Orthogonal bases

A collection of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$ in a finite dimensional vector space $\mathcal{S}$ is called an **orthogonal basis** if

1. $\operatorname{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}) = \mathcal{S}$,
2. $\boldsymbol{v}_j \perp \boldsymbol{v}_k$ (i.e. $\langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle = 0$) for all $j \neq k$.

If in addition the vectors are normalized (under the induced norm),

$$\|\boldsymbol{v}_n\| = 1, \quad \text{for} \ \ n = 1, \ldots, N,$$

we will call it an **orthonormal basis** or **orthobasis**.

A note on infinite dimensions

In infinite dimensions, we need to be a little more careful with what we mean by "span". If $\mathcal{B} = \{\boldsymbol{v}_n\}_{n \in \mathbb{Z}}$ is an infinite sequence of orthogonal vectors in a Hilbert space $\mathcal{S}$, it is an orthobasis if the *closure* of $\operatorname{span}(\mathcal{B})$ is $\mathcal{S}$; this is written

$$\operatorname{cl} \operatorname{Span}\left(\{\boldsymbol{v}_n\}_n\right) = \mathcal{S}.$$

We don't need to get into too much, but basically this means that every vector in $\mathcal{S}$ can be approximated arbitrarily well by a finite linear combination of vectors in $\mathcal{B}$.

Here is an example which illustrates the point: Let $x(t)$ be any function on $[0, 1]$ which is not a polynomial — say $x(t) = \sin(2\pi t)$. Let $\mathcal{B} = \{1, t, t^2, t^3, \ldots\}$; the span (set of a finite linear combinations of elements) of $\mathcal{B}$ is all polynomials on $[0, 1]$. So $\boldsymbol{x} \notin \operatorname{span}(\mathcal{B})$. But $x(t)$ can be approximated arbitrarily well by elements in $\mathcal{B}$ (using higher and higher order polynomials) so $\boldsymbol{x} \in \operatorname{cl}\operatorname{Span}(\mathcal{B})$).

Examples.

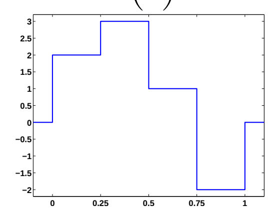
1. $\mathcal{S} = \mathbb{R}^2$, equipped with the standard inner product

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \qquad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$$

2. $\mathcal{S}$ = space of piecewise constant functions on $[0, 1/4),\ [1/4, 1/2),\ [1/2, 3/4),\ [3/4, 1]$

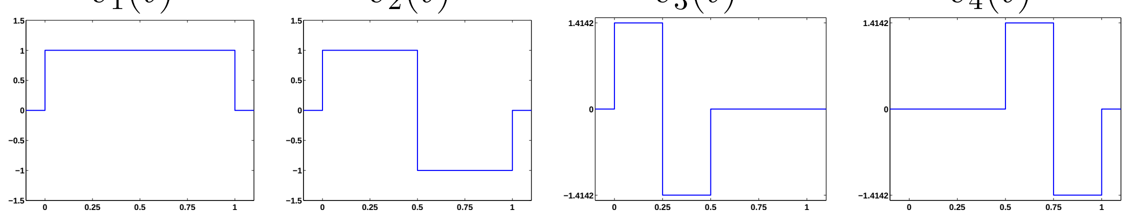
   Example signal:

$x(t)$



   The following four functions form an orthobasis for this space

$v_1(t)$ $v_2(t)$ $v_3(t)$ $v_4(t)$



3. **Fourier series**

$$\left\{ v_k(t) = \frac{1}{\sqrt{2\pi}} e^{jkt} \ , \ k \in \mathbb{Z} \right\} \quad \text{is an orthobasis for } L_2([0, 2\pi])$$

(with the standard inner product).

Let's quickly check the orthogonality:

$$\begin{aligned} \left\langle \frac{1}{\sqrt{2\pi}} e^{jk_1 t}, \frac{1}{\sqrt{2\pi}} e^{jk_2 t} \right\rangle &= \frac{1}{2\pi} \int_0^{2\pi} e^{j(k_1 - k_2)t} \ \mathrm{d}t \\ &= \begin{cases} 1, & k_1 = k_2 \\ 0, & k_1 \neq k_2 \end{cases}. \end{aligned}$$

It is also true that the closure of $\operatorname{span}(\{(2\pi)^{-1/2} e^{jkt}\}_{k=-\infty}^{\infty})$ is $L_2([0, 2\pi])$. The proof of this is a bit involved; if you are interested, see Chapter 5 of Young's *Introduction to Hilbert Space*.

4. **Sampling**

One of the fundamental results in information theory is that a signal that is *bandlimited* can be reconstructed exactly from uniformly spaced samples — this is known as the Shannon-Nyquist sampling theorem, and it has played a major role in the advances in our understanding of data acquisition, imaging, and digital communications.

To state this carefully, we need the notion of a **Fourier transform**. If $f(t)$ is a function of a continuous variable, its Fourier transform is

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\ \mathrm{d}t.$$

Roughly speaking, $\hat{f}(\omega)$ is a density that describes how the energy in $f(t)$ is spread over different frequencies. There is a unique mapping between $f(t)$ and $\hat{f}(\omega)$ — given the latter, we can recover $f(t)$ using

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)\, e^{j\omega t}\ \mathrm{d}\omega.$$

The mapping also preserves the standard inner product (to within a constant), in that

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle_{L_2} = \int_{-\infty}^{\infty} f(t)g(t)\ \mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)\overline{\hat{g}(\omega)}\,\mathrm{d}\omega = \frac{1}{2\pi}\langle \hat{\boldsymbol{f}}, \hat{\boldsymbol{g}}\rangle_{L_2}.$$

(This is known as the classical Parseval indentity.)

We say that a function $f(t)$ is **bandlimited** to Ω if

$$\hat{f}(\omega) = 0 \quad \text{for all} \quad |\omega| > \Omega.$$

Roughly speaking, this means that $f(t)$ contains no spectral content at frequencies greater than Ω.

Let $B_\Omega(\mathbb{R})$ = real-valued functions which are bandlimited to Ω, equipped with the standard inner product. The set of functions

$$\left\{ v_n(t) = \sqrt{T}\,\frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)},\ n\in\mathbb{Z} \right\},$$

with $T = \pi/\Omega$ is an orthobasis for $B_\Omega(\mathbb{R})$. It is (a sort of) easily checked fact that

$$\hat{v}_n(\omega) = \begin{cases} \sqrt{T}\,e^{-j\omega Tn} & |\omega| \leq \Omega \\ 0 & \text{otherwise.} \end{cases}$$

We check the orthogonality of the $\boldsymbol{v}_n$:

$$\begin{aligned}
&\left\langle \sqrt{T}\frac{\sin(\pi(t-n_1T)/T)}{\pi(t-n_1T)}\ ,\ \sqrt{T}\frac{\sin(\pi(t-n_2T)/T)}{\pi(t-n_2T)} \right\rangle \\
&= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} Te^{-j\omega Tn_1}e^{j\omega Tn_2}\ \mathrm{d}\omega \qquad \text{(Parseval)} \\
&= \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} e^{j\omega T(n_1-n_2)}\ \mathrm{d}\omega \\
&= \begin{cases} 1, & n_1 = n_2 \\ 0, & n_1 \neq n_2 \end{cases}.
\end{aligned}$$

That the (closure of the) span of this set is $B_\Omega(\mathbb{R})$ is mathematically equivalent to the Fourier series, where functions on an interval are being composed from harmonic sinusoids. In a few pages, we will show that the expansion coefficients in this basis are samples, thus giving us another way to reinterpret the Shannon-Nyquist sampling theorem.

5. **Legendre Polynomials** Define

$$p_0(t) = 1, \quad p_1(t) = t,$$

and then for $n = 1, 2, \ldots$

$$p_{n+1}(t) = \frac{2n+1}{n+1}\, t\, p_n(t) - \frac{n}{n+1} p_{n-1}(t),$$

and so

$$\begin{aligned} p_2(t) &= \frac{1}{2}(3t^2 - 1) \\ p_3(t) &= \frac{1}{2}(5t^3 - 3t) \\ p_4(t) &= \frac{1}{8}(35t^4 - 30t^2 + 3) \\ &\vdots \qquad \text{etc.} \end{aligned}$$

These $p_n(t)$ are called *Legendre polynomials*, and if we renormalize them, taking

$$v_n(t) = \sqrt{\frac{2n+1}{2}}\, p_n(t),$$

then $v_0(t), \ldots, v_N(t)$ are an orthobasis for polynomials of degree N on $[-1, 1]$.

Computing approximations with the Legendre basis is far more stable than computing the approximation in the standard basis.

Linear approximation and orthobases

Let's return to our linear approximation problem:
Given $\boldsymbol{x} \in \mathcal{S}$, we want to find the closest point in a subspace $\mathcal{T}$.

Suppose we have an orthobasis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ for $\mathcal{T}$. Then solving this problem is easy. Here's why: we know the solution is

$$\hat{\boldsymbol{x}} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_N\boldsymbol{v}_N \tag{1}$$

where the a_n are given by

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \boldsymbol{G}^{-1}\boldsymbol{b}, \quad \text{with } \boldsymbol{G} = \begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_2 \rangle \\ \vdots & & \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_N \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_N \rangle \end{bmatrix}, \ \boldsymbol{b} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}$$

Now since $\langle \boldsymbol{v}_n, \boldsymbol{v}_k \rangle = 1$ if $n = k$ and 0 otherwise, $\boldsymbol{G} = \mathbf{I}$ (the identity matrix), and so $\boldsymbol{G}^{-1} = \mathbf{I}$ as well, and

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}. \tag{2}$$

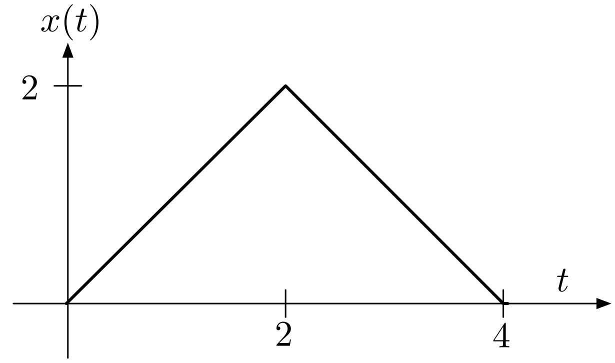
So calculating the closest point is as easy as computing N inner products — no matrix inversion necessary.

Combining the expressions (1) and (2) gives us the compact expression

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

Example. Suppose $x(t) \in L_2([0,4])$ is

$$x(t) = \begin{cases} t & 0 \leq t \leq 2 \\ 4-t & 2 \leq t \leq 4 \end{cases}$$



Let $\mathcal{T}$ = piecewise constant functions on $[0,1),\ [1,2),\ [2,3),\ [3,4]$.

Find the closest point in $\mathcal{T}$ to $\boldsymbol{x}$. A good orthobasis to use is

$$v_n(t) = \begin{cases} 1 & (n-1) \leq t \leq n \\ 0 & \text{otherwise} \end{cases}, \quad n = 1, 2, 3, 4.$$

Orthobasis expansions

The orthogonality principle (easily) gives us an expression for the **expansion coefficients** if a vector in an orthobasis.

Suppose a finite dimensional space $\mathcal{S}$ has an orthobasis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. Given any $\boldsymbol{x} \in \mathcal{S}$, the closest point in $\mathcal{S}$ to $\boldsymbol{x}$ is $\boldsymbol{x}$ itself (of course). This gives us the following **reproducing formula**:

$$\boldsymbol{x} = \sum_{n=1}^{N} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle \, \boldsymbol{v}_n, \quad \text{for all } \boldsymbol{x} \in \mathcal{S}.$$

In infinite dimensions, if $\mathcal{S}$ has an orthobasis $\{\boldsymbol{v}_n\}_{n=-\infty}^{\infty}$ and $\boldsymbol{x} \in \mathcal{S}$ obeys

$$\sum_{n=-\infty}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle|^2 \; < \; \infty,$$

then we can write

$$\boldsymbol{x} = \sum_{n=-\infty}^{\infty} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle \, \boldsymbol{v}_n.$$

(We need the sequence of expansion coefficients to be square-summable to make sure the sum of vectors above converges to something.)

In other words, $\boldsymbol{x} \in \mathcal{S}$ is captured without loss by the discrete list of numbers

$$\ldots, \; \langle \boldsymbol{x}, \boldsymbol{v}_{-1} \rangle, \; \langle \boldsymbol{x}, \boldsymbol{v}_0 \rangle, \; \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle, \; \ldots$$

An orthobasis gives us a natural way to discretize vectors in $\mathcal{S}$ through a set of expansion coefficients. Moreover, there is a straightforward and explicit way to compute these expansion coefficients — you simply take an inner product with the corresponding basis vector.

Example: Sampling a bandlimited function.

$B_{\pi/T}$ = space of bandlimited signals equipped with the standard inner product. We have seen already that

$$v_n(t) = \sqrt{T}\,\frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)}, \quad n \in \mathbb{Z}$$

is an orthobasis for $B_{\pi/T}$. This means that any $\boldsymbol{x} \in B_{\pi/T}$ can be written

$$\boldsymbol{x} = \sum_{n=-\infty}^{\infty} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle\, \boldsymbol{v}_n.$$

What are the $\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle$? As we saw above

$$\begin{aligned}
\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle &= \left\langle x(t)\ ,\ \sqrt{T}\,\frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)} \right\rangle \\
&= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \hat{x}(\omega)\,\sqrt{T}e^{jn\omega T}\ \mathrm{d}\omega \\
&= \sqrt{T}x(nT),
\end{aligned}$$

which is simply a sample scaled by $\sqrt{T}$. The reproducing formula in this case is a restatement of the Shannon-Nyquist sampling theorem:

$$\begin{aligned}
x(t) &= \sum_{n=-\infty}^{\infty} \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle\, \boldsymbol{v}_n \\
&= \sum_{n=-\infty}^{\infty} x(nT)\,\frac{T\sin(\pi(t-nT)/T)}{\pi(t-nT)},
\end{aligned}$$

meaning that knowing $x(t)$ only at the discrete set of points $\{nT,\ n \in \mathbb{Z}\}$ is enough to recover $x(t)$ for all $t \in \mathbb{R}$ using "sinc interpolation".

The moral of the story is that we can recreate a vector in a Hilbert space from the sequence of numbers $\{\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle\}$. We can think of every different orthobasis for $\mathcal{S}$ as a different **transform**, and the $\{\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle\}$ as **transform coefficients**.

Next we will see that our notions of **distance** and **angle** also carry over to this discrete space.