Prelude to orthobases: moving to infinite dimensions

The last set of notes gave us a computationally concrete technique for *approximating* an arbitrary function $\boldsymbol{f}$ from a Hilbert space using a linear combination of *fixed* elements $\boldsymbol{v}_n$ from the space. The best approximation (in terms of the norm induced by the inner product in the Hilbert space) to f is given by

$$\hat{f}(t) = \sum_{n=1}^{N} \hat{\alpha}_n v_n(t), \quad \hat{\boldsymbol{\alpha}} = \boldsymbol{G}^{-1}\boldsymbol{b}, \tag{1}$$

where

$$G_{i,j} = \langle \boldsymbol{v}_j, \boldsymbol{v}_i \rangle, \quad b_i = \langle \boldsymbol{f}, \boldsymbol{v}_i \rangle.$$

This is a well-posed problem in finite dimensions since we know that $\boldsymbol{G}$ is invertible when the $\{\boldsymbol{v}_n\}$ are linearly independent. If $\boldsymbol{f} \in \text{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\})$, then $\hat{\boldsymbol{f}} = \boldsymbol{f}$, and (1) can be viewed as a "reproducing formula" for $\boldsymbol{f}$.

What we would like to do is take this idea to infinite dimensions. That is, we would like an infinite sequence of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ so that for any fixed $\boldsymbol{f} \in \mathcal{H}$, when we create the approximation

$$\hat{\boldsymbol{f}}_N = \text{best approximation to } \boldsymbol{f} \text{ in } \text{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\})$$

using (1) above, $\hat{\boldsymbol{f}}_N$ gets closer and closer to $\boldsymbol{f}$ as N increases. That is, we would like

$$\lim_{N\to\infty} \|\hat{\boldsymbol{f}}_N - \boldsymbol{f}_N\| = 0,$$

or equivalently $\lim_{N\to\infty} \hat{\boldsymbol{f}}_N = \boldsymbol{f}$. As N increases, the space spanned by the $\{\boldsymbol{v}_n\}_{n=1}^N$ is becoming richer and richer, eventually encompass-

ing the entire Hilbert space. In the limit, we would like to write

$$f(t) = \sum_{n=1}^{\infty} \alpha_n \boldsymbol{v}_n(t),$$

i.e. represent $\boldsymbol{f}$ by the (infinite) list of numbers $\{\alpha_n\}$.

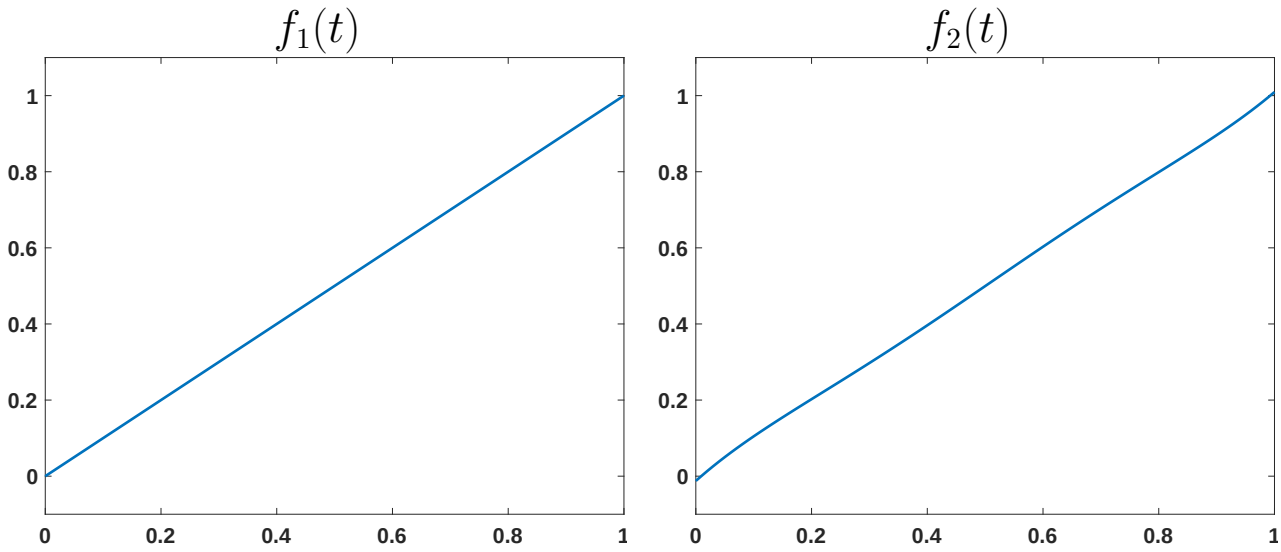
This process, however, is not always as straightforward as it sounds. Things can go wrong, even in familiar, well-defined cases.

The example below illustrates the problem clearly in one such familiar setting. On the left, we plot

$$f_1(t) = t,$$

while on the right, we plot

$$f_2(t) = 2.716\,t^5 - 6.898\,t^4 + 6.255\,t^3 - 2.4068\,t^2 + 1.356\,t - 0.013$$



Although they have very different expansion coefficients, these functions are almost exactly the same. In fact, the relative error in the

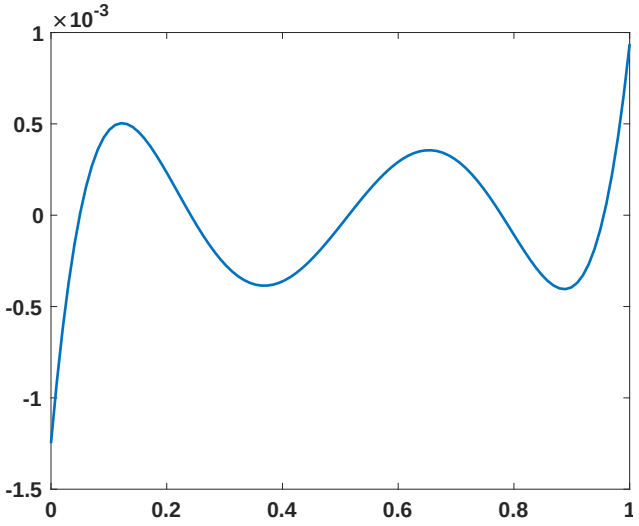
L_2 norm is

$$\frac{\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_2}{\|\boldsymbol{f}_1\|} \approx 3.25 \cdot 10^{-5}.$$

One conclusion we can draw from this is while the vectors $\{1, t, t^2, t^3, t^4, t^5\}$ are technically linearly independent, functionally they are not — we can build up very close approximations of members of that set through linear combinations of the other elements. Equivalently, there is a good 5th-order approximation of the $\mathbf{0}$ function on $[0, 1]$; we have

$$f_2(t) - f_1(t) = 2.716\,t^5 - 6.898\,t^4 + 6.255\,t^3 - 2.4068\,t^2 + 0.356\,t - 0.013,$$

which is plotted below,



Notice that the scale of this plot is 10^{-3}, i.e. three orders of magnitude below the plots above (which is why f_1 looks pretty similar to f_2 in those plots).

For polynomials, this effect becomes far more dramatic as we increase their order. While we were able to get within $1.5 \cdot 10^{-3}$ pointwise with a 5th order polynomial, we can get within $1.5 \cdot 10^{-6}$ with a 9th order polynomial, and within $6 \cdot 10^{-8}$ with a 11th order polynomial.

So while we might be able to say things like $f(t) \approx \sum_{n=0}^{N} \alpha_n t^n$ for large enough N, we do not really have confidence in the $\{\alpha_n\}$ as a representation for $\boldsymbol{f}$, as the mapping from the function $\boldsymbol{f}$ to the list of numbers $\{\alpha_n\}$ is unstable — very small changes in $\{\alpha_n\}$ lead to wildly different functions.

In the next section of these notes, we start to get a handle on this process. We start with the concept of an **orthobasis**, where the mapping from vector in a Hilbert space to sequence of coefficients is perfectly stable — changes to the sizes of the coefficients amount to a comparable sized change in reconstructed vector.

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. $\text{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}) = \mathcal{S}$,
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 $\text{span}(\mathcal{B})$ is $\mathcal{S}$; this is written

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

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 $\boldsymbol{x} \notin \text{span}(\mathcal{B})$. But $x(t)$ can be approximated arbitrarily well by elements in $\mathcal{B}$ (using higher and higher order polynomials) so $\boldsymbol{x} \in \text{cl Span}(\mathcal{B})$).

Examples.

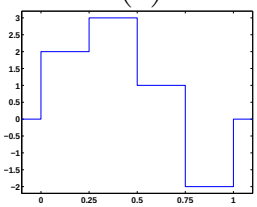
1. $\mathcal{S} = \mathbb{R}^2$, equipped with the standard inner product

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

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

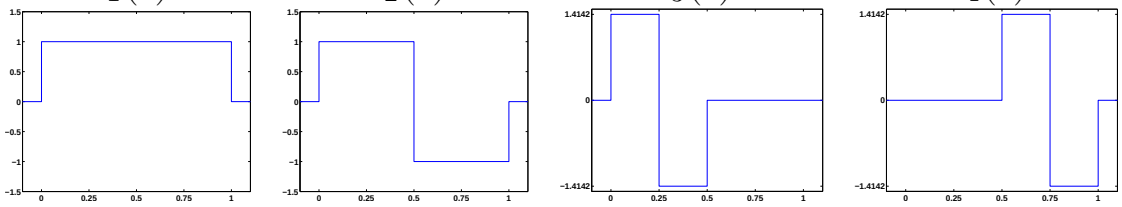
   Example signal:

$x(t)$



   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**

$$\left\{v_k(t) = e^{j2\pi kt} \ , \ k \in \mathbb{Z}\right\} \quad \text{is an orthobasis for } L_2([0,1])$$

(with the standard inner product).

Let's quickly check the orthogonality:

$$\begin{aligned}\langle \boldsymbol{v}_{k_1}, \boldsymbol{v}_{k_2} \rangle &= \int_0^1 e^{j2\pi(k_1-k_2)t} \, dt \\ &= \begin{cases} 1, & k_1 = k_2 \\ 0, & k_1 \neq k_2 \end{cases}.\end{aligned}$$

It is also true that the closure of span$(\{e^{j2\pi kt}\}_{k=-\infty}^{\infty})$ is $L_2([0,1])$. The proof of this is a bit involved; if you are interested, see Chapter 5 of Young's *Introduction to Hilbert Space*.

For real-valued functions, we can equivalently use

$$v_0(t) = 1, \quad v_k(t) = \begin{cases} \sqrt{2}\cos(\pi(k+1)t), & k \text{ odd} \\ \sqrt{2}\sin(\pi kt), & k \text{ even} \end{cases}$$

where we have introduced the factors of $\sqrt{2}$ to ensure that $\|\boldsymbol{v}_k\|_{L_2} = 1$.

4. **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

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

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.

5. Sampling (Additional reading)

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

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

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

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

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

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

(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

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

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

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

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.

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

$$\hat{\boldsymbol{x}} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_N \boldsymbol{v}_N \tag{2}$$

where the α_n are given by

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \boldsymbol{G}^{-1}\boldsymbol{b}, \quad \text{with } \boldsymbol{G} = \begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_2 \rangle \\ \vdots & & \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_N \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_N \rangle \end{bmatrix}, \; \boldsymbol{b} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}$$

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

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}. \tag{3}$$

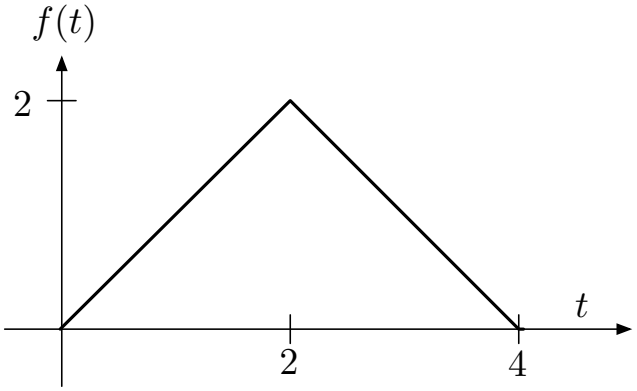
So calculating the closest point is as easy as computing N inner products — no matrix inversion necessary.

Combining the expressions (2) and (3) gives us the compact expression

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

Example. Suppose $\boldsymbol{f} \in L_2([0,4])$ is

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



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

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

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

Example Now let

$$f(t) = \begin{cases} t+1, & -1 \le t < 0, \\ 1-t, & 0 \le t \le 1. \end{cases}$$

What is the best 4th-order polynomial approximation of $\boldsymbol{f}$ on $[-1, 1]$?

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 will write the rigorous version of this shortly.)

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: Change of basis in $\mathbb{R}^2$.

Consider $\mathbb{R}^2$ equipped with the standard Euclidean inner product and the orthobasis

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

Let $\boldsymbol{x} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$. Find α_1, α_2 such that

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2.$$

Example: Fourier Series

We have seen that the functions

$$v_k(t) = e^{j2\pi kt},$$

form an orthobasis for $L_2([0,1])$. Given a function $\boldsymbol{f}$ on $[0,1]$, we also know that

$$f(t) = \sum_{k=-\infty}^{\infty} \alpha_k e^{j2\pi kt}, \quad \alpha_k = \int_0^1 f(t) e^{-j2\pi kt}\, dt.$$

The equation above is equivalent to

$$\boldsymbol{f} = \sum_{k=-\infty}^{\infty} \langle \boldsymbol{f}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k$$

where $\langle \cdot, \cdot \rangle$ is the L_2 inner product for functions on $[0,1]$.

For real-valued functions, we can equivalently take

$$f(t) = \sum_{k=0}^{\infty} \alpha_k v_k(t), \quad \alpha_k = \int_0^1 f(t) v_k(t)\, dt,$$

with

$$v_k(t) = \begin{cases} 1, & k = 0, \\ \sqrt{2}\cos(\pi(k+1)t), & k \text{ odd}, \\ \sqrt{2}\sin(\pi kt), & k \text{ even}. \end{cases}$$

Example: Change of basis for a polynomial

Let $f(t) = 5t^2 - 3t + 2$ for $t \in [0, 1]$. Find $\alpha_0, \alpha_1, \alpha_2$ such that

$$f(t) = \alpha_0 v_0(t) + \alpha_1 v_1(t) + \alpha_2 v_2(t),$$

where the $v_n(t)$ are the Legendre polynomials given earlier in these notes.

Example: Sampling a bandlimited function (additional reading).

$B_{\pi/T}$ = 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 $\boldsymbol{f} \in B_{\pi/T}$ can be written

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

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

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

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

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

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

The moral of the story is that we can recreate a vector $\boldsymbol{x}$ 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**. Let us put down the technical details under which this is true.

Technical details: Convergence of orthonormal sequences in Hilbert space

We now solidify the mathematics to make sense of what we have called the reproducing formula

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

There are an infinite number of terms in this sum, so we will have to take care in showing that the sum converges to something meaningful. Let $\boldsymbol{x}_N$ be the partial sum

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

We want to show that the sequence

$$\{\boldsymbol{x}_N\}_{N=1}^{\infty} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\}$$

converges to $\boldsymbol{x}$ under the norm induced by the inner product. Technically, this means that for every $\epsilon > 0$, there exists an m such that

$$\|\boldsymbol{x} - \boldsymbol{x}_N\| \leq \epsilon \quad \text{for all} \quad N \geq m.$$

We usually write this as

$$\lim_{N \to \infty} \|\boldsymbol{x} - \boldsymbol{x}_N\| = 0$$

or $\boldsymbol{x}_N \to \boldsymbol{x}$ as $N \to \infty$.

We establish this with a series of three propositions. The first two are technical, and are used as an end to prove the third.

Proposition 1: Inner product with a fixed vector is a continuous function. Let $\boldsymbol{c}$ be a vector in a Hilbert space $\mathcal{S}$, and let $\{\boldsymbol{x}_N\}$ be a sequence of vectors in $\mathcal{S}$ that converge to a point $\boldsymbol{x}$. Then

$$\lim_{N\to\infty} \langle \boldsymbol{x}_N, \boldsymbol{c}\rangle = \langle \boldsymbol{x}, \boldsymbol{c}\rangle.$$

Proof. This follows directly from Cauchy-Schwarz:

$$\begin{aligned} |\langle \boldsymbol{x}_N, \boldsymbol{c}\rangle - \langle \boldsymbol{x}, \boldsymbol{c}\rangle| &= |\langle \boldsymbol{x}_N - \boldsymbol{x}, \boldsymbol{c}\rangle| \\ &\leq \|\boldsymbol{x}_N - \boldsymbol{x}\|\,\|\boldsymbol{c}\|. \end{aligned}$$

Thus

$$\|\boldsymbol{x}_N - \boldsymbol{x}\| \leq \frac{\epsilon}{\|\boldsymbol{c}\|} \quad \Rightarrow \quad |\langle \boldsymbol{x}_N, \boldsymbol{c}\rangle - \langle \boldsymbol{x}, \boldsymbol{c}\rangle| \leq \epsilon,$$

and so $\langle \boldsymbol{x}_N, \boldsymbol{c}\rangle \to \langle \boldsymbol{x}, \boldsymbol{x}\rangle$ as $\boldsymbol{x}_N \to \boldsymbol{x}$.

■

Proposition 2: Bessel's inequality. Let $\{\boldsymbol{v}_n\}_{n=1}^{\infty}$ be a sequence of orthonormal vectors in a Hilbert space $\mathcal{S}$. Then for any $\boldsymbol{x} \in \mathcal{S}$,

$$\sum_{n=1}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_n\rangle|^2 \;\leq\; \|\boldsymbol{x}\|^2. \tag{4}$$

Proof. Fix N and let

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

We know that $\boldsymbol{y}_N$ is the closest point in Span$(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\})$ to $\boldsymbol{x}$, and that

$$\langle \boldsymbol{x} - \boldsymbol{y}_N, \boldsymbol{y}_N \rangle = 0.$$

Thus by the Pythagorean rule,

$$\|\boldsymbol{x}\|^2 = \|\boldsymbol{x} - \boldsymbol{y}_N\|_2 + \|\boldsymbol{y}_N\|^2.$$

Since $\|\boldsymbol{y}_N\|^2 = \sum_{n=1}^{N} |\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle|^2$, we have

$$\begin{aligned} \sum_{n=1}^{N} |\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle|^2 &= \|\boldsymbol{x}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}_N\|^2 \\ &\leq \|\boldsymbol{x}\|^2. \end{aligned}$$

Since the bound holds uniformly for all N, (4) follows.

■

Note that we are not requiring $\{\boldsymbol{v}_n\}_{n=1}^{\infty}$ to span the entire space — if it does span the space, it is an orthobasis, and the inequality in (4) becomes equality.

Proposition 3: The reproducing formula. Let $\{\boldsymbol{v}_n\}_{n=1}^{\infty}$ be a sequence of orthonormal vectors in a Hilbert space $\boldsymbol{\mathcal{S}}$, and let $\{\alpha_n\}$ be a sequence of scalars. Set

$$\boldsymbol{x}_N = \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n. \tag{5}$$

Then the sequence of vectors $\{\boldsymbol{x}_N\}$ converges to a vector $\boldsymbol{x} \in \boldsymbol{\mathcal{S}}$ as $N \to \infty$ if and only if

$$\sum_{n=1}^{\infty} |\alpha_n|^2 \; < \; \infty. \tag{6}$$

Moreover, the point $\boldsymbol{x}$ obeys

$$\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle = \alpha_n \quad \text{for all} \quad n = 1, 2, \ldots$$

Proof. By "converge" we mean that there is a vector $\boldsymbol{x} \in \mathcal{S}$ such that

$$\lim_{N \to \infty} \|\boldsymbol{x} - \boldsymbol{x}_N\|_S = 0.$$

First, suppose that (6) holds, and take $\boldsymbol{x}_N$ as in (5). Then for any $m \geq 1$,

$$\begin{aligned} \|\boldsymbol{x}_{N+m} - \boldsymbol{x}_N\|^2 &= \left\| \sum_{n=N+1}^{N+m} \alpha_n \boldsymbol{v}_n \right\|^2 \\ &= \sum_{n=N+1}^{N+m} |\alpha_n|^2, \end{aligned}$$

where the second step follows from the Pythagorean theorem and the fact that $\|\boldsymbol{v}_n\| = 1$. Since the $\{\alpha_n\}$ are square-summable, we have

$$\|\boldsymbol{x}_{N+m} - \boldsymbol{x}_N\|^2 \to 0 \quad \text{as} \quad N \to \infty.$$

Thus $\{\boldsymbol{x}_N\}_{N=1}^{\infty}$ is a Cauchy sequence, and since $\mathcal{S}$ is a Hilbert space, $\{\boldsymbol{x}_N\}$ converges.

Now suppose that the sequence $\{\boldsymbol{x}_N\}$ in (5) converges; that is, there exists a $\boldsymbol{x}$ such that

$$\lim_{N \to \infty} \|\boldsymbol{x} - \boldsymbol{x}_N\| = 0.$$

For a fixed $k \leq N$, we have

$$\begin{aligned} \langle \boldsymbol{x}_N, \boldsymbol{v}_k \rangle &= \sum_{n=1}^{N} \alpha_n \langle \boldsymbol{v}_n, \boldsymbol{v}_k \rangle \\ &= \alpha_k, \end{aligned}$$

since $\langle \boldsymbol{v}_n, \boldsymbol{v}_k \rangle = 1$ when $n = k$ and is zero otherwise. Since the above holds for all $N \geq k$, we have

$$\begin{aligned}
\alpha_k &= \lim_{N\to\infty} \langle \boldsymbol{x}_N, \boldsymbol{v}_k \rangle \\
&= \left\langle \lim_{N\to\infty} \boldsymbol{x}_N, \boldsymbol{v}_k \right\rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle.
\end{aligned}$$

(The first step above follows from the continuity on the inner product; see the lemma above.) Then applying the Bessel inequality,

$$\begin{aligned}
\sum_{n=1}^{\infty} |\alpha_n|^2 &= \sum_{n=1}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle|^2 \\
&\leq \|\boldsymbol{x}\|^2.
\end{aligned}$$

Since we started with the supposition that $\boldsymbol{x} \in \mathcal{S}$, we know $\|\boldsymbol{x}\| \leq \infty$.

A useful property: Parseval's Theorem

Once an orthobasis is introduced into any Hilbert space, it can be mapped to standard Euclidean space. As we will see, inner products between vectors become standard dot products between their expansion coefficients, and (induced) norms become standard sum-of-squares norms. The key identity is called Parseval's theorem[1], and it is relatively easy to establish.

Let $\mathcal{S}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_S$ which induces the norm $\| \cdot \|_S$. Let $\{v_k\}_{k \in \Gamma}$ be an orthobasis[2] for $\mathcal{S}$. Then for every $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_S = \sum_{k \in \Gamma} \alpha_k \overline{\beta_k},$$

where

$$\alpha_k = \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle_S, \qquad \beta_k = \langle \boldsymbol{y}, \boldsymbol{v}_k \rangle_S.$$

You can think of the $\{\alpha_k\}$ as the transform coefficients of $\boldsymbol{x}$ and the $\{\beta_k\}$ as the transform coefficients of $\boldsymbol{y}$. So we have

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle_S &= \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\ell_2}, \\ \|\boldsymbol{x}\|_S^2 &= \|\boldsymbol{\alpha}\|_2^2. \end{aligned}$$

$\Rightarrow$ An orthobasis makes every Hilbert space **equivalent** to ℓ_2.

[1] In its most common usage, the name Parseval is usually associated with the preservation of energy (to within a constant) between a function $f(t)$ and its Fourier transform, as we discussed in the last set of notes. Here, we will use this terminology more broadly.

[2] We are using Γ to be an arbitrary index set here; it can be either finite, e.g. $\Gamma = 1, 2, \ldots, N$, or infinite, e.g. $\Gamma = \mathbb{Z}$.

All of the geometry (lengths, angles) maps into standard Euclidean geometry in coefficient space. As you can imagine, this is a pretty useful fact.

Proof of Parseval. With $\alpha_k = \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle$ and $\beta_k = \langle \boldsymbol{y}, \boldsymbol{v}_k \rangle$, we can write

$$\boldsymbol{x} = \sum_{k \in \Gamma} \alpha_k \, \boldsymbol{v}_k, \quad \text{and} \quad \boldsymbol{y} = \sum_{k \in \Gamma} \beta_k \, \boldsymbol{v}_k,$$

and so

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle_S &= \left\langle \sum_{k \in \Gamma} \alpha_k \boldsymbol{v}_k, \sum_{\ell \in \Gamma} \beta_\ell \boldsymbol{v}_\ell \right\rangle_S \\ &= \sum_{k \in \Gamma} \alpha_k \left\langle \boldsymbol{v}_k, \sum_{\ell \in \Gamma} \beta_\ell \boldsymbol{v}_\ell \right\rangle_S \\ &= \sum_{k \in \Gamma} \sum_{\ell \in \Gamma} \alpha_k \overline{\beta_\ell} \langle \boldsymbol{v}_k, \boldsymbol{v}_\ell \rangle_S. \end{aligned}$$

For a fixed value of k, only one term in the inner sum above will be nonzero, as $\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell \rangle = 0$ unless $\ell = k$. Thus

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_S = \sum_{k \in \Gamma} \alpha_k \overline{\beta_k}.$$

A straightforward consequence of the result above is that distances in $\mathcal{S}$ under the induced norm are equivalent to Euclidean (ℓ_2) distances in coefficient space:

$$\|\boldsymbol{x} - \boldsymbol{y}\|_S = \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 = \left(\sum_{k \in \Gamma} (\alpha_k - \beta_k)^2 \right)^{1/2}.$$

Thus changing the value of an orthobasis expansion coefficient by an amount ϵ will change the signal by an amount (as measured in $\|\cdot\|_S$) ϵ.

To be more precise about this, suppose $\boldsymbol{x}$ has transform coefficients $\{\alpha_k = \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle_S\}$. If I perturb one of them, say at location k_0, by setting

$$\tilde{\alpha}_k = \begin{cases} \alpha_{k_0} + \epsilon & k = k_0 \\ \alpha_k & k \neq k_0 \end{cases},$$

and then synthesizing

$$\tilde{\boldsymbol{x}} = \sum_{k \in \Gamma} \tilde{\alpha}_k \boldsymbol{v}_k,$$

we will have

$$\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_S = \epsilon.$$

Notice that while the error is localized to one expansion coefficient, it could effect the entire reconstruction, but its net effect will still be ϵ.

More generally, suppose we add an error to every coefficient:

$$\tilde{\alpha}_k = \alpha_k + \epsilon_k,$$

then re-synthesize the vector

$$\tilde{\boldsymbol{x}} = \sum_{k \in \Gamma} \tilde{\alpha}_k \boldsymbol{v}_k.$$

The total error (as measured by the norm $\| \cdots \|_{\mathcal{S}}$) in the reconstructed vector will be the same as the total error (as measured by the Euclidean norm $\| \cdot \|_2$):

$$\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_{\mathcal{S}} = \|\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\|_2 = \left(\sum_{k \in \Gamma} |\epsilon_k|^2 \right)^{1/2}.$$

The upshot of this is that as we manipulate the expansion coefficients $\{\alpha_k\}$, we know what the net effect will be in the reconstructed function.

Truncating ortho expansions 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 $\{\boldsymbol{v}_k\}$:

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

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

$$\hat{\boldsymbol{x}} = \sum_{k=0}^{9} \langle \boldsymbol{x}, \boldsymbol{v}_k \rangle \boldsymbol{v}_k,$$

where $\langle \cdot, \cdot \rangle$ is the inner product for $\mathcal{S}$.

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

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

Since also

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

the 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

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

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 below.

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

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

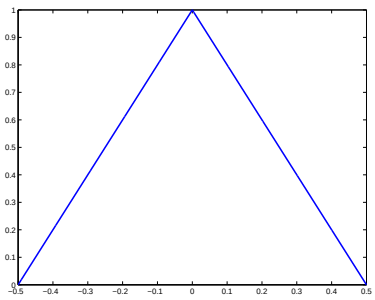
(That this is true follows directly from the observation that every signal on $[-1/2, 1/2]$ that is real-valued and even has a Fourier series which is real-valued and even.) This is an orthobasis expansion in the standard inner product with

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

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

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



the expansion coefficients are

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

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

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

and

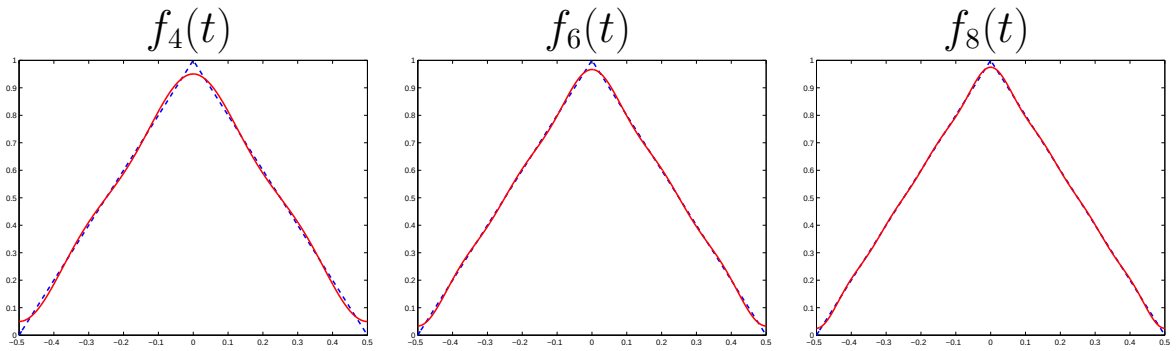
$$\begin{aligned}
\sum_{k=0}^{\infty} |\alpha_k|^2 &= \frac{1}{4} + \frac{8}{\pi^4}\sum_{k'=0}^{\infty}\frac{1}{(1+2k')^4} \\
&= \frac{1}{4} + \frac{8}{\pi^4}\left(\frac{\pi^4}{96}\right) \\
&= \frac{1}{3}.
\end{aligned}$$

When we truncate the expansion at K terms,

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

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

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



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

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

and so

$$\|\boldsymbol{f} - \boldsymbol{f}_K\|_2^2 = \sum_{k=K}^{\infty} |\alpha_k|^2,$$

or, since $\boldsymbol{f}_K \perp \boldsymbol{f} - \boldsymbol{f}_K$,

$$\|\boldsymbol{f} - \boldsymbol{f}_K\|_2^2 = \|\boldsymbol{f}\|_2^2 - \|\boldsymbol{f}_K\|_2^2.$$

In the three examples above, we have

$$\|\boldsymbol{f} - \boldsymbol{f}_4\|_2^2 \approx 1.92 \cdot 10^{-4}, \quad \|\boldsymbol{f} - \boldsymbol{f}_6\|_2^2 \approx 6.01 \cdot 10^{-5},$$
$$\|\boldsymbol{f} - \boldsymbol{f}_8\|_2^2 \approx 2.59 \cdot 10^{-5}.$$

Application: The DCT and JPEG

A great example of how manipulating orthobasis expansion coefficients can lead to something useful is given by the JPEG image compression standard. This is a complicated compression standard, but it is based on a simple idea: break the image into pieces, represent each piece using a cosine basis, then achieve compression by truncating/quantizing the expansion.

Lets start by recalling the discrete Fourier transform for vectors in $\mathbb{C}^N$. This is basically the same as Fourier series, except that the sinusoidal basis vectors are discrete (instead of functions of a continuous variable). Any $\boldsymbol{x} \in \mathbb{C}^N$ can be written as

$$\boldsymbol{x} = \sum_{k=0}^{N-1} \alpha_k \boldsymbol{\psi}_k, \quad \psi_k[n] = \frac{1}{\sqrt{N}} e^{j2\pi nk/N}, \quad n = 0, \ldots, N-1.$$

Computing $\boldsymbol{\alpha}$ for a given $\boldsymbol{x}$ can be done in $O(N \log_2 N)$ time using the *fast Fourier transform* (FFT), one of the most important algorithms in engineering and applied mathematics.

Closely related is the *discrete cosine transform* (DCT). This is a basis for $\mathbb{R}^N$, and has slightly more favorable symmetry properties than the standard DFT. The DCT basis functions for $\mathbb{R}^N$ are

$$\psi_k[n] = \begin{cases} \sqrt{\frac{1}{N}} & k = 0 \\ \sqrt{\frac{2}{N}} \cos\left(\frac{\pi k}{N}(n + \frac{1}{2})\right) & k = 1, \ldots, N-1 \end{cases} \tag{7}$$

for sample indices $n = 0, 1, \ldots, N-1$. Showing that

$$\sum_{n=0}^{N-1} \psi_k[n]\psi_\ell[n] = \begin{cases} 1 & k = \ell \\ 0 & k \neq \ell \end{cases}$$

is an exercise you can do at home. Notice that the samples of the cosines are on the half-sample points (we see $(n+1/2)$ in the expression above instead of n).

The DCT can be quickly computed from the DFT of a symmetric extension of $\boldsymbol{x}$. That means we have a **fast algorithm** for computing the DCT — the cost is essentially the same as for an FFT, $O(N \log N)$.

An *image* is an array of pixels arranged on a grid indexed by n_1, n_2. We can think of $x[n_1, n_2]$ as signifying the intensity of an image at location (n_1, n_2). If we want more than black and white images, we can use three intensity arrays (three "channels") to represent the color at every point.

Here, we will assume that these arrays are square, $N \times N$, although the discussion below is easily adapted for rectangular arrays. We will still call $\{x[n_1, n_2],\ 0 \leq n_1, n_2 \leq N-1\}$ a vector, even though it is naturally indexed by two variables . (It should be clear anyway that arrays of numbers like this obey our requirements for an abstract vector space.) We will call this space of 2D arrays $\mathbb{R}^N \times \mathbb{R}^N$.

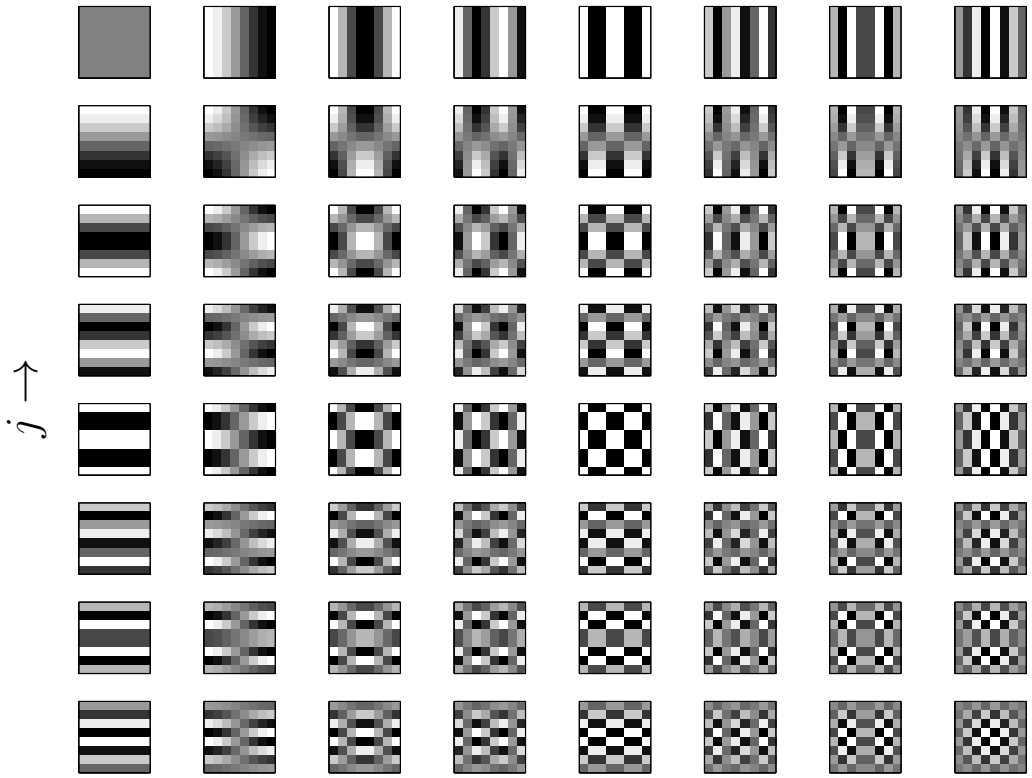
There is a straightforward way to create an orthobasis for $\mathbb{R}^N \times \mathbb{R}^N$ from an orthobasis for $\mathbb{R}^N$. We simply create *separable 2D arrays* from the 1D basis vectors. More precisely, if $\{\boldsymbol{\psi}_k,\ k = 0, \ldots, N-1\}$ is an orthobasis for $\mathbb{R}^N$, then $\{\boldsymbol{\psi}^{\text{2D}}_{j,k},\ 0 \leq j, k \leq N\}$ is an orthbobasis for $\mathbb{R}^N \times \mathbb{R}^N$, where

$$\psi^{\text{2D}}_{j,k}[n_1, n_2] = \psi_j[n_1]\psi_k[n_2].$$

You will prove this (and actually something much more general than

this) on next week's homework.

Let's take a look at the 2D DCT orthobasis for 8×8 image patches. The 64 2D DCT basis functions for $N = 8$ are shown below:



$\psi_{j,k}[m, n]$ for $j, k = 0, \ldots, 7$

2D DCT coefficients are indexed by two integers, and so are naturally arranged on a grid as well:

$$\begin{bmatrix} \alpha_{0,0} & \alpha_{0,1} & \cdots & \alpha_{0,N-1} \\ \alpha_{1,0} & \alpha_{1,1} & \cdots & \alpha_{1,N-1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{N-1,0} & \alpha_{N-1,1} & \cdots & \alpha_{N-1,N-1} \end{bmatrix}$$

The DCT in image and video compression

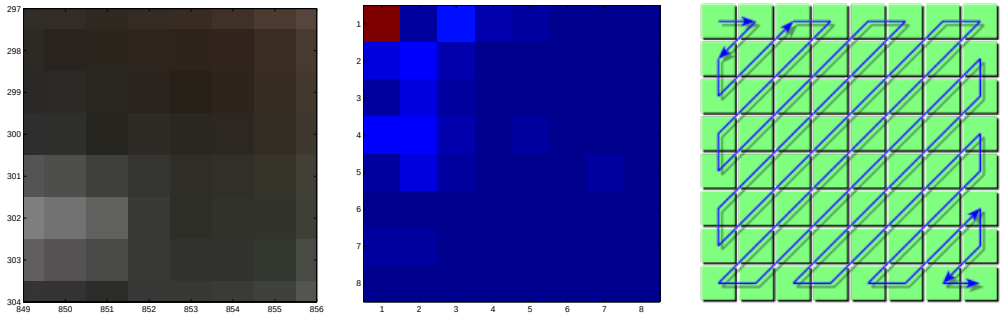
The DCT is basis of the popular JPEG image compression standard. The central idea is that while energy in a picture is distributed more or less evenly throughout, in the DCT transform domain it tends to be *concentrated* at low frequencies.

JPEG compression work roughly as follows:

1. Divide the image into 8×8 blocks of pixels
2. Take a DCT within each block
3. Quantize the coefficients — the rough effect of this is to keep the larger coefficients and remove the samller ones
4. Bitstream (losslessly) encode the result.

There are some details we are leaving out here, probably the most important of which is how the three different color bands are dealt with, but the above outlines the essential ideas.

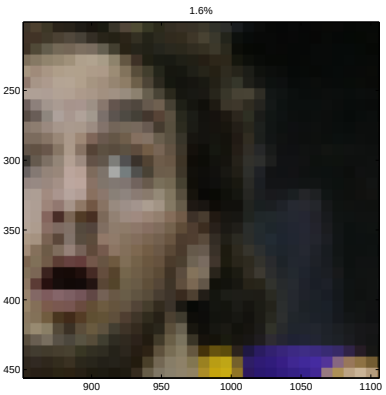
The basic idea is that while the energy within an 8×8 block of pixels tends to be more or less evenly distributed, the DCT concentrates this energy onto a relatively small number of transform coefficients. Moreover, the significant coefficients tend to be at the same place in the transform domain (low spatial frequencies).



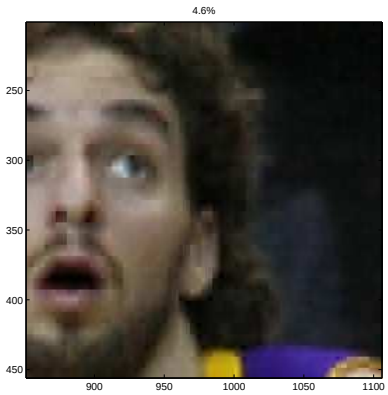
8×8 block 2D DCT coeffs ordering

To get a rough feel for how closely this model matches reality, let's look at a simple example. Here we have an original image 2048×2048, and a zoom into a 256×256 piece of the image:

Here is the same piece after using 1 of the 64 coefficients per block ($1/64 \approx 1.6\%$), $3/64 \approx 4.6\%$ of the coefficients, and $10/64 \approx 15.62\%$:



1/64



3/64

10/64

So the "low frequency" heuristic appears to be a good one.

JPEG does not just "keep or kill" coefficients in this manner, it quantizes them using a fixed quantization mask. Here is a common

example:

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}.$$

The quantization simply maps $\alpha_{j,k} \to \tilde{\alpha}_{j,k}$ using

$$\tilde{\alpha}_{j,k} = Q_{j,k} \cdot \text{round}\left(\frac{\alpha_{j,k}}{Q_{j,k}}\right)$$

You can see that the coefficients at low frequencies (upper left) are being treated much more gently than those at higher frequencies (lower right).

The **decoder** simply reconstructs each 8×8 block $\boldsymbol{x}_b$ using the synthesis formula

$$\tilde{\boldsymbol{x}}_b[m,n] = \sum_{k=0}^{7} \sum_{\ell=0}^{7} \tilde{\alpha}_{k,\ell}\, \phi_{k,\ell}[m,n]$$

By the Parseval theorem, we know exactly what the effect of quantizing each coefficient is going to be on the total error, as

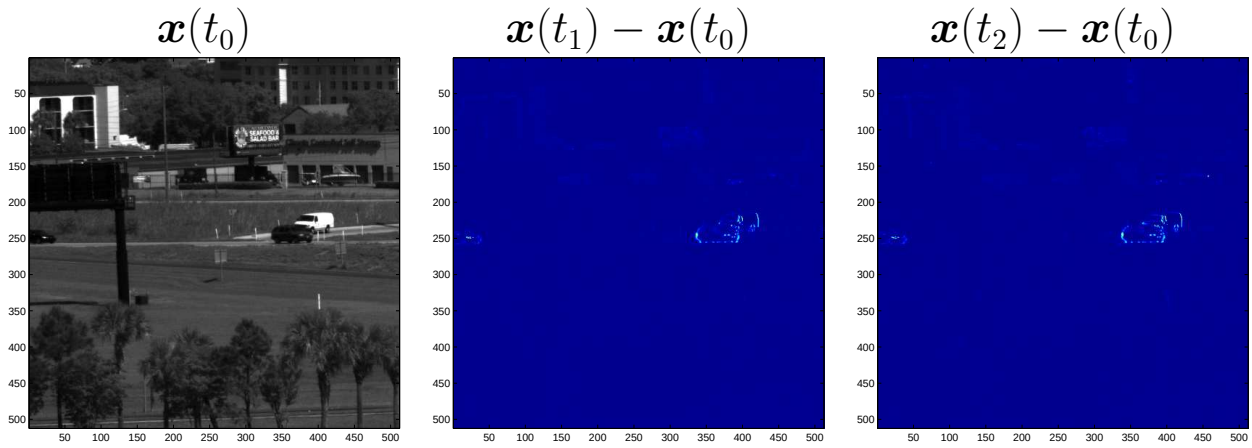
$$\|\boldsymbol{x}_b - \tilde{\boldsymbol{x}}_b\|_2^2 = \|\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\|_2^2 = \sum_{k=0}^{7} \sum_{\ell=0}^{7} |\alpha_{k,\ell} - \tilde{\alpha}_{k,\ell}|^2.$$

Video compression

The DCT also plays a fundamental role in video compression (e.g. MPEG, H.264, etc.), but in a slightly different way. Video codecs are complicated, but here is essentially what they do:

1. Estimate, describe, and quantize the motion in between frames.
2. Use the motion estimate to "predict" the next frame.
3. Use the (block-based) DCT to code the residual.

Here is an example video frame, along with the differences between this frame and the next two frames (in false color):



The only activity is where the car is moving from left to right.