Recall from Week 10

- The characteristic polynomial of a square matrix A is $\det(\lambda I - A)$.
- Its roots are the values of λ for which the equation $Av = \lambda v$ is satisfied by a non-zero vector v.
- We were in the process of describing how a 3×3 determinant can be calculated from the equation

$$A \times \text{adj}(A) = \det(A) I.$$

The 3×3 case

The version of the above equation for a 3×3 matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \underbrace{\begin{bmatrix} ei - fh & -bi + ch & bf - ce \\ -di + fg & ai - cg & -af + cd \\ dh - eg & -ah + bg & ae - bd \end{bmatrix}}_{\text{adj}(A)} = (\underbrace{aei - afh - bdi + bfg + cdh - ceg}_{\det(A)})I_3.$$

Definitions

- The minor $M_{i,j}$ of the (i,j)-entry of A is the determinant of the 2×2 matrix that remains when Row i and Column j are deleted from A.
- The cofactor $C_{i,j}$ is either $M_{i,j}$ or $-M_{i,j}$, according to $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$
- The adjugate of A is has $C_{j,i}$ in the (i,j)-position. It is the transpose of the matrix of cofactors of A.
- The determinant A can be found by multiplying each entry of any chosen row or column by its own cofactor, and adding the results.

The 3×3 case

The version of the above equation for a 3×3 matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} \begin{vmatrix} e & f \\ h & i \end{vmatrix} & -\begin{vmatrix} b & c \\ h & i \end{vmatrix} & \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ -\begin{vmatrix} d & f \\ g & i \end{vmatrix} & \begin{vmatrix} a & c \\ g & i \end{vmatrix} & -\begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ \begin{vmatrix} d & e \\ g & h \end{vmatrix} & -\begin{vmatrix} a & b \\ g & h \end{vmatrix} & \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix} = \underbrace{(aei - afh - bdi + bfg + cdh - ceg)}_{\det(A)} I_3.$$

Definitions

- The minor $M_{i,j}$ of the (i,j)-entry of A is the determinant of the 2×2 matrix that remains when Row i and Column j are deleted from A.
- The cofactor $C_{i,j}$ is either $M_{i,j}$ or $-M_{i,j}$, according to $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$
- The adjugate of A is has $C_{j,i}$ in the (i,j)-position. It is the transpose of the matrix of cofactors of A.
- The determinant A can be found by multiplying each entry of any chosen row or column by its own cofactor, and adding the results.

Determinant Properties

1. Each of the definitions above applies to $n \times n$ matrices in general, and gives us a way to recursively define a $n \times n$ determinant, in terms of $(n-1) \times (n-1)$ determinants.
2. The cofactor expansion method is not generally the most efficient way to compute a determinant (it is ok in the 3×3 case). But it can be taken as the definition of a determinant.
3. In some special cases, the determinant is easier to compute. If A is upper or lower triangular, then $\det(A)$ is the product of the entries on the main diagonal of A. If A has a square $k \times k$ block A_1 in the upper left, a square $(n-k) \times (n-k)$ block in the lower right, and only zeros in the lower left $(n-k) \times k$ region, then $\det(A) = \det(A_1)\det(A_2)$.
4. For a pair of $n \times n$ matrices A and B, $\det(AB) = \det(A)\det(B)$. This is the multplicative property of the determinant, or the Cauchy-Binet formula. It is not obvious at all.

The characteristic polynomial of a 3×3 matrix

Example Using cofactor expansion by the first column, we find that the characteristic polynomial of $B = \begin{bmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{bmatrix}$ is

$$\begin{aligned} \det(\lambda I_3 - B) &= \det \begin{bmatrix} \lambda - 5 & -6 & -2 \\ 0 & \lambda + 1 & 8 \\ -1 & 0 & \lambda + 2 \end{bmatrix} \\ &= (\lambda - 5)\left((\lambda+1)(\lambda+2) - 0(8)\right) + (-1)\left((-6)(8) - (\lambda + \right. \\ &= (\lambda - 5)(\lambda^2 + 3\lambda + 2) - (2\lambda - 46) \\ &= \lambda^3 - 2\lambda^2 - 15\lambda + 36 \\ &= (\lambda - 3)(\lambda^2 + \lambda - 12) \\ &= (\lambda - 3)(\lambda + 4)(\lambda - 3) \\ &= (\lambda - 3)^2(\lambda + 4) \end{aligned}$$

Algebraic and Geometric Multiplicity

The eigenvalues of B are 3 (occurring twice as a root of the characteristic polynomial), and -4 (occurring once). We say that 3 has algebraic multplicity 2 and -4 has algebraic multiplicity 1 as an eigenvalue of B.

The *geometric multplicity* of each eigenvalue is the dimension of its corresponding eigenspace.

$$3I_3 - B = \begin{bmatrix} -2 & -6 & -2 \\ 0 & 4 & 8 \\ -1 & 0 & 5 \end{bmatrix}$$

The RREF of this matrix is $\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ and the nullspace consists of all vectors $\begin{bmatrix} 5t \\ -2t \\ t \end{bmatrix}$, where $t \in \mathbb{R}$. This is the eigenspace of B corresponding to $\lambda = 3$. It has dimension 1, so 3 has geometric multiplicity 1 as an eigenvector of B.

Algebraic and Geometric Multiplicity

The eigenvalues of B are 3 (occurring twice as a root of the characteristic polynomial), and -4 (occurring once). We say that 3 has algebraic multplicity 2 and -4 has algebraic multiplicity 1 as an eigenvalue of B.
The *geometric multplicity* of each eigenvalue is the dimension of its corresponding eigenspace.

$$3I_3 - B = \begin{bmatrix} -2 & -6 & -2 \\ 0 & 4 & 8 \\ -1 & 0 & 5 \end{bmatrix}$$

The RREF of this matrix is $\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ and the nullspace consists of all vectors $\begin{bmatrix} 5t \\ -2t \\ t \end{bmatrix}$, where $t \in \mathbb{R}$. This is the eigenspace of B corresponding to $\lambda = 3$. It has dimension 1, so 3 has geometric multiplicity 1 as an eigenvector of B.

Algebraic and Geometric Multiplicity

Theorem The geometric multplicity of an eigenvalue is at most equal to its algebraic multiplicity.

Proof: Suppose that μ has geometric multiplicity k as an eigenvalue of the square matrix $A \in M_n(\mathbb{R})$, and let $\{v_1, \dots, v_k\}$ be a basis for the μ-eigenspace of A. Extend this to a basis $\mathcal{B}$ of $\mathbb{R}^n$, and let P be the matrix whose columns are the elements of $\mathcal{B}$. Then the first k columns of $P^{-1}AP$ have μ in the diagonal position and zeros elsewhere. It follows that $\lambda - \mu$ occurs at least k times as a factor of $\det(\lambda I_n - P^{-1}AP)$.

Corollary A matrix is diagonalizable if and only if the geometric multiplicity of each of its eigenvalues is equal to the algebraic multiplicity.

In $\mathbb{R}^2$, the scalar (or dot) product of the vectors $x = \binom{x_1}{x_2}$ and $y = \binom{y_1}{y_2}$ is given by

$$x \cdot y = x_1 y_1 + x_2 y_2 = x^T y = y^T x = y \cdot x.$$

We can interpret the *length* $||x||$ of the vector x as the length of the directed line segment from the origin to (x_1, x_2), which by the Theorem of Pythagoras is $\sqrt{x_1^2 + x_2^2}$ or $\sqrt{x \cdot x}$.

Once we have a concept of length of a vector, we can define the *distance* $d(x, y)$ between two vectors x and y as the length of their difference: $d(x, y) = ||x - y||$.

In $\mathbb{R}^2$, the scalar (or dot) product of the vectors $x = \binom{x_1}{x_2}$ and $y = \binom{y_1}{y_2}$ is given by

$$x \cdot y = x_1 y_1 + x_2 y_2 = x^T y = y^T x = y \cdot x.$$

Similarly, from the Cosine Rule we can observe that $x \cdot y = ||x||\,||y|| \cos\theta$, where θ is the angle between the directed line segments representing x and y. In particular, x is orthogonal to y (or $x \perp y$) if and only if $x \cdot y = 0$.

So the scalar product encodes geometric information in $\mathbb{R}^2$, and it also provides a mechanism for defining concepts of length, distance and orthogonality on real vector spaces that do not necessarily have an obvious geometric structure.

Real Inner Products

Let V be a vector space over $\mathbb{R}$. An inner product on V is a function from $V \times V$ to $\mathbb{R}$ that assigns an element of $\mathbb{R}$ to every ordered pair of elements of V, and has the following properties. We write $\langle x, y \rangle$ for the inner product of x and y, and write the function as $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$.

1. Symmetry: $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$
2. Linearity in both slots (bilinearity): For all $x, y, z \in V$ and all $a, b \in \mathbb{R}$, we have $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ and $\langle x, ay + bz \rangle = a\langle x, y \rangle + b\langle x, z \rangle$.
3. Non-negativity: $\langle x, x \rangle \geq 0$ for all $x \in V$, and $\langle x, x \rangle = 0$ only if $x = 0_V$.

The ordinary scalar product on $\mathbb{R}^n$ is the best known example of an inner product.

Examples of inner products

1. The ordinary scalar product on $\mathbb{R}^n$.
2. Let C be the vector space of all continuous real-valued functions on the interval $[0,1]$. The analogue of the ordinary scalar product on C is the inner product given by
$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx, \text{ for } f, g \in C.$$
3. On the space $M_{m\times n}(\mathbb{R})$, the *Frobenius inner product* or *trace inner product* is defined by $\langle A, B\rangle = \text{trace}(A^T B)$. Note that trace$A^T B$ is the sum over all positions (i,j) of the products $A_{ij}B_{ij}$. So this is closely related to the ordinary scalar product, if the matrices A and B were regarded as vectors with mn entries over $\mathbb{R}$.

It is possible for a single vector space to have many different inner products defined on it, and if there is any risk of ambiguity we need to specify which one we are considering.

Length, Distance and Orthogonality

Given a real vector space and equipped with an inner product $\langle \cdot, \cdot \rangle$, we make the following two definitions.

Definition We define the length or norm of any vector v by

$$||v|| = \sqrt{\langle v, v \rangle},$$

and we define the distance between the vectors u and v by

$$d(u, v) = ||u - v||.$$

Definition We say that vectors u and v are orthogonal (with respect to $\langle \cdot, \cdot \rangle$) if $\langle u, v \rangle = 0$.

These definitions are consistent with "typical" geometrically motivated concepts of distance and orthogonality.

Unit Vectors and Scaling

An element v of V is referred to as a unit vector if $||v|| = 1$.
The norm of elements of V has the property that $||cv|| = |c|\,||v||$ for any vector v and real scalar c. To see this we can note that

$$||cv|| = \sqrt{\langle cv, cv\rangle} = \sqrt{c^2\langle v, v\rangle} = c||v||.$$

So we can adjust the norm of any element of V, while preserving its direction, by multplying it by a positive scalar.

Definition If v is a non-zero vector in an inner product space V, then

$$\hat{v} := \frac{1}{||v||}v$$

is a unit vector in the same direction as v, referred to as the *normalization* of v.

Orthogonal Projection

Lemma Let u and v be non-zero vectors in an inner product space V. Then it is possible to write (in a unique way) $v = au + v'$, where a is scalar and v' is orthogonal to u.

- If v is orthogonal to u, take $a = 0$ and $v' = v$.
- If v is a scalar multiple of u, take $au = v$ and $v' = 0$.
- Otherwise, to solve for a and v' in the equation $v = au + v'$ (with $u \perp v'$), take the inner product with u on both sides. Then

$$\langle u, v\rangle = a\langle u, u\rangle + 0 \Longrightarrow a = \frac{\langle u, v\rangle}{||u||^2},\ \ v' = v - \frac{\langle u, v\rangle}{||u||^2}u.$$

We can verify directly that the two components in this expression are orthogonal to each other.

Example In $\mathbb{R}^2$, write $u = \binom{2}{1}$ and $v = \binom{6}{-2}$.

Orthogonal projection of one vector on another

Definition

For non-zero vectors u and v in an inner product space V, the vector $\frac{\langle u, v\rangle}{||u||^2}u$ is called the projection of v on the 1-dimensional space spanned by u. It is denoted by $\mathrm{proj}_u(v)$ and it has the property that $v - \mathrm{proj}_u(v)$ is orthogonal to u.

Lemma

$\mathrm{proj}_u(v)$ is the unique element of $\langle u\rangle$ whose distance from v is minimal.

Proof Let au be a scalar multiple of u. Then

$$d(au, v)^2 = \langle au - v, au - v\rangle = a^2\langle u, u\rangle - 2a\langle u, v\rangle + \langle v, v\rangle$$

Regarded as a quadratic function of a, this has a minimum when its derivative is 0, i.e. when $2a\langle u, u\rangle - 2\langle u, v\rangle = 0$, when $a = \frac{\langle u, v\rangle}{||u||^2}$.

Orthogonal Bases (the Gram-Schmidt process)

Let V be a finite-dimensional inner product space, with a given basis $\mathcal{B} = \{b_1, b_2, \dots b_n\}$.

A basis $\mathcal{B}$ is called orthogonal if its elements are all orthogonal to each other.

We can adjust $\mathcal{B}$ to an orthogonal basis $\mathcal{B}' = \{v_1, \dots, v_n\}$ as follows.

1. Write $v_1 = b_1$.
2. Write $v_2 = b_2 - \text{proj}_{v_1}(v_2) = b_2 - \dfrac{\langle b_1, b_2\rangle}{||b_1||^2} b_1$.

   Then the pairs b_1, b_2 and v_1, v_2 span the same space, and $v_1 \perp v_2$.
3. Write $v_3 = b_3 - \text{proj}_{v_1}(b_3) - \text{proj}_{v_2}(b_3)$.

   Then the sets v_1, v_2, v_3 and b_1, b_2, b_3 span the same space, and v_3 is orthogonal to both v_1 and v_2.

   To see this note that

$$\langle v_1, v_3\rangle = \langle v_1, b_3\rangle - \frac{\langle v_1, b_3\rangle}{\langle v_1, v_1\rangle}\langle v_1, v_1\rangle - c\cancel{\langle v_1, v_2\rangle}$$

4. Continue in this way - at the kth step, form v_k by subtracting from b_k its projections on each of $v_1, \dots, v_n$.

Orthogonal projection on a subspace

The result of this process is a basis $\{v_1, \dots, v_n\}$ whose elements satisfy

$$\langle v_i, v_j \rangle = 0 \text{ for } i \neq j$$

We can adjust this basis to a orthonormal basis (consisting of orthogonal unit vectors) by replacing each v_i with its normalization $\hat{v}_i$.
From the Gram-Schmidt process, we have

Theorem

If V is a finite-dimensional inner product space, then V has an orthogonal (or orthonormal) basis.

Now let W be a subspace of V, and let $v \in V$. The orthogonal projection of v on W, denoted $\text{proj}_W(v)$, is defined to be the unique element u of W for which

$$v = u + v',$$

and $v' \perp w$ for all $w \in W$.

How to calculate a projection from an orthogonal basis

That $\text{proj}_W(v)$ exists follows from the fact that an orthogonal basis $\{b_1, \dots, b_k\}$ of W may be extended to an orthogonal basis $\mathcal{B} = \{b_1, \dots, b_k, c_{k+1}, \dots, c_n\}$ of W. Then v has a unique expression of the form

$$v = a_1 b_1 + \cdots + a_k b_k + a_{k+1} c_{k+1} + \cdots + a_n c_n, \text{ for scalars } a_i,$$

and $\text{proj}_W(v) = a_1 b_1 + \cdots + a_k b_k$.
Moreover, taking inner products with b_i gives $\langle v, b_i \rangle = a_i \langle b_i, b_i \rangle$, so that

$$\text{proj}_W(v) = \sum_{i=1}^{k} \frac{\langle v, b_i \rangle}{\langle b_i, b_i \rangle} b_i,$$

where $\{b_1, \dots, b_k\}$ is an orthogonal basis of W.

$\text{proj}_W(v)$ is the nearest point of W to v

Let $u = \text{proj}_W(v)$ and let w be any element of W. Then

$$\begin{aligned} d(v,w)^2 &= \langle v-w, v-w\rangle \\ &= \langle (v-u)+(u-w), (v-u)+(u-w)\rangle \\ &= \langle v-u, v-u\rangle + \cancel{2\langle v-u, u-w\rangle} + \langle u-w, u-w\rangle \\ &= \langle v-u, v-u\rangle + \langle u-w, u-w\rangle \\ &\geq d(v,u)^2, \end{aligned}$$

with equality only if $w = \text{proj}_W(v)$.

Example In $\mathbb{R}^3$, find the unique point of the plane $x + 2y - z = 0$ that is nearest to the point $(1, 2, 2)$.