to* u is the vector

$$\boxed{z = y - \widehat{y} = y - \frac{y \bullet u}{u \bullet u} u.}$$

By construction it holds that $y = \widehat{y} + z$. Moreover, as its name suggests, we have $z \bullet u = 0$ since

$$z \bullet u = y \bullet u - \frac{y \bullet u}{u \bullet u} u \bullet u = y \bullet u - y \bullet u = 0.$$

Observation. The vectors $\widehat{y}$ and z do not change if u is replaced by a nonzero scalar multiple: if we change u to cu for some $0 \neq c \in \mathbb{R}$ then all the factors of c cancel:

$$\frac{y \bullet cu}{cu \bullet cu} cu = \frac{c(y \bullet u)}{c^2(u \bullet u)} cu = \frac{y \bullet u}{u \bullet u} u = \widehat{y}.$$

Let $L = \mathbb{R}\text{-span}\{u\}$. Then $\widehat{y}$ and z may also be called the *orthogonal projection* of y onto L the *component* of y orthogonal to L. We will write $\boxed{\text{proj}_L(y) = \widehat{y} \in L.}$

In $\mathbb{R}^2$, the distance from a point (x, y) to a line $L = \mathbb{R}\text{-span}\{u\}$ is the length the

$$\left\| \begin{bmatrix} x \\ y \end{bmatrix} - \text{proj}_L\left(\begin{bmatrix} x \\ y \end{bmatrix} \right) \right\|.$$

(Try drawing a picture to understand this fact!)

Example. To find the distance from the point $(x, y) = (7, 6)$ to the line L defined by $y = \frac{1}{2}x$, note that L contains the vector $u = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Let $y = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$. Then

$$\text{proj}_L\left(\begin{bmatrix} 7 \\ 6 \end{bmatrix} \right) = \frac{y \bullet u}{u \bullet u} u = \frac{28 + 12}{16 + 4} u = \frac{40}{20} u = 2u = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

so the distance is

$$\left\| \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\| = \sqrt{1 + 4} = \sqrt{5}.$$