1 Wed 9/16 Orthogonality

So far we have a vector space with operations + and $\cdot$. Now we define another operation, **inner product**: $x^T y = x_1 y_1 + \cdots + x_n y_n$. In general an inner product is a map $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$ with the following property:

(1) Positive definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle \geqslant 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(2) Scalar multiplication: $\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$.

(3) Addition: $\langle \mathbf{v_1} + \mathbf{v_2}, \mathbf{w} \rangle = \langle \mathbf{v_1}, \mathbf{w} \rangle + \langle \mathbf{v_2}, \mathbf{w} \rangle$.

Example 1.1

In $P_2(\mathbb{R})$, suppose $r_1, r_2 \in P_2(\mathbb{R})$. Suppose r_1, r_2 are continuous on the interval $[-1, 1]$. One way to define inner product is

$$\langle r_1, r_2 \rangle = \int_{-1}^{1} r_1 r_2 dx$$

Obviously $\int_{-1}^{1} (r_1)^2 dx \geqslant 0$ and the integral is 0 if and only if r is the zero polynomial. The other two are also obvious.

2 Fri 9.18

Theorem 1

If $\{v_1, \ldots, v_n\}$ are orthogonal, then they are linearly independent.

Proof

First note that orthogonality implies $v_i^T v_j = 0$ for $i \neq j$.

Suppose $c_1 v_1 + \cdots + c_n v_n = 0$. To show the vectors are linearly independent, we want to show that all the coefficients are 0.

Since $c_1 v_1 + \cdots + c_n v_n = 0$, multiplying everything by v_i^T gives

$$v_i^T [c_1 v_1 + \cdots + c_n v_n] = v_i T = 0 \implies c_i v_i^T v_i = 0$$

Clearly $v_i \neq 0 \implies v_i^T v_i \neq 0$. Hence $c_i = 0$. Since we've chosen c_i arbitrarily, all the coefficients are 0. Hence the vectors are linearly independent. □

Remark

Note that orthogonal $\implies$ linearly independent, but the converse is not true: linearly independent $\not\Longrightarrow$ orthogonal.

Proposition 2

Suppose $\{v_1, \ldots, v_n\}$ is an orthogonal basis for $\mathbb{R}^n$. Then we can write $\mathbf{v} \in \mathbb{R}^n$ as a linear combination of v_i's: $\mathbf{v} = c_1 v_1 + \cdots + c_n v_n$. Then

$$v^T v_i = (c_1 v_1 + \cdots + c_n v_n)^T v_i \implies v^T v_i = c_i v_i^T v_i \implies c_i = \frac{v^T v_i}{v_i^T v_i}$$

Therefore,

$$\mathbf{v} = \left(\frac{v^T v_1}{v_1^T v_1}\right) v_1 + \left(\frac{v^T v_2}{v_2^T v_2}\right) v_2 + \cdots + \left(\frac{v^T v_n}{v_n^T v_n}\right) v_n.$$

Gram-Schmidt Orthonormalization

Next question: how to construct an orthogonal basis from a set of linearly independent vectors?

Example 2.1

An example in $\mathbb{R}^2$: suppose we have $\mathbf{v_1}, \mathbf{v_2} \in \mathbb{R}^2$. To come up with an orthogonal basis, we fix $\mathbf{w_1} = \mathbf{v_1}$. To get $\mathbf{w_2}$, we set it as $\mathbf{v_2} - \text{proj}_{\mathbf{v_1}} \mathbf{v_2}$. In this way $\mathbf{w_2} \perp \mathbf{w_1}$. Simple algebra:

$$\mathbf{w_2}^T \mathbf{w_1} = 0 \implies (\mathbf{v_2} - c_1 \mathbf{v_1})^T \mathbf{v_1} = 0 \implies \mathbf{v_2}^T \mathbf{v_1} = c_1 \mathbf{v_1}^T \mathbf{v_1} \implies c_1 = \frac{\mathbf{v_2}^T \mathbf{v_1}}{\mathbf{v_1}^T \mathbf{v_1}}$$

Therefore $\mathbf{w_2} = \mathbf{v_2} - \left(\frac{\mathbf{v_2}^T \mathbf{v_1}}{\mathbf{v_1}^T \mathbf{v_1}}\right) \mathbf{v_1}$.

Example 2.2

For $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$, a possible way to construct a set of orthogonal vectors is

$$\mathbf{w_1} = \mathbf{v_1}$$
$$\mathbf{w_2} = \mathbf{v_2} - \left(\frac{\mathbf{v_2}^T \mathbf{v_1}}{\mathbf{v_1}^T \mathbf{v_1}}\right) \mathbf{v_1}$$
$$\mathbf{w_3} = \mathbf{v_3} - \left(\frac{\mathbf{v_3}^T \mathbf{v_1}}{\mathbf{v_1}^T \mathbf{v_1}}\right) \mathbf{v_1} - \left(\frac{\mathbf{v_3}^T \mathbf{v_2}}{\mathbf{v_2}^T \mathbf{v_2}}\right) \mathbf{v_2}$$

taking all the projections off and we get orthogonal vectors.

Example 2.3

Suppose $\mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{3} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. Apply Gram-Schmidt.

Solution

Let $\mathbf{w_1} = \mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Then

$$\mathbf{w_2} = \mathbf{v_2} - \left(\frac{\mathbf{v_2}^T \mathbf{v_1}}{\mathbf{v_1}^T \mathbf{v_1}} \right) \mathbf{v_1} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 0 = \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix}.$$

Finally,

$$\mathbf{w_3} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \frac{1}{1} \mathbf{v_1} - \frac{2}{1} \mathbf{v_2} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

3 Mon 9/21 Projections and Least Squares

More on projection and Gram-Schmidt. Note that

$$\text{proj}_a b = \left(\frac{b^T a}{a^T a} \right) a = a \left(\frac{b^T a}{a^T a} \right) = \frac{a a^T b}{a^T a} = \underbrace{\left(\frac{a a^T}{a^T a} \right)}_{\text{projection matrix}} b$$

Properties of the projection matrix P :

(1) $P^2 = P$: projecting twice is the same as projecting once.

(2) $P^T = P$: disregard the constant denominators, then $(aa^T)^T = (a^T)^T a^T = aa^T$.

Least Square Approximation

Suppose we are trying to solve

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix} x = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

which clearly has no solution. However, we can try to find $\overline{x}$ such that the error is minimized:

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} \overline{x}$$

is minimized. In other words, we want to make $e = b - Ax$ as small as possible. This can be achieved when e is perpendicular to b. Back to projection.

Example 3.1

Suppose we try to find a line through the points $(t_1, b_1), (t_2, b_2), (t_3, b_3)$ and the line has equation $Ct + D$. Then

$$\begin{cases} Ct_1 + D = b_1 \\ Ct_2 + D = b_2 \\ Ct_3 + D = t_3 \end{cases} \implies \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}.$$

Most of the time $\mathbf{t}$ may not lie on the plane spanned by the column space of the first matrix. The error is inevitable, but it can be minimized if we take the projection of $\mathbf{t}$ onto the plane. Therefore we are looking for $\overline{x}$ such that $e = b - A\overline{x}$ is orthogonal to the subspace and thus the column vectors.

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}^T e = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T e = 0 \implies A^T e = 0 \implies A^T(b - A\overline{x}) = 0 \implies A^T A\overline{x} = A^T b \implies \overline{x} = (A^T A)^{-1} A^T b$$

and

$$A\overline{x} = A(A^T A)^{-1} A^T b\overline{x} = Px \implies P = A(A^T A)^{-1} A^T (\text{compare this with } P = \frac{aa^T}{a^T a} \text{ for one column vector } a)$$

4 Wed 9/23 Least Squares

Example 4.1

Suppose we have measured some data of a spring – displacement vs force:

Force (N)	1	2	3	4
Displacement (m)	3	4	7	8

Find the best fit line for these data where the line can be written as $y = AF + B$.

Solution

It is obvious that we cannot find a straight line that goes through all four points $(1,3),(2,4),(3,7)$, and $(4,8)$. There has to be some errors. Nevertheless, we proceed and write out the equations first:

$$\begin{cases} 1A + B = 3 \\ 2A + B = 4 \\ 3A + B = 7 \\ 4A + B = 8 \end{cases} \implies \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 7 \\ 8 \end{bmatrix} \sim A\mathbf{x} = \mathbf{b}$$

We cannot reduce error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ to 0, but we can minimize it by making its components the least, that is, being orthogonal to each respective vector from A.

Then,

$$A^T(\mathbf{b} - A\overline{x}) = 0 \implies A^T\mathbf{b} = A^TA\overline{x} \implies \overline{x} = (A^TA)^{-1}A^T\mathbf{b}.$$

Time to compute:

$$A^TA = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \implies (A^TA)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix}$$

Then

$$\overline{x} = \frac{1}{20} \begin{bmatrix} 3 & -10 \\ -10 & 30 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 1 \end{bmatrix}.$$

Then the best fit line has equation $y = 0.8x + 1$.

Remark

Think of the column space of A. It is a plane in $\mathbb{R}^4$ spanned by the two column vectors. Since $A\mathbf{x} = \mathbf{b}$ has no solution, we know $\mathbf{b}$ does not lie on that plane. Therefore, to minimize the error $\mathbf{e}$, we find the *shortest* distance between $\mathbf{b}$ and $C(A)$, and this is the orthogonal vector between $\mathbf{b}$ and $C(A)$. Since $\mathbf{e}$ is orthogonal to the plane $C(A)$, it is orthogonal to both column vectors since they lie on the plane. Therefore we have two equations and we can write them as one matrix multiplication: $A^T\mathbf{e} = 0$.