Why Do we Care about Orthogonality?

Feb 11, 2025

There are infinitely many reasons to care about orthogonality, we only present one used in function approximation.

Orthogonality helps us because finding a "best approximate" will be equivalent to solving a big system of linear equations. This is very computationally involved. But using orthogonal basis make these computation much more efficient.

Goal: We are given a complicated function f, and want to approximate / estimate f using simpler or better understood function.

Our setting is in Inner Product spaces, and we will say $\hat{f}$ is a good approximation / close to f when $\|f-\hat{f}\|$ small.

Setup: Let V be an inner product space, over $\mathbb{R}$, and $\{w_1, w_2, \ldots, w_n\} \subseteq V$ linearly independent, s.t $W = \text{span}\{w_1, \ldots, w_n\}$ a n-dim vector space.
The complicated function f lives in V, and we want to find $\tilde{f} \in W$ to is a "good" approximation for f.
The vectors $\{w_1, \ldots, w_n\}$ will always be known and "easy to understand" functions.

We say $\tilde{f} \in W$ is a <u>best approximation</u> to f wrt W to mean

$$\|f - \tilde{f}\| = \inf_{w \in W} \|f - w\|.$$

⊛ One can show that in inner-product spaces, we always have a unique best approximation (Friedberg-Insel-Spence, section 6.2)

Process: Fix $v \in V$, and for any $w \in W$

write $w = \sum_{i=1}^{n} \alpha_i w_i \in W$. (since $\{w_i\}$ basis for W.)

We want to estimate $\| v - w \|$:

$$\| v-w \|^2 = \langle v-w, v-w \rangle$$

$$= \langle v - \sum_{i=1}^{n} \alpha_i w_i , v - \sum_{j=1}^{n} \alpha_j w_j \rangle$$

$$= \langle v, v \rangle - \sum_{j=1}^{n} \alpha_j \langle v, w_j \rangle$$

$$- \sum_{i=1}^{n} \alpha_i \langle w_i, v \rangle + \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \alpha_j \langle w_i, w_j \rangle$$

working over $\mathbb{R}$

$$= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \alpha_j \langle w_i, w_j \rangle - 2 \sum_{j=1}^{n} \alpha_j \langle v, w_j \rangle$$

$$+ \langle v, v \rangle$$

$$= F(\alpha_1, \ldots, \alpha_n)$$

Since $\langle v, v \rangle$, $\langle w_i, w_j \rangle$, $\langle v, w_j \rangle$ known and fixed

we can treat them as constants, and view

$F : \mathbb{R}^n \to \mathbb{R}$ as a polynomial in $\alpha_1, \dots, \alpha_n$.

Recall we want to minimize $\|v - w\|^2$. That is we want to minimize F. So set $\frac{\partial F}{\partial x_k} = 0$, $\forall k$ and solve to find critical points.

Now observe that calculating $\frac{\partial F}{\partial \alpha_k}$ we get

$$\frac{\partial}{\partial \alpha_k}\left(\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \langle w_i, w_j \rangle - 2 \sum_{j=1}^{n} \alpha_j \langle v, w_j \rangle + \langle v, v \rangle \right)$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_k}\left(\sum_{j=1}^{n} \alpha_i \alpha_j \langle w_i, w_j \rangle \right) - 2\langle v, w_k \rangle + 0$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_k}\left(\alpha_i \sum_{j=1}^{n} \alpha_j \langle w_i, w_j \rangle \right) - 2\langle v, w_k \rangle$$

$i = k$ $\qquad\qquad$ $i \neq k$

$$= \frac{\partial}{\partial \alpha_k}\left(\alpha_k^2 \langle w_k, w_k \rangle + \alpha_k \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j \langle w_k, w_j \rangle + \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{\partial}{\partial \alpha_k}\left(\alpha_i \sum_{j=1}^{n} \alpha_j \langle w_i, w_j \rangle \right) \right.$$

$$- 2\langle v, w_k \rangle$$

$$= 2\alpha_k \langle w_k, w_k \rangle + \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j \langle w_k, w_j \rangle + \sum_{\substack{i=1 \\ i \neq k}}^{n} \alpha_i \langle w_i, w_k \rangle$$

$$- 2\langle v, w_k \rangle$$

$$= 2 \sum_{j=1}^{n} \alpha_j \langle w_j, w_k \rangle - 2\langle v, w_k \rangle$$

• Now setting $\frac{F}{\partial x_k} = 0$ means $\forall k$.

$$\begin{cases} \sum_{j=1}^{n} \alpha_j \langle w_1, w_j \rangle = \langle v, w_1 \rangle \\ \vdots \\ \sum_{j=1}^{n} \alpha_j \langle w_n, w_j \rangle = \langle v, w_n \rangle \end{cases}$$

(n-equations)

(n-unknowns)

But this can be represented as a Matrix sol:

$$\Rightarrow \begin{bmatrix} \langle w_k, w_1 \rangle & \langle w_k, w_2 \rangle & \cdots & \langle w_k, w_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \langle v, w_k \rangle$$

The above tells us what the k^{th} row of our matrix looks like.

I.e

$$\underbrace{\begin{bmatrix} \langle w_1, w_1 \rangle & \langle w_1, w_2 \rangle & \cdots & \langle w_1, w_n \rangle \\ \langle w_2, w_1 \rangle & \langle w_2, w_2 \rangle & & \vdots \\ \vdots & & & \\ \langle w_n, w_1 \rangle & & \cdots & \langle w_n, w_n \rangle \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\alpha} = \underbrace{\begin{bmatrix} \langle v, w_1 \rangle \\ \langle v, w_2 \rangle \\ \vdots \\ \langle v, w_n \rangle \end{bmatrix}}_{b} \quad (*)$$

(symmetric if V a $\mathbb{R}$ vspace, conj-symmetric if V over $\mathbb{C}$)

Where we call the matrix above the **Gram matrix** associated to the basis w.

- Now find $(\alpha_1, \dots, \alpha_n)$ that minimize $F(\alpha_1, \dots, \alpha_n)$.

- So far: Solving ⊛ gives us access to finding the best approximate, since the best approximate is just

$$w_0 = \sum_{i=1}^{n} \tilde{\alpha}_i w_i \quad , \text{ where } \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$$

solution to ⊛.

Now ⊛ is very computationally heavy. Solving systems of equation where the matrix is diagonal is much more efficient!

So if we started with $\{w_1, \dots, w_n\}$ an orthogonal set, then the Gram matrix would look like.

$$\begin{bmatrix} \langle w_1, w_1 \rangle & & & 0 \\ & \langle w_2, w_2 \rangle & & \\ & & \ddots & \\ 0 & & & \langle w_n, w_n \rangle \end{bmatrix} \quad !$$

Practical Example yet to come!