4.1 Inner products

Let V be a vector space with field of scalars $\mathbb{R}$.

An *inner product on V* is a function

$$\langle\, ,\, \rangle\colon V \times V \to \mathbb{R}$$

satisfying

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \lambda\mathbf{u}, \mathbf{v} \rangle = \lambda\langle \mathbf{u}, \mathbf{v} \rangle$
3. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
4. (a) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$

   (b) $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}$.

An *inner product space* is a vector space V together with a choice of inner product.

If W is a subspace of V, then W is itself an inner product space with respect to the inner product of V.

Example 4.1. Consider

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + 2u_2v_2$$

on $\mathbb{R}^2$.

Example 4.2. Consider

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_1v_2 + u_2v_2$$

on $\mathbb{R}^2$.

Example 4.3. Consider

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 - u_2v_2 + u_3v_3$$

on $\mathbb{R}^3$.

Example 4.4. Consider

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$$

on the vector space of continuous functions $\mathcal{C}([0,1])$.

4.2 The Cauchy–Schwarz inequality

Let V be an inner product space.

The *length* of a vector $\mathbf{u} \in V$ is

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

The *distance* between two vectors $\mathbf{u}, \mathbf{v} \in V$ is

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

To define the angle between two vectors, we need the Cauchy–Schwarz inequality:

Theorem 4.5. Let $\mathbf{u}$, $\mathbf{v}$ be vectors in an inner product space V. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

Proof.

The *angle θ* between two vectors $\mathbf{u}, \mathbf{v} \in V$ is defined by the equation

$$\cos(\theta) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

Example 4.6. Consider the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 2 u_2 v_2$$

on $\mathbb{R}^2$.

Compute $\|\mathbf{u}\|$, $d(\mathbf{u}, \mathbf{v})$, and the angle between $\mathbf{u}$ and $\mathbf{v}$ for $\mathbf{u} = (2, 5)$ and $\mathbf{v} = (-1, 3)$.

4.3 Orthogonality

Let V be an inner product space.

We say that $\mathbf{u}, \mathbf{v} \in V$ are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

We say that a subset $S \subset V$ is *orthogonal* if any two distinct vectors in S are orthogonal.

We say that a subset $S \subset V$ is *orthonormal* if it is orthogonal and every vector in S has length 1.

Example 4.7. In $\mathbb{R}^2$ with the dot product,

the vectors $(1, 1)$ and $(1, -1)$ are

the vectors $(1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$ are

Example 4.8. In $\mathcal{C}([-1,1])$ with inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx,$$

the functions x^2 and x^3 are

Theorem 4.9. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthogonal set of nonzero vectors in an inner product space. Then S is linearly independent.

(We call S an *orthogonal basis* for $\mathrm{Span}(S)$. Moreover, if S is an orthonormal set, we call it an *orthonormal basis* for $\mathrm{Span}(S)$.)

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for V, then every $\mathbf{v} \in V$ can be written

$$\mathbf{v} =$$

We will soon see that every inner product space has orthonormal bases.

Proposition 4.10. Let $\mathbf{u}, \mathbf{v}$ be vectors in an inner product space V, with $\mathbf{u} \neq \mathbf{0}$. Then the vector

$$\mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$$

is orthogonal to $\mathbf{u}$.

This motivates the definition: the *orthogonal projection of* $\mathbf{v}$ *onto* $\mathbf{u}$ is

$$\operatorname{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

If $\mathbf{u}$ happens to be a unit vector, then the formula simplifies to

$$\operatorname{proj}_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u}.$$

Moreover, we can project onto a subspace W of V as follows: let $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be an orthonormal basis for W. The *orthogonal projection of* $\mathbf{v}$ *onto* W is

$$\begin{aligned} \operatorname{proj}_W(\mathbf{v}) &= \operatorname{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \operatorname{proj}_{\mathbf{u}_m}(\mathbf{v}) \\ &= \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}, \mathbf{u}_m \rangle \mathbf{u}_m. \end{aligned}$$

Note that this defines a linear transformation $\operatorname{proj}_W \colon V \to V$ with image W.

4.4 The Gram–Schmidt orthonormalisation process

Let V be an inner product space.

There is a procedure that starts with an arbitrary basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V and returns an orthonormal basis $\mathcal{C} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$.

Example 4.11. Let $W = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \subset \mathbb{R}^4$ with the dot product, where

$$\mathbf{v}_1 = (1,1,1,1) \qquad \mathbf{v}_2 = (2,4,2,4) \qquad \mathbf{v}_3 = (1,5,-1,3).$$

(a) Find an orthonormal basis for W.

(b) Find the point of W closest to the point $\mathbf{v} = (1,2,3,0)$.

4.5 Curve fitting

Given data points $(x_1, y_1), \ldots, (x_n, y_n)$ and a linear model of the form

$$y = a + bx,$$

we wish to find the values of the parameters a and b that exhibit the best fit for the data.

For an exact fit we would have

$$y_1 = a + bx_1$$
$$\vdots$$
$$y_n = a + bx_n,$$

in other words

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = a \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + b \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$y = A \quad u$$

It will typically be impossible to find $\mathbf{u}$ such that $\mathbf{y} - A\mathbf{u} = \mathbf{0}$, so we attempt the next best thing: Find $\mathbf{u} = \mathbf{u}_{\min}$ such that $\|\mathbf{y} - A\mathbf{u}\|$ is as small as possible.

Letting

$$W = \{A\mathbf{u} \mid \mathbf{u} \in \mathbb{R}^2\},$$

we are after $\mathbf{u}_{\min}$ such that $A\mathbf{u}_{\min}$ is the point of W closest to $\mathbf{y}$.

vertible In other words,

the orthogonal projection of y onto W

Note that $Au_{min} = \text{proj}_W y$

$$\|\mathbf{y} - A\mathbf{u}\|^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

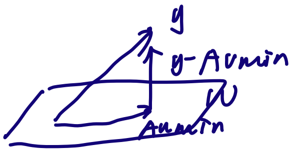
$y - Au = (y_1 - (a+bx_1),\ y_2 - (a+bx_2) \cdots,\ y_n - (a+bx_n))$

which is why this is known as the method of $\boxed{\text{least squares.}}$

$\begin{bmatrix} 1 & 0 & \frac{9}{7} \\ 0 & 1 & \frac{4}{7} \end{bmatrix}$

$y = \frac{9}{7}x + \frac{4}{7}$

Since $A\mathbf{u}_{\min} = \mathrm{proj}_W(\mathbf{y})$, we have that $\mathbf{y} - A\mathbf{u}_{\min}$ is orthogonal to W, therefore



when we take any vector in W $u \in \mathbb{R}^2$

$$(Au)\cdot(y - Au_{min}) = 0$$

$$u \cdot w = v_1 w_1 + \cdots + v_i w_i + \cdots + v_n w_n$$

$$= v^T \cdot w \quad \left(\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \right)$$

$$(Au)^T (y - Au_{min}) = 0$$

$$u^T A^T (y - Au_{min}) = 0$$

We conclude that $\mathbf{u}_{\min}$ is a solution of the equation

$$u^T(A^Ty - A^TA\,u_{min}) = 0$$

$$\underbrace{}_{\text{fixed}} = 0$$

$$A^T A\mathbf{u}_{\min} = A^T\mathbf{y},$$

or, if (A^TA) is invertible,

$$\boxed{\mathbf{u}_{\min} = (A^TA)^{-1}A^T\mathbf{y}.}$$

$\rightarrow$ A is not a square mat

A is n

in

we ca

simpli

Example 4.12. Find the straight line $y = a + bx$ that best fits the data points $(-1, 1)$, $(1, 1)$, $(2, 3)$.

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 3 & 2 & 5 \\ 2 & 6 & 6 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 1 & -4 & -1 \\ 1 & 3 & 3 \end{array}\right] \rightarrow$$

$$u_{min} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$u_{min} = \left(\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$\rightarrow \begin{bmatrix} 1 & 3 \\ 0 & -7 \end{bmatrix}$

$\star$ A^TA is invertible
need to be test

$$= \left(\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$$= \frac{1}{14}\begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 18 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{9}{7} \\ \frac{4}{7} \end{bmatrix}$$

The same method extends, with small adjustments, to more complicated models.

Suppose, for instance, that you want to fit the model

$$y = a + bx^2 + ce^x$$

to data points $(x_1, y_1), \ldots, (x_n, y_n)$.

...j = 0 for all $i \neq j$
...i = 1 for all i

4.6 Orthogonal matrices

An $n \times n$ matrix Q is an *orthogonal matrix* if

$$Q^T Q = I.$$

In particular, Q is invertible with $Q^T = Q^{-1}$

Example 4.13. $Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

$$Q^T Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

so Q is an orthogonal matrix

Proposition 4.14. An $n \times n$ matrix is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$.

proof:

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & \cdots & \cdots & q_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

$$Q^T = \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix}$$

$$Q^T Q = \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

$\underbrace{(n \times 1) \times (1 \times n)} \Rightarrow (n \times n)$

$$= \begin{bmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\ v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & v_2 \cdot v_n \\ \vdots & & & \\ v_n \cdot v_1 & \cdots & \cdots & v_n \cdot v_n \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix} \Longleftrightarrow v_i \cdot$$

$v_i \cdot$

Proposition 4.15. If Q is an $n \times n$ orthogonal matrix and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

$$(Q\mathbf{u}) \cdot (Q\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}.$$

ices

proof:

$$= (Qv)^T (Qv)$$

So orthogonal matrices preserve the dot product of vectors. Therefore they also preserve lengths and angles.

$$\|Qv\| = \sqrt{(Qv)\cdot(Qv)} = \sqrt{v\cdot v} = \|v\|$$

Geometrically, they correspond to

$$Q^TQ = I$$

$$\det(Q^TQ) = \det(I)$$

$$\det(Q)\cdot\det(Q) = \det(I) = 1$$

$\det(Q) = 1$ or -1 → reflection correspond with a rotation in R^n.

like $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ → rotate

273

An $n \times n$ matrix A is *symmetric* if $A^T = A$.

The spectral theorem for real matr

Theorem 4.16. If A is a symmetric $n \times n$ real matrix then

(a) all the eigenvalues of A are real $(A-\lambda I)=0$

(b) A has n orthonormal eigenvectors.

linear independent → diagonalisable

$-\lambda$

4

$\lambda)$

$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}s + \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}t$

So if A is symmetric then there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Hence A is diagonalisable:

$$Q^{-1}AQ = D \qquad \| \quad Q^TAQ = D$$

and the change of basis matrix Q is orthogonal.

(We say that A is *orthogonally diagonalisable*.)

eigenspace

Example 4.17. Given $A = \begin{bmatrix} 3 & 2 & -2 \\ 2 & 0 & 4 \\ -2 & 4 & 0 \end{bmatrix}$, orthogonally diagonalise A.

$$\det(A-\lambda I) = \begin{vmatrix} 3-\lambda & 2 & -2 \\ 2 & -\lambda & 4 \\ -2 & 4 & -\lambda \end{vmatrix}$$

$$= 3-\lambda \begin{vmatrix} -\lambda & 4 \\ 4 & -\lambda \end{vmatrix} - 2 \begin{vmatrix} 2 & 4 \\ -2 & -\lambda \end{vmatrix} - 2 \begin{vmatrix} 2 & \\ -2 & \end{vmatrix}$$

$$= (3-\lambda)(\lambda^2-16) - 2(-2\lambda+8) - 2(8-$$

$$= -(\lambda-4)^2(\lambda+5)$$

① $\lambda = 4$ ② $\lambda = -5$

$\lambda = 4$

$$A-\lambda I = \begin{bmatrix} -1 & 2 & -2 \\ 2 & -4 & 4 \\ -2 & 4 & -4 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

two free parameter

$x_3 = t$

$x_2 = s$

$x_1 = 2s - 2t$

$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} =$

for $\lambda = 4$

$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ eigenvector

$\lambda = -5$

$$\begin{bmatrix} 8 & 2 & -2 \\ 2 & +5 & 4 \\ -2 & 0 & +5 \end{bmatrix} \to \begin{bmatrix} 0 & -18 & -18 \\ 2 & 5 & 4 \\ 0 & 5 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 5 & 4 \\ 0 & 1 & 1 \\ 0 & 5 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$\begin{bmatrix} -2 & -\frac{1}{2} \\ 0 & -1 \\ 1 & 1 \end{bmatrix}$ & $P^{-1}AP = D$

not orthogonal diagonalisable

Apply Gram - Smith to each

$$D = \begin{bmatrix} 4 & & \\ & 4 & \\ & & -5 \end{bmatrix} \qquad P = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -1 \\ 1 \end{bmatrix} t$$

$x_3 = t$
$x_2 = -t$
$x_1 = \frac{1}{2}t$