Orthogonal projections

Once again, suppose that given $\boldsymbol{x} \in \mathcal{S}$, we want to find the closest point in a subspace $\mathcal{T}$. Recall that if we have an orthobasis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ for $\mathcal{T}$, then the closest point $\hat{\boldsymbol{x}}$ can be obtained via the simple formula

$$\hat{\boldsymbol{x}} = \sum_{n=1}^{N} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle \, \boldsymbol{v}_n.$$

We can also think of $\hat{\boldsymbol{x}}$ as the **orthogonal projection** of $\boldsymbol{x}$ onto $\mathcal{T}$. Specifically, we will use the notation $\boldsymbol{P}_{\mathcal{T}}[\cdot]$ for the **projection operator** onto $\mathcal{T}$. $\boldsymbol{P}_{\mathcal{T}}[\cdot]$ takes a signal and returns the signal in $\mathcal{T}$ closest to the input. Using this notation, we have

$$\boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}] = \sum_{n=1}^{N} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle \, \boldsymbol{v}_n.$$

We note that, by virtue of being a projection, $\boldsymbol{P}_{\mathcal{T}}$ satisfies a number of useful properties that will come in handy:

1. For any $\boldsymbol{x} \in \mathcal{T}$, $\boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}] = \boldsymbol{x}$. This can easily be verified by noting that if $\boldsymbol{x} \in \mathcal{T}$ we can write $\boldsymbol{x} = \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n$ and thus

$$\begin{aligned}
\boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}] &= \boldsymbol{P}_{\mathcal{T}}\left[\sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n\right] \\
&= \sum_{\ell=1}^{N} \left\langle \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n, \boldsymbol{v}_\ell \right\rangle \boldsymbol{v}_\ell \\
&= \sum_{\ell=1}^{N} \sum_{n=1}^{N} \alpha_n \langle \boldsymbol{v}_n, \boldsymbol{v}_\ell \rangle \, \boldsymbol{v}_\ell = \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n.
\end{aligned}$$

2. As a consequence, we also have that $\boldsymbol{P}_{\mathcal{T}}$ is **idempotent**, meaning that $\boldsymbol{P}_{\mathcal{T}}^2 = \boldsymbol{P}_{\mathcal{T}}$.
3. We can also define the complementary projection $\boldsymbol{Q}_{\mathcal{T}} = \mathbf{I} - \boldsymbol{P}_{\mathcal{T}}$, which computes the residual $\boldsymbol{x} - \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}]$. From the orthogonality principle we know that for any $\boldsymbol{x}$, $\boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}]$ and $\boldsymbol{Q}_{\mathcal{T}}[\boldsymbol{x}]$ are orthogonal. It is not difficult to show that $\boldsymbol{Q}_{\mathcal{T}}$ is also an orthogonal projection. Indeed, $\boldsymbol{Q}_{\mathcal{T}}$ can be constructed similarly to $\boldsymbol{P}_{\mathcal{T}}$ provided we have an orthobasis for the subspace of $\mathcal{S}$ which is orthogonal to $\mathcal{T}$.

We can say just a little more about the last property. What we are essentially doing here is decomposing the space $\mathcal{S}$ into two orthogonal subspaces, $\mathcal{T}$ and all of the vectors in $\mathcal{S}$ which are orthogonal to $\mathcal{T}$. We denote this set by $\mathcal{T}^\perp = \mathcal{S} \ominus \mathcal{T}$. One can also view this as building up the space $\mathcal{S}$ via the **direct sum** $\mathcal{S} = \mathcal{T} \oplus \mathcal{T}^\perp$.

One consequence of the orthogonality between the projections onto $\mathcal{T}$ and $\mathcal{T}^\perp$ is that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$, we have that

$$\langle \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}], (\mathbf{I} - \boldsymbol{P}_{\mathcal{T}})[\boldsymbol{y}] \rangle = \langle \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}], \boldsymbol{y} - \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{y}] \rangle = 0.$$

Similarly,

$$\langle \boldsymbol{x} - \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}], \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{y}] \rangle = 0.$$

From these we have the useful and intuitive facts that

$$\langle \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}], \boldsymbol{y} \rangle = \langle \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}], \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{y}] \rangle = \langle \boldsymbol{x}, \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{y}] \rangle.$$

Note also that since $\mathcal{T}$ and $\mathcal{T}^\perp$ are orthogonal, if $\|\cdot\|_S$ denotes the induced norm, then from Pythagoras we have that for any $\boldsymbol{x} \in \mathcal{S}$,

$$\|\boldsymbol{x}\|_S^2 = \|\boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}]\|_S^2 + \|(\mathbf{I} - \boldsymbol{P}_{\mathcal{T}})[\boldsymbol{x}]\|_S^2.$$

Subspace projections and linear approximation

Say $\{\boldsymbol{v}_k\}_{k=0}^{\infty}$ is an orthobasis for a Hilbert space $\mathcal{S}$. Let $\mathcal{T}$ be the subspace spanned by the first 10 elements of $\{\boldsymbol{v}_k\}$:

$$\mathcal{T} = \operatorname{span}\left(\{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_9\}\right).$$

1. Given $\boldsymbol{x} \in \mathcal{S}$, what is the closest point in $\mathcal{T}$ (call it $\hat{\boldsymbol{x}}$) to $\boldsymbol{x}$? We have seen that it is given by the projection

$$\hat{\boldsymbol{x}} = \boldsymbol{P}_{\mathcal{T}}[\boldsymbol{x}] = \sum_{k=0}^{9} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k.$$

2. How good an approximation is $\hat{\boldsymbol{x}}$ to $\boldsymbol{x}$? If we measure this in the induced norm $\|\cdot\|_S$, then

$$\begin{aligned}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_S^2 &= \left\| \sum_{k=0}^{\infty} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k - \sum_{k=0}^{9} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k \right\|_S^2 \\ &= \left\| \sum_{k=10}^{\infty} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k \right\|_S^2 \\ &= \sum_{k=10}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_k \rangle|^2.\end{aligned}$$

Since we also have

$$\|\boldsymbol{x}\|_S^2 = \sum_{k=0}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_k \rangle|^2$$

the (relative) approximation error for $\hat{x}$ will be small if the first 10 transform coefficients

$$\langle \boldsymbol{x}, \boldsymbol{v}_0 \rangle,\ \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle,\ \ldots,\ \langle \boldsymbol{x}, \boldsymbol{v}_9 \rangle,$$

contain "most" of the total energy.

Of course, there is nothing special about taking the first 10 coefficients. We can just as easily form a K term approximation using

$$\hat{\boldsymbol{x}}_K = \sum_{k=0}^{K-1} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k$$

which has error

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}_K\|_S^2 = \sum_{k=K}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_k \rangle|^2.$$

If the sum above is small for moderately large K, we can "compress" $\boldsymbol{x}$ by using just the first K terms in the expansion.

This is precisely what is done in image and video compression — more details on this to come soon!

Example:
Any real-valued function on $[-1/2, 1/2]$ with even symmetry can be built up out of harmonic cosines:

$$x(t) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(2\pi k t).$$

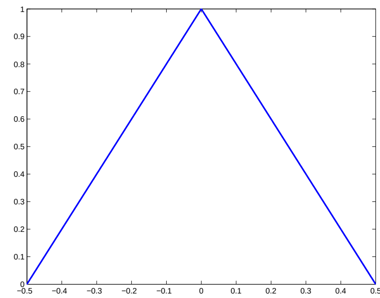
(That this is true follows directly from the observation that every signal on $[-1/2, 1/2]$ that is real-valued and even has a Fourier series which is real-valued and even.) This is an orthobasis expansion in the standard inner product with

$$v_0(t) = 1,\ v_1(t) = \sqrt{2}\cos(2\pi t),\ \ldots,\ v_k(t) = \sqrt{2}\cos(2\pi k t),\ \ldots$$

It is easy to check that $\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell \rangle = 0,\ k \neq \ell$ and $\langle \boldsymbol{v}_k, \boldsymbol{v}_k \rangle = 1$.

For the triangle function below

$$x(t) = \begin{cases} 1+2t, & -1/2 \le t \le 0 \\ 1-2t, & 0 \le t \le 1/2 \end{cases}$$



the expansion coefficients are

$$\begin{aligned} \alpha_0 &= 1/2, \\ \alpha_k &= \int_{-1/2}^{1/2} x(t)\sqrt{2}\cos(2\pi k t)\,\mathrm{d}t \\ &= 2\sqrt{2}\int_0^{1/2}(1-2t)\cos(2\pi k t)\,\mathrm{d}t \\ &= \begin{cases} 0 & k \text{ even, } k \neq 0 \\ \frac{2\sqrt{2}}{\pi^2 k^2} & k \text{ odd} \end{cases}. \end{aligned}$$

First, let's compute the norm in time and coefficient space just to make sure they agree:

$$\|\boldsymbol{x}\|_2^2 = \int_{-1/2}^{1/2} |x(t)|^2\,\mathrm{d}t = 2\int_0^{1/2}(1-2t)^2\,\mathrm{d}t = 1/3,$$

and

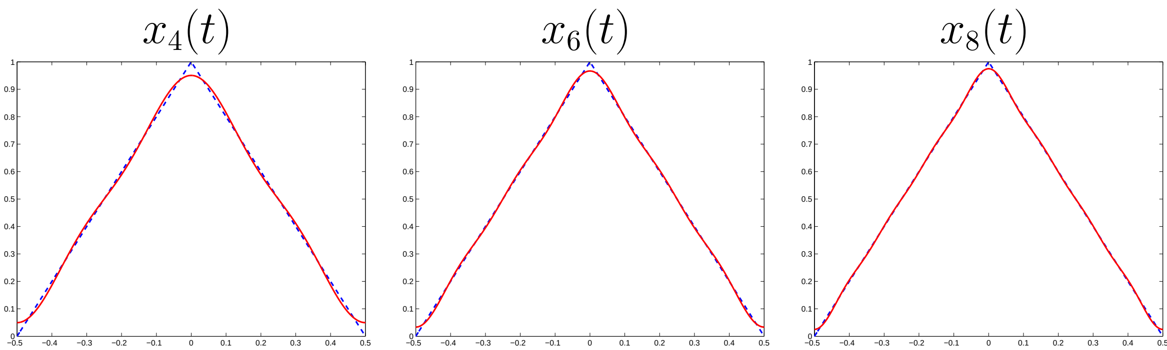
$$\begin{aligned} \sum_{k=0}^{\infty} |\alpha_k|^2 &= \frac{1}{4} + \frac{8}{\pi^4}\sum_{k'=0}^{\infty}\frac{1}{(1+2k')^4} \\ &= \frac{1}{4} + \frac{8}{\pi^4}\left(\frac{\pi^4}{96}\right) \\ &= \frac{1}{3}. \end{aligned}$$

When we truncate the expansion at K terms,

$$x_K(t) = \frac{1}{2} + \sum_{k=1}^{K-1} \alpha_k \sqrt{2}\cos(2\pi k t),$$

we can interpret the result as an **approximation** of $x(t)$ that is a member of the K-dimensional subspace $\operatorname{span}(\{\sqrt{2}\cos(2\pi k t\}_{k=0}^{K-1})$, and we know that it is the best approximation in that subspace.

Here are the approximation for $K = 4, 6, 8$:



We can compute the error in each of these approximations explicitly, as

$$\begin{aligned} x(t) - x_K(t) &= \sum_{k=0}^{\infty} \alpha_k \sqrt{2}\cos(2\pi k t) - \sum_{k=0}^{K-1} \alpha_k \sqrt{2}\cos(2\pi k t) \\ &= \sum_{k=K}^{\infty} \alpha_k \sqrt{2}\cos(2\pi k t), \end{aligned}$$

and so

$$\|x(t) - x_K(t)\|_2^2 = \sum_{k=K}^{\infty} |\alpha_k|^2,$$

or, since $x_K(t) \perp x(t) - x_K(t)$,

$$\|x(t) - x_K(t)\|_2^2 = \|x(t)\|_2^2 - \|x_K(t)\|_2^2.$$

In the three examples above, we have

$$\|x(t) - x_4(t)\|_2^2 \approx 1.92 \cdot 10^{-4}, \quad \|x(t) - x_6(t)\|_2^2 \approx 6.01 \cdot 10^{-5},$$
$$\|x(t) - x_8(t)\|_2^2 \approx 2.59 \cdot 10^{-5}.$$

The Gram-Schmidt algorithm

We have seen that orthobases for a Hilbert space (or a subspace) have many nice properties. Given any basis $\{\boldsymbol{v}_n\}_{n=1}^N$ for an N-dimensional space (or subspace), we can turn it into and orthobasis using the **Gram-Schmidt algorithm**.

The goal is to take a sequence of signals $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ and produce $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_N\}$ such that

$$\operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}) \;=\; \operatorname{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\})$$

and

$$\langle \boldsymbol{u}_n, \boldsymbol{u}_\ell \rangle = \begin{cases} 1 & n = \ell, \\ 0 & n \neq \ell \end{cases}.$$

That is, $\{\boldsymbol{u}_n\}$ spans the same space as $\{\boldsymbol{v}_n\}$, but it is a orthobasis.

1. Choose $\boldsymbol{w}_1 = \boldsymbol{v}_1$ and normalize it to get[1]

   $$\boldsymbol{u}_1 = \frac{\boldsymbol{w}_1}{\|\boldsymbol{w}_1\|}.$$

   Clearly, $\boldsymbol{u}_1$ is an orthobasis for span($\{\boldsymbol{v}_1\}$).

2. To get $\boldsymbol{u}_2$, we subtract from $\boldsymbol{v}_2$ its projection onto $\boldsymbol{u}_1$:

   $$\begin{aligned} \boldsymbol{w}_2 &= \boldsymbol{v}_2 - \langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle \boldsymbol{u}_1 \\ \boldsymbol{u}_2 &= \frac{\boldsymbol{w}_2}{\|\boldsymbol{w}_2\|} \end{aligned}$$

[1] The norm here and below is the one induced by the inner product.

Note that $\boldsymbol{u}_2$ is orthogonal to $\boldsymbol{u}_1$ by the orthogonality principle, but just to make sure

$$\begin{aligned}\langle \boldsymbol{u}_2, \boldsymbol{u}_1 \rangle &= \frac{1}{\|\boldsymbol{w}_2\|}\langle \boldsymbol{w}_2, \boldsymbol{u}_1 \rangle \\ &= \frac{1}{\|\boldsymbol{w}_2\|}\left(\langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle - \langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle \langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle\right) \\ &= 0.\end{aligned}$$

So $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthobasis for $\operatorname{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\})$.

3. At the beginning of the k^{th} step, $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}\}$ is an orthobasis for $\operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}\})$. We get $\boldsymbol{u}_k$ by subtracting off its projection onto $\operatorname{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}\})$ and normalizing:

$$\begin{aligned}\boldsymbol{w}_k &= \boldsymbol{v}_k - \sum_{\ell=1}^{k-1} \langle \boldsymbol{v}_k, \boldsymbol{u}_\ell \rangle \boldsymbol{u}_\ell, \\ \boldsymbol{u}_k &= \frac{\boldsymbol{w}_k}{\|\boldsymbol{w}_k\|}.\end{aligned}$$

By induction, $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is and orthobasis for $\operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\})$.

Note: If at any point

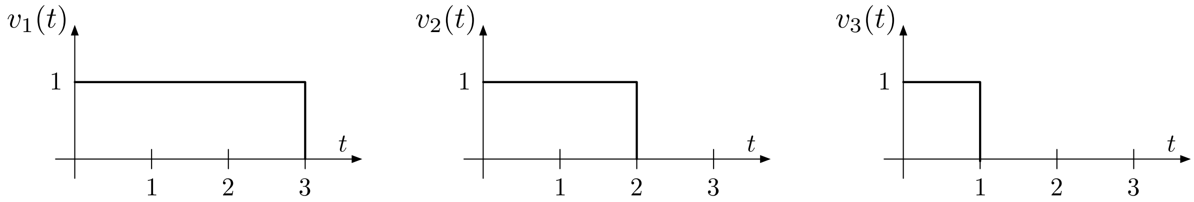
$$\boldsymbol{v}_k \in \operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}\})$$

(which means the $\{\boldsymbol{v}_n\}$ are linearly dependent — and not a basis), we will have

$$\boldsymbol{u}_k = \mathbf{0}.$$

When this happens, we can simply throw away $\boldsymbol{u}_k, \boldsymbol{v}_k$ and move on. The set of $\{\boldsymbol{u}_k\}$ will be smaller than N, but will still be an orthobasis for $\operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\})$.

Exercise: Let $\mathcal{S}$ be the space of piecewise-constant signals on $[0,1),[1,2),[2,3]$ with the standard L_2 inner product. Turn the following basis



into an orthobasis using Gram-Schmidt.