Hilbert Spaces and the Projection Theorem in a Nutshell

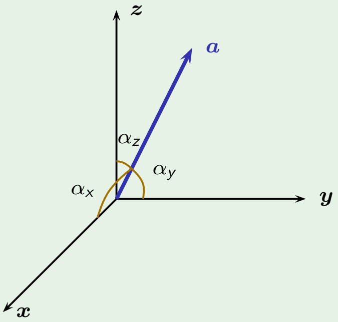
What is a Hilbert Space?

A Hilbert space is nothing but a collection of elements with very nice properties, over which we can define two functions:

- An *inner product* $< \cdot, \cdot >$ between two elements.
  The inner product is related to the notion of projection of one element onto the other.
  When the inner product between two elements is zero, they are said to be *orthogonal*;
- A *norm* $\|\cdot\|$ of an element.
  The norm is related to the notion of energy.
  The norm of the difference between two elements represents the *distance* between them, that is, the *energy of the error* between them.

Did you now that ...

The Euclidean space is a Hilbert space



Given a vector $\boldsymbol{a} = a_x\boldsymbol{x} + a_y\boldsymbol{y} + a_z\boldsymbol{z}$

- The norm is the length of the vector, *i.e.*, $\|\boldsymbol{a}\| = \left(a_x^2 + a_y^2 + a_z^2\right)^{\frac{1}{2}}$
- The inner product is the scalar product, *e.g.*,

$$< \boldsymbol{a}, \boldsymbol{c} >= \|\boldsymbol{a}\| \, \|\boldsymbol{c}\| \cos \alpha_{ab}$$

where α_{ab} is the angle between the two vectors $\boldsymbol{a}$ and $\boldsymbol{c}$. In particular

$$< \boldsymbol{a}, \boldsymbol{x} >= \|\boldsymbol{a}\| \cos \alpha_x = a_x$$
$$< \boldsymbol{a}, \boldsymbol{y} >= \|\boldsymbol{a}\| \cos \alpha_y = a_y$$
$$< \boldsymbol{a}, \boldsymbol{z} >= \|\boldsymbol{a}\| \cos \alpha_z = a_z$$

Did you now that ...

Random variables with finite variance form a Hilbert space

The collection of random variables X such that $\mathrm{Var}\left(X\right) < \infty$ is denoted as $\mathrm{L}^2(P)$.

- The inner product is

$$\langle A, B \rangle_{\mathrm{L}^2(P)} = \mathrm{cov}\left(A, B\right) = \mathrm{E}\left[(A - \mathrm{E}\left[A\right])(B - \mathrm{E}\left[B\right])^*\right] .$$

- The norm is

$$\|A\|_{\mathrm{L}^2(P)} = \mathrm{Var}\left(A\right) = \mathrm{E}\left[\left|A - \mathrm{E}\left[A\right]\right|^2\right] .$$

w.s.s. processes form a Hilbert space

This is just a particular case of the space of random variables with finite variance: For a w.s.s. process $X[n]$ we have, by definition, $\mathrm{E}\left[\left|X[n]\right|^2\right] < \infty$.

A Key Tool in Hilbert Spaces: The Projection Theorem

The Projection Theorem

Let E and S be two Hilbert spaces such that $S \subset E$. Then, for every element $\boldsymbol{a}$ of E it exists a unique element $\boldsymbol{b}$ of S such that

- $\boldsymbol{b}$ is the element of S that minimizes the norm of the difference $\boldsymbol{a} - \boldsymbol{b}$, that is

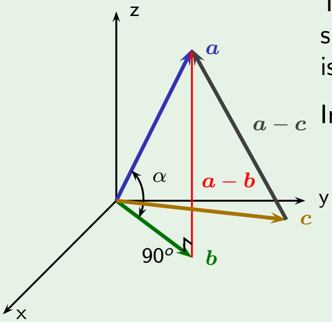
$$\boldsymbol{b} = \arg\min_{\boldsymbol{c} \in S} \|\boldsymbol{a} - \boldsymbol{c}\| \ ;$$

- The difference $\boldsymbol{a} - \boldsymbol{b}$ is orthogonal to any element in S, that is

$$\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{c} \rangle = 0 \, , \quad \forall \ \boldsymbol{c} \in S \, ,$$

The element $\boldsymbol{b}$ is called the *orthogonal projection of $\boldsymbol{a}$ onto S*.

The Projection Theorem in the Euclidean space $\mathbb{R}^3$



The projection of the vector $\boldsymbol{a}$, defined in the space $E = \mathbb{R}^3$, onto the plane $S = \mathbb{R}^2$ ($S \subset E$), is given by $\boldsymbol{b}$.

In particular:

- $\boldsymbol{b}$ is the vector in the plane that minimizes the norm of the difference $\boldsymbol{a}-\boldsymbol{b}$

$$\|\boldsymbol{a}-\boldsymbol{b}\| \leq \|\boldsymbol{a}-\boldsymbol{c}\| \ \forall \ \boldsymbol{c}$$

- The difference $\boldsymbol{a}-\boldsymbol{b}$ is orthogonal to any vector in the plane $S = \mathbb{R}^2$.

The Projection Theorem in the space of w.s.s. processes

The projection theorem can be rephrased as follows.

Let E and S be two spaces of w.s.s. signals such that $S \subset E \subseteq \mathrm{L}^2(P)$. Then, for every w.s.s. signal $X[n]$ in E it exists a unique w.s.s. signal $Y[n]$ in S such that

- $Y[n]$ is the signal in S minimizing the norm of $X[n] - Y[n]$, that is

$$Y[n] = \arg\min_{U[n] \in S} \mathrm{E}\left[\left| X[n] - U[n] \right|^2 \right] ;$$

- The difference $X[n] - Y[n]$ is orthogonal to any element in S, that is

$$\mathrm{E}\left[(X[n] - Y[n]) U[n]^* \right] = 0 , \quad \forall\ U[n] \in S ,$$

The w.s.s. signal $Y[n]$ is the *orthogonal projection of $X[n]$ onto S*.

The Projection Theorem in the space of w.s.s. processes

Consider the particular case of an AR process (of order 3) in canonical form

$$X[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] + W[n]\,.$$

Call E the space generated by $X[k]$, for all $k \leq n$, and by $W[n]$.
Call S the space generated by $X[k]$, for all $k \leq n-1$, where $S \subset E$.

Then the orthogonal projection of $X[n] \in E$ onto S is given by

$$Y[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3]\,.$$

Indeed, for every vector in S, that is for every $X[k]$ with $k \leq n-1$,

$$\begin{aligned} &< X[n] - Y[n], X[k] >= \\ &\qquad < X[n] - (-p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3]), X[k] >= \\ &\qquad\qquad < W[n], X[k] >= \operatorname{cov}(W[n], X[k]) = 0\,. \end{aligned}$$

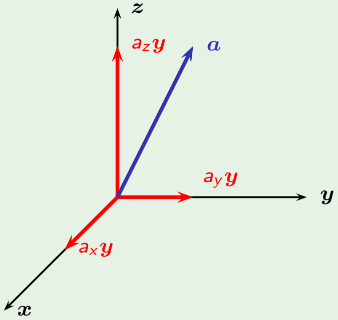
Intuitive property

Let $\boldsymbol{a}$ be a element of E that we can write as

$$\boldsymbol{a} = \boldsymbol{v} + \boldsymbol{u}$$

where $\boldsymbol{u} \in S$ and $\boldsymbol{v}$ is orthogonal to any element in S. Then, the orthogonal projection of $\boldsymbol{a}$ onto S is straightforwardly given by $\boldsymbol{u}$.

Intuitive property in the Euclidean space



$$\boldsymbol{a} = a_x\boldsymbol{x} + a_y\boldsymbol{y} + a_z\boldsymbol{z}$$

where $\boldsymbol{u} = a_x\boldsymbol{x} + a_y\boldsymbol{y} \in S = \mathbb{R}^2$ and $\boldsymbol{v} = a_z\boldsymbol{z}$ orthogonal to any vector in $S = \mathbb{R}^2$.

Therefore, the orthogonal projection of $\boldsymbol{a}$ onto $S = \mathbb{R}^2$ gives

$$\boldsymbol{b} = \boldsymbol{u} = a_x\boldsymbol{x} + a_y\boldsymbol{y}\,.$$

Intuitive property for w.s.s. Signals

Let $X[n]$ be a element of E that we can write as $X[n] = U[n] + V[n]$, where $U[n] \in S$ and $V[n]$ is orthogonal to any element in S. Then, the orthogonal projection of $X[n]$ onto S is straightforwardly given by $U[n]$.

In the specific case of AR process (of order 3) in canonical form

$$X[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] + W[n]\,,$$

we have $U[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] \in S$ and $V[n] = W[n]$ orthogonal any vector in S.

Therefore the orthogonal projection of $X[n]$ onto S gives

$$Y[n] = U[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3]\,.$$

Applications of the Projection Theorem for w.s.s. Signals

Optimal Predictor as a Projection onto the Past

Let $X[n]$ be a w.s.s. stochastic process.

Call S=space formed by the linear combinations of $X[n-1], \ldots, X[n-N]$, and E=space formed by the linear combinations of $X[n+k], \ldots, X[n], \ldots, X[n-N]$.

To predict $X[n+k] \in E$ we use a linear combination of its past $a_1 X[n-1] + \ldots + a_N X[n-N]$. Notice that, clearly, such a linear combination belongs to S.

Among all the possible linear combinations we choose as optimal the one that minimizes the mean square error (MSE)

$$\mathrm{E}\left[\left| X[n+k] - (a_1 X[n-1] + \ldots + a_N X[n-N]) \right|^2 \right] .$$

Gosh! Do I now really have to minimize the MSE by computiting all the derivatives w.r.t. a_n and set them to zero?

Optimal Predictor as a Projection onto the Past

No effort with the Projection Theorem!

Recall that to predict $X[n+k] \in E$ we search for the linear combination of its past $a_1 X[n-1] + \ldots + a_N X[n-N]$ that minimizes the MSE

$$\mathrm{E}\left[\left| X[n+k] - (a_1 X[n-1] + \ldots + a_N X[n-N]) \right|^2 \right].$$

In Summary: we want to estimate an element of E with the element $Y[n]$ of S that minimizes the MSE $\mathrm{E}\left[\left| X[n+k] - Y[n] \right|^2 \right]$. By the projection theorem

- $Y[n]$ is the orthogonal projection of $X[n+k]$ onto S;
- $Y[n]$ is such that $\mathrm{E}\left[(X[n+k] - Y[n]) U[n]^* \right] = 0$, for all $U[n] \in S$.

Remark: Since $S=$ all the linear combinations of $X[n-1], \ldots, X[n-N]$ the last property is equivalent to $\mathrm{E}\left[(X[n+k] - Y[n]) X[n-l]^* \right] = 0$ for $l = 1, \ldots, N$.

Therefore instead of computing annoying derivatives we just have to solve the linear system

$$\mathrm{E}\left[(X[n+k] - Y[n]) X[n-l]^* \right] = 0\,, \quad l = 1, \ldots, N\,.$$