Basis representation fundamentals

Having a basis representation for our signals of interest allows us to do two very nice things:

- Take the signal apart, writing it as a discrete linear combination of "atoms":

  $$x(t) = \sum_{\gamma\in\Gamma} \alpha(\gamma)\psi_\gamma(t)$$

  for some fixed set of *basis* signals $\{\psi_\gamma(t)\}_{\gamma\in\Gamma}$. Here Γ is a discrete index set (for example $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{Z}\times\mathbb{Z}$, $\mathbb{N}\times\mathbb{Z}$ etc.) which will be different depending on the application.

  Conceptually, we are breaking the signal up into manageable "chunks" that are either easier to compute with or have some semantic interpretation.

- Translate (linearly) the signal into into a discrete list of numbers in such a way that it can be reconstructed (i.e. the translation is lossless). Linear transform = series of inner products, so this mapping looks like:

  $$x(t) \longrightarrow \left\{\begin{array}{c} \langle x(t), \psi_1(t)\rangle \\ \langle x(t), \psi_2(t)\rangle \\ \vdots \\ \langle x(t), \psi_\gamma(t)\rangle \\ \vdots \end{array}\right\}$$

  for some fixed set of signals $\{\psi_\gamma(t)\}_{\gamma\in\Gamma}$.

  Having a discrete representation of the signal has a number of advantages, not the least of which is that they can be inputs to and outputs from digital computers.

Here are two very familiar examples:

1) Fourier series:
Let $x(t) \in L_2([0,1])$. Then we can build up $x(t)$ using harmonic complex sinusoids:

$$x(t) = \sum_{k \in \mathbb{Z}} \alpha(k)\, \mathrm{e}^{\mathrm{j}2\pi k t}$$

where

$$\begin{aligned}\alpha(k) &= \int_0^1 x(t)\, \mathrm{e}^{-\mathrm{j}2\pi k t}\, \mathrm{d}t \\ &= \langle x(t), \mathrm{e}^{\mathrm{j}2\pi k t}\rangle.\end{aligned}$$

Fourier series has two nice properties:

1. The $\{\alpha(k)\}$ carry semantic information about which frequencies are in the signal.
2. If $x(t)$ is smooth, the magnitudes $|\alpha(k)|$ fall off quickly as k increases. This energy compaction provides a kind of implicit *compression*.

If $x(t)$ is real, it might be sort of annoying that we are representing it using a list of complex numbers. An equivalent decomposition is

$$x(t) = \alpha(0)\psi_{0,0}(t) + \sum_{m \in \{0,1\}} \sum_{k=1}^{\infty} \alpha(m,k)\psi_{m,k}(t),$$

where $\alpha(m,k) = \langle x(t), \psi_{m,k}(t)\rangle$ with

$$\begin{aligned}\psi_{0,k}(t) &= \begin{cases} \mathbf{1} & k = 0 \\ \sqrt{2}\cos(2\pi k t) & k \geq 1 \end{cases} \\ \psi_{1,k}(t) &= \sqrt{2}\sin(2\pi k t).\end{aligned}$$

2) Sampling a bandlimited signal:

Suppose that $x(t)$ is *bandlimited* to $[-\pi/T, \pi/T]$:

$$\hat{x}(\omega) = \int x(t)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = 0 \quad \text{for} \quad |\omega| > \pi/T.$$

Then the Shannon-Nyquist sampling theorem tells us that we can reconstruct $x(t)$ from point samples that are equally spaced by T:

$$\begin{aligned} x[n] &= x(nT), \\ x(t) &= \sum_{n=-\infty}^{\infty} x[n]\,\frac{\sin(\pi(t-nT))}{\pi(t-nT)/T}. \end{aligned}$$

We can re-interpret this as a basis decomposition

$$x(t) = \sum_{n=\infty}^{\infty} \alpha(n)\,\psi_n(t)$$

with

$$\begin{aligned} \psi_n(t) &= \sqrt{T}\,\frac{\sin(\pi(t-nT))}{\pi(t-nT)} \\ \alpha(n) &= \sqrt{T}\,x(nT). \end{aligned}$$

If $x(t)$ is bandlimited, then the $\alpha(n)$ are also inner products against the $\psi_n(t)$:

$$\begin{aligned} \alpha(n) &= \sqrt{T}\,x(nT) \\ &= \frac{\sqrt{T}}{2\pi}\int_{-\pi/T}^{\pi/T} \hat{x}(\omega)\,\mathrm{e}^{\mathrm{j}\omega nT}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\langle \hat{x}(\omega), \hat{\psi}_n(\omega)\rangle, \end{aligned}$$

where

$$\hat{\psi}_n(\omega) = \begin{cases} \sqrt{T}\,e^{-\mathrm{j}\omega nT} & |\omega| \le \pi/T \\ 0 & |\omega| > \pi/T. \end{cases}$$

Then by the classical Parseval theorem for Fourier transforms:

$$\alpha(n) = \langle x(t), \psi_n(t) \rangle,$$

where

$$\begin{aligned}
\psi_n(t) &= \frac{\sqrt{T}}{2\pi} \int_{-\pi/T}^{\pi/T} \mathrm{e}^{-\mathrm{j}\omega n T} \mathrm{e}^{\mathrm{j}\omega t} \, \mathrm{d}\omega \\
&= \frac{\sqrt{T}}{2\pi} \int_{-\pi/T}^{\pi/T} \mathrm{e}^{\mathrm{j}\omega (t - nT)} \, \mathrm{d}\omega \\
&= \sqrt{T} \cdot \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)}.
\end{aligned}$$

Thus we can interpret the Shannon-Nyquist sampling theorem as an expansion of a bandlimited signal in an basis of shifted sinc functions. We offer two additional notes about this result:

- Sampling a signal is a fundamental operation in applications. Analog-to-digital converters (ADCs) are prevalent and relatively cheap — ADCs operating at 10s of MHz cost on the order of a few dollars/euros.
- The sinc representation for bandlimited signals is mathematically the same as the Fourier series for signals with finite support, just with the roles of time and frequency reversed.

Orthobasis expansions

Fourier series and the sampling theorem are both examples of expansions in an *orthonormal basis* ("orthobasis expansion" for short). The set of signals $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is an orthobasis for a space H if

1. $$\langle \psi_\gamma, \psi_{\gamma'} \rangle = \begin{cases} 1 & \gamma = \gamma' \\ 0 & \gamma \neq \gamma' \end{cases}.$$

2. $\mathrm{span}\{\psi_\gamma\}_{\gamma\in\Gamma} = H$. That is, there is no $x \in H$ such that $\langle \psi_\gamma, x \rangle = 0$ for all $\gamma \in \Gamma$. (In infinite dimensions, this should technically read the *closure* of the span).

If $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is an orthobasis for H, then every $x(t) \in H$ can be written as

$$x(t) = \sum_{\gamma\in\Gamma} \langle x(t), \psi_\gamma(t) \rangle \, \psi_\gamma(t).$$

This is called the **reproducing formula**.

Orthobases are nice since they not only allow every signal to be decomposed as a linear combination of elements, but we have a simple and explicit way of computing the coefficients (the $\alpha(\gamma) = \langle x, \psi_\gamma \rangle$) in this expansion.

Associated with an orthobasis $\{\psi_\gamma\}_{\gamma\in\Gamma}$ for a space H are two linear operators. The first operator $\Psi^* : H \to \ell_2(\Gamma)$ maps the signal $x(t)$ in H to the sequence of expansion coefficients in $\ell_2(\Gamma)$ (of course, if H is finite dimensional, it may be more appropriate to write the range of this mapping as $\mathbb{R}^N$ rather than $\ell_2(\Gamma)$). The mapping Ψ is called the **analysis operator**, and its action is given by

$$\Psi^*[x(t)] \;=\; \{\langle x(t), \psi_\gamma(t) \rangle\}_{\gamma\in\Gamma} = \{\alpha(\gamma)\}_{\gamma\in\Gamma}.$$

The second operator $\Psi : \ell_2(\Gamma) \to H$ takes a sequence of coefficients in $\ell_2(\Gamma)$ and uses them to build up a signal. The mapping Ψ is called the **synthesis operator**, and its action is given by

$$\Psi[\{\alpha(\gamma)\}_{\gamma\in\Gamma}] \;=\; \sum_{\gamma\in\Gamma} \alpha(\gamma)\,\psi_\gamma(t).$$

Formally, Ψ and Ψ^* are *adjoint operators* — in finite dimensions, they can be represented as matrices where the basis functions $\psi_\gamma(t)$ are the columns of Ψ and rows of Ψ^*.

The generalized Parseval theorem

The (generalized) Parseval theorem says that the mapping from a signal $x(t)$ to its basis coefficients **preserves inner products** (and hence energy). If $x(t)$ is a continuous-time signal, then the relation is between two different types of inner products, one continuous and one discrete. Here is the precise statement:

Theorem. Let $\{\psi_\gamma\}_{\gamma\in\Gamma}$ be an orthobasis for a space H. Then for any two signals $x, \in H$

$$\langle x, y\rangle_H = \sum_{\gamma\in\Gamma} \alpha(\gamma)\beta(\gamma)^*$$

where

$$\alpha(\gamma) \;=\; \langle x, \psi_\gamma\rangle_H \quad \text{and} \quad \beta(\gamma) \;=\; \langle y, \psi_\gamma\rangle_H.$$

Proof.

$$\begin{aligned}
\langle x, y\rangle_H &= \left\langle \sum_\gamma \alpha(\gamma)\psi_\gamma, \sum_{\gamma'} \beta(\gamma')\psi_{\gamma'} \right\rangle_H \\
&= \sum_\gamma \sum_{\gamma'} \alpha(\gamma)\beta(\gamma')^* \langle \psi_\gamma, \psi_{\gamma'}\rangle_H \\
&= \sum_\gamma \alpha(\gamma)\beta(\gamma)^*,
\end{aligned}$$

since $\langle \psi_\gamma, \psi_{\gamma'}\rangle_H = 0$ unless $\gamma = \gamma'$, in which case $\langle \psi_\gamma, \psi_{\gamma'}\rangle_H = 1$.

Of course, this also means that the energy in the original signal is preserved in its coefficients. For example, if $x(t) \in L_2(\mathbb{R})$ is a continuous-time signal and $\alpha_\gamma = \langle x, \psi_\gamma\rangle$, then

$$\begin{aligned}
\|x(t)\|^2_{L_2(\mathbb{R})} = \int |x(t)|^2 \,\mathrm{d}t = \int x(t)x(t)^*\mathrm{d}t &= \sum_{\gamma\in\Gamma} \alpha(\gamma)\alpha(\gamma)^* = \sum_{\gamma\in\Gamma} |\alpha(\gamma)|^2 \\
&= \|\alpha\|^2_{\ell_2(\Gamma)}.
\end{aligned}$$

Everything is discrete

An amazing consequence of the Parseval theorem is that every space of signals for which we can find any orthobasis can be discretized. That the mapping from (continuous) signal space into (discrete) coefficient space preserves inner products essentially means that it preserves all of the geometrical relationships between the signals (i.e. distances and angles). In some sense, this means that all signal processing can be done by manipulating discrete sequences of numbers.

For our primary continuous spaces of interest, $L_2(\mathbb{R})$ and $L_2([0,1])$ which are equipped with the standard inner product, there are many orthobases from which to choose, and so many ways in which we can "sample" the signal to make it discrete.

Here is an example of the power of the Parseval theorem. Suppose that I have samples $\{x[n] = x(nT)\}_n$ of a bandlimited signal $x(t)$. Suppose one of the samples is perturbed by a known amount ϵ, forming

$$\tilde{x}[n] = \begin{cases} x[n] + \epsilon & n = n_0 \\ x[n] & \text{otherwise} \end{cases}.$$

What is the effect on the reconstructed signal? That is, if

$$\tilde{x}(t) = \sum_{n \in \mathbb{Z}} \tilde{x}[n] \, \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T}$$

what is the energy in the error

$$\|x - \tilde{x}\|_{L_2}^2 = \int |x(t) - \tilde{x}(t)|^2 \, \mathrm{d}t \quad ?$$

Projections and the closest point problem

A fundamental problem is to find the closest point in a fixed subspace to a given signal. If we have an orthobasis for this subspace, this problem is easy to solve.

Formally, let $\psi_1(t), \ldots, \psi_N(t)$ be a finite set of orthogonal vectors in H, and set

$$\mathcal{V} = \operatorname{span}\{\psi_1, \ldots, \psi_N\}.$$

Given a fixed signal $x_0(t) \in H$, the solution $\tilde{x}_0(t)$ to

$$\min_{x \in \mathcal{V}} \ \|x_0(t) - x(t)\|_2^2 \tag{1}$$

is given by

$$\tilde{x}_0(t) = \sum_{k=1}^{N} \langle x_0(t), \psi_k(t)\rangle \psi_k(t).$$

We will prove this statement a little later.

The result can be extended to infinite dimensional subspaces as well. If $\{\psi_k(t)\}_{k\in\mathbb{Z}}$ is a set of (not necessarily complete) orthogonal signals in H, and we let $\mathcal{V}$ be the closure of the span of $\{\psi_k\}_{k\in\mathbb{Z}}$, then the solution to (1) is simply

$$\tilde{x}_0(t) = \sum_{k\in\mathbb{Z}} \langle x_0(t), \psi_k(t)\rangle \psi_k(t).$$

Example: Let $x(t) \in L_2(\mathbb{R})$ be an arbitrary continuous-time signal. What is the closest bandlimited signal to $x(t)$?

The solution of (1) is called the *projection* of x_0 onto $\mathcal{V}$. There is a linear relationship between a point $x_0 \in H$ and the corresponding closest point $\tilde{x}_0 \in \mathcal{V}$. If Ψ^* is the (linear) mapping

$$\Psi^*[x_0] = \{\langle x_0, \psi_k \rangle\}_k,$$

and Ψ is the corresponding adjoint, then $\tilde{x}_0$ can be compactly written as

$$\tilde{x}_0 = \Psi[\Psi^*[x_0]].$$

We can define the linear operator $P_{\mathcal{V}}$ that maps x_0 to its closest point as

$$P_{\mathcal{V}} = \Psi^*\Psi.$$

It is easy to check that $\Psi^*[\Psi[\{\alpha(k)\}_k]] = \{\alpha(k)\}_k$ for any set of coefficients $\{\alpha(k)\}_k$, and so

$$P_{\mathcal{V}} P_{\mathcal{V}} = P_{\mathcal{V}}.$$

It is also easy to see that $P_{\mathcal{V}}$ is self-adjoint:

$$P_{\mathcal{V}}^* = P_{\mathcal{V}}.$$

Cosine transforms

The cosine-I transform is an alternative to Fourier series; it is an expansion in an orthobasis for functions on $[0, 1]$ (or any interval on the real line) where the basis functions look like sinusoids. There are two main differences that make it more attractive than Fourier series for certain applications:

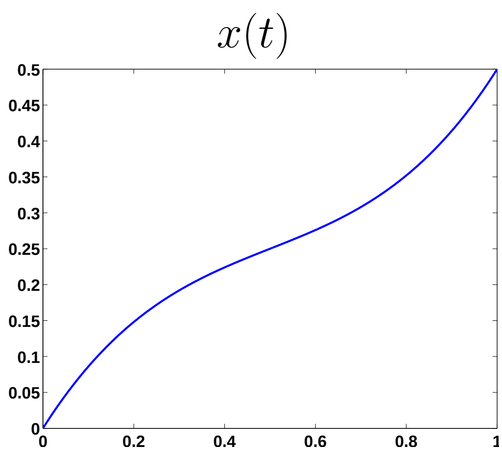
1. the basis functions and the expansion coefficients are real-valued;
2. the basis functions have different symmetries.

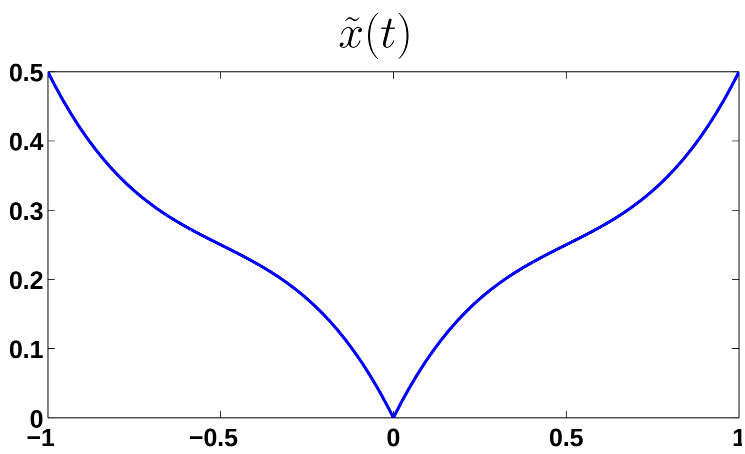
Definition. The cosine-I basis functions for $t \in [0, 1]$ are

$$\psi_k(t) = \begin{cases} 1 & k = 0 \\ \sqrt{2}\cos(\pi k t) & k > 0 \end{cases}. \tag{2}$$

We can derive the cosine-I basis from the Fourier series in the following manner. Let $x(t)$ be a signal on the interval $[0, 1]$. Let $\tilde{x}(t)$ be its "reflection extension" on $[-1, 1]$. That is

$$\tilde{x}(t) = \begin{cases} x(-t) & -1 \le t \le 0 \\ x(t) & 0 \le t \le 1 \end{cases}$$





We can use Fourier series to synthesis $\tilde{x}(t)$:

$$\tilde{x}(t) = \sum_{k=-\infty}^{\infty} \alpha_k \, \mathrm{e}^{\mathrm{j}\pi k t}.$$

Since $\tilde{x}(t)$ is real, we will have $\alpha_{-k} = \overline{\alpha_k}$, and so we can rewrite this as

$$\tilde{x}(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(\pi k t) + \sum_{k=1}^{\infty} b_k \sin(\pi k t),$$

where $a_0 = \alpha_0$, $a_k = 2\,\mathrm{Re}\,\{\alpha_k\}$, and $b_k = -2\,\mathrm{Im}\,\{\alpha_k\}$. Since $\tilde{x}(t)$ is even and $\sin(\pi k t)$ is odd, $\langle \tilde{x}(t), \sin(\pi k t)\rangle = 0$ and so

$$b_k = 0, \quad \text{for all } k = 1, 2, 3, \ldots,$$

and so $\tilde{x}(t)$ on $[-1, 1]$ can be written as

$$\tilde{x}(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(\pi k t).$$

Since we can use this expansion to build up **any** symmetric function on $[-1, 1]$, it means that the right hand side of the function on $[0, 1]$ is arbitrary, so **any** $x(t)$ on $[0, 1]$ can be written as

$$x(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(\pi k t).$$

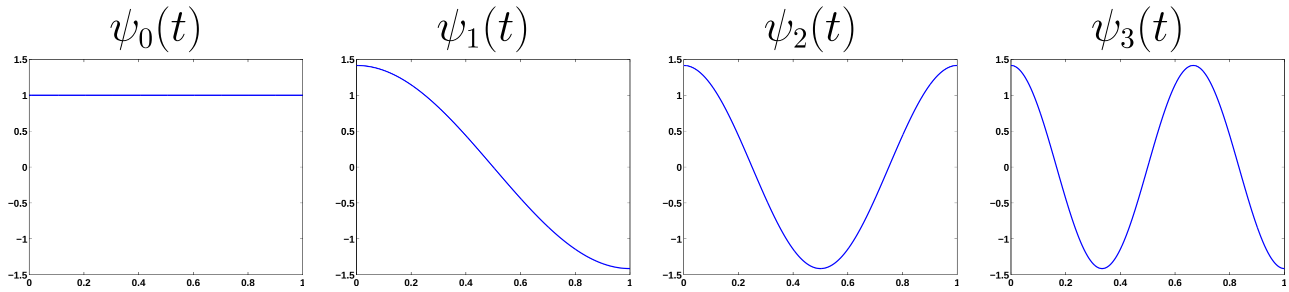
All that remains to show that $\{\psi_k : k = 1, 2, \ldots\}$ is an orthobasis is

$$\langle \psi_k, \psi_\ell \rangle = 2 \int_0^1 \cos(\pi k t) \cos(\pi \ell t) \mathrm{d}t = \begin{cases} 1 & k = \ell \\ 0 & k \neq \ell \end{cases}.$$

I will let you do this at home.

One way to think about the cosine-I expansion is that we are taking an **oversampled** Fourier series, with frequencies spaced at multiples of π rather than 2π, but then only using the real part.

Here are the first four cosine-I basis functions:



The Discrete Cosine Transform (DCT)

Just like there is a discrete Fourier transform, there is also a discrete cosine transform:
Definition: The DCT basis functions for $\mathbb{R}^N$ are

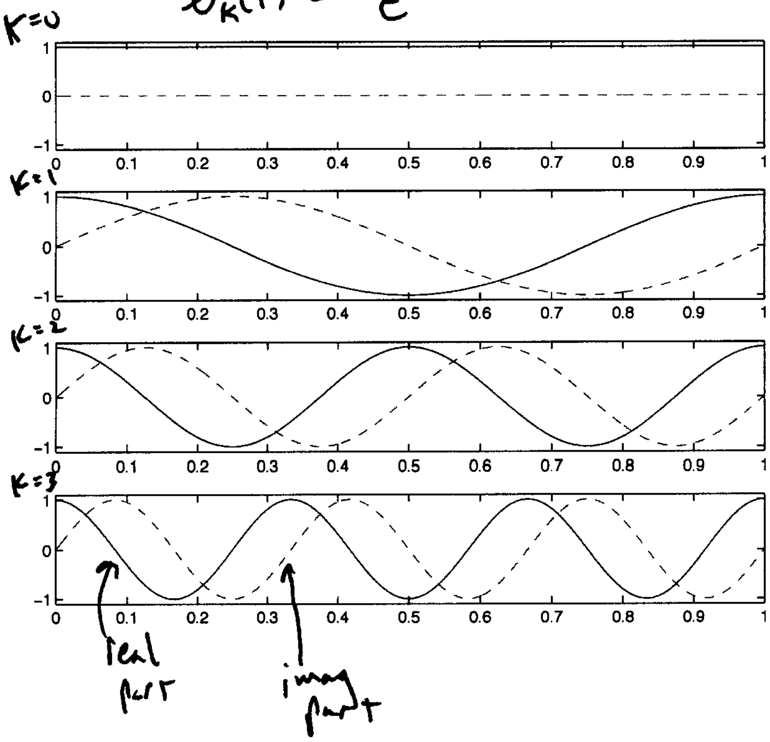
$$\psi_k[n] = \begin{cases} \sqrt{\frac{1}{N}} & k = 0 \\ \sqrt{\frac{2}{N}} \cos\left(\frac{\pi k}{N}\left(n + \frac{1}{2}\right)\right) & k = 1, \ldots, N-1 \end{cases}, \quad n = 0, 1, \ldots, N-1. \tag{3}$$

The cosine-I transform has "even" symmetry at both endpoints. There is a variation on this, called the cosine-IV transform, that has even symmetry at one endpoint and odd symmetry at the other:
Definition. The cosine-IV basis functions for $t \in [0, 1]$ are

$$\psi_k(t) = \sqrt{2}\cos\left(\left(k + \frac{1}{2}\right)\pi t\right), \qquad k = 0, 1, 2, \ldots. \tag{4}$$

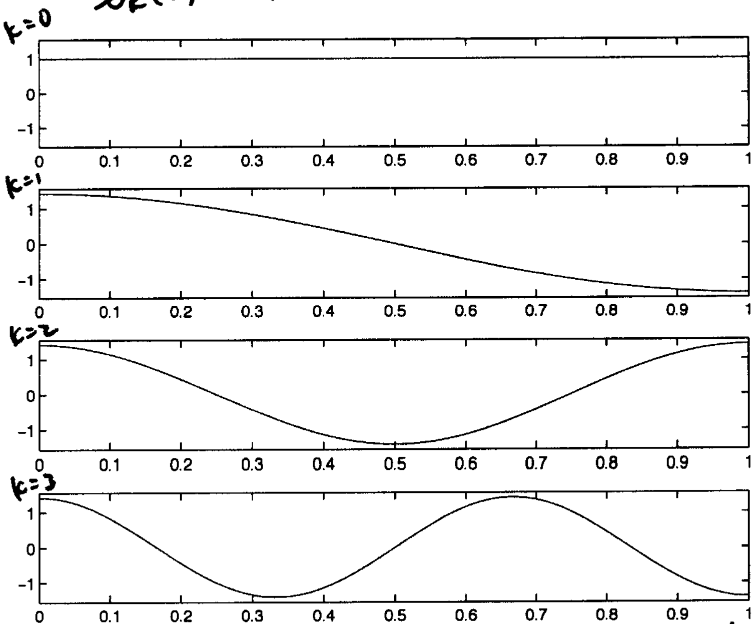
Fourier Series

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



Cosine-I

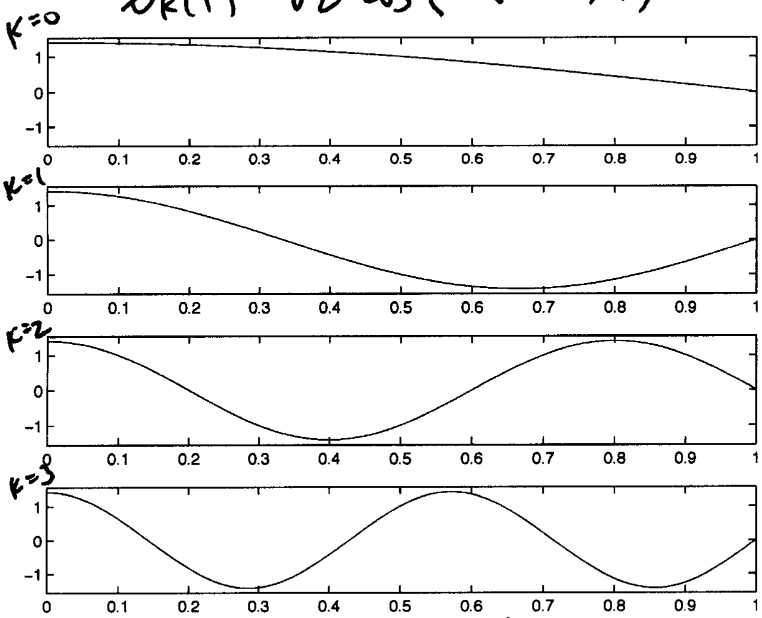
$\phi_k(t) = \lambda_k \sqrt{2} \cos(\pi k t)$



Note: even symmetry, real valued
↳ around 0 & 1
cycles = $k/2$

Cosine-IV

$\phi_k(t) = \sqrt{2} \cos(\pi(k+1/2)t)$



Note: even sym around 0
odd sym around 1
cycles = $\frac{k+1/2}{2}$

```
untitled text                                        Page 1 of 1
Printed: 1/28/09 10:03:07 AM              Printed For: Justin Romberg

t = linspace(0,1,500);

% dft
figure(1)
clf
for kk = 0:3
  subplot(4,1,kk+1)
  plot(t, cos(2*pi*t*kk), 'b-', t, sin(2*pi*t*kk), 'r--');
  axis([0 1 -1.1 1.1])
end

% dct-i
figure(2)
clf
for kk = 0:3
  subplot(4,1,kk+1)
  if (kk==0), lk = 1/sqrt(2); else lk = 1; end
  plot(t, lk*sqrt(2)*cos(pi*kk*t), 'b-');
  axis([0 1 -1.1*sqrt(2) 1.1*sqrt(2)])
end

% dct-iv
figure(3)
clf
for kk = 0:3
  subplot(4,1,kk+1)
  plot(t, sqrt(2)*cos(pi*(kk+1/2)*t), 'b-');
  axis([0 1 -1.1*sqrt(2) 1.1*sqrt(2)])
end
```

code to produce plots

The cosine-I and DCT for 2D images

Just as for Fourier series and the discrete Fourier transform, we can leverage the 1D cosine-I basis and the DCT into separable bases for 2D images.

Definition. Let $\{\psi_k(t)\}_{k\geq 0}$ be the cosine-I basis in (2). Set

$$\psi^{\text{2D}}_{k_1,k_2}(s,t) = \psi_{k_1}(s)\psi_{k_2}(t).$$

Then $\{\psi^{\text{2D}}_{k_1,k_2}(s,t)\}_{k_1,k_2\in\mathbb{N}}$ is an orthonormal basis for $L_2([0,1]^2)$

This is just a particular instance of a general fact. It is straightforward to argue (you can do so at home) that if $\{\psi_\gamma(t)\}_{\gamma\in\Gamma}$ is an orthonormal basis for $L_2([0,1])$, then $\{\psi_{\gamma_1}(s)\psi_{\gamma_2}(t)\}_{\gamma_1,\gamma_2\in\Gamma}$ is an orthonormal basis for $L_2([0,1]^2)$.

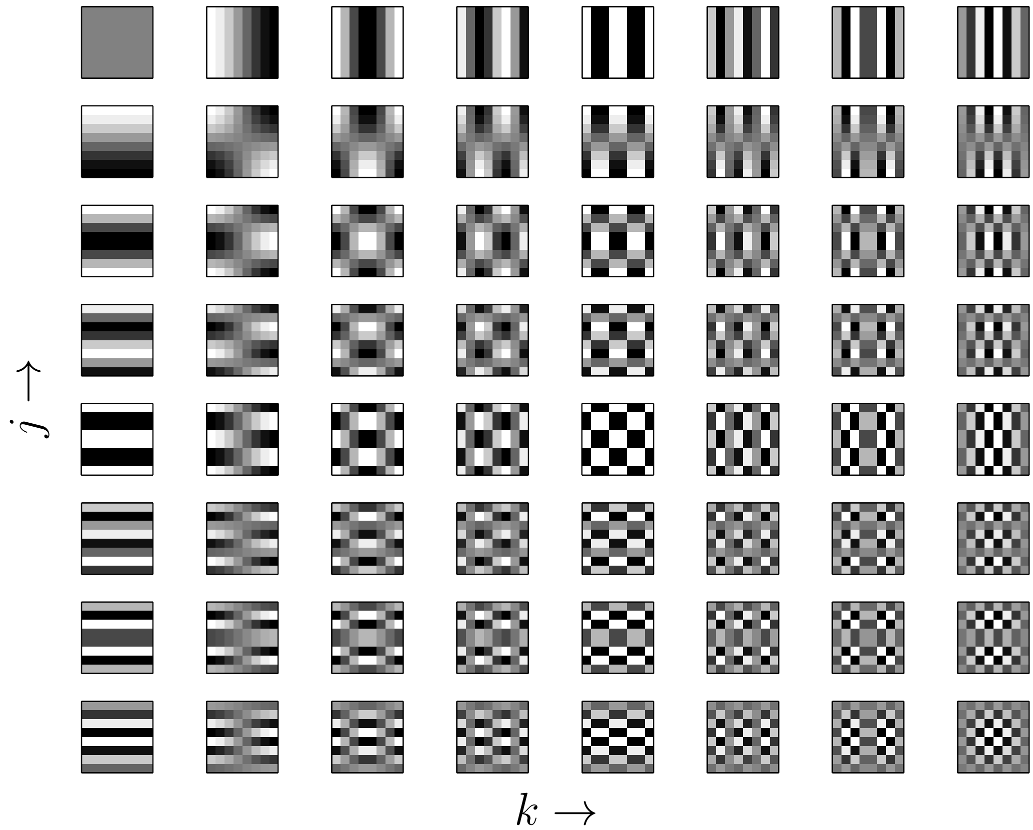
The DCT extends to 2D in the same way.

Definition. Let $\{\psi_k[n]\}_{0\leq k\leq N-1}$ be the DCT basis in (3). Set

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

Then $\{\psi^{\text{2D}}_{j,k}[m,n]\}_{0\leq j,k\leq N-1}$ is an orthonormal basis for $\mathbb{R}^N\times\mathbb{R}^N$.

The 64 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:

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

The DCT in image and video compression

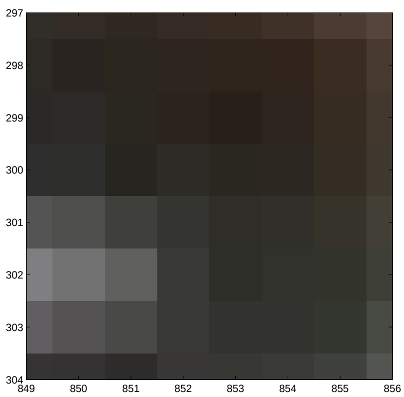
The DCT is basis of the popular JPEG image compression standard. The central idea is that while energy in a picture is distributed more or less evenly throughout, in the DCT transform domain it tends to be *concentrated* at low frequencies.

JPEG compression work roughly as follows:

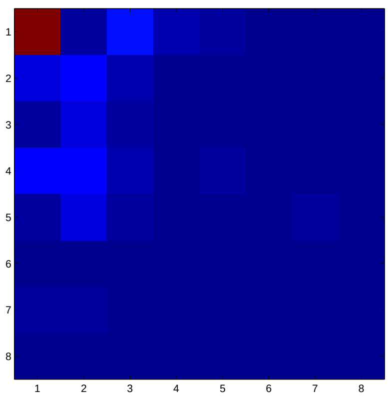
1. Divide the image into 8×8 blocks of pixels
2. Take a DCT within each block
3. Quantize the coefficients — the rough effect of this is to keep the larger coefficients and remove the samller ones
4. Bitstream (losslessly) encode the result.

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.

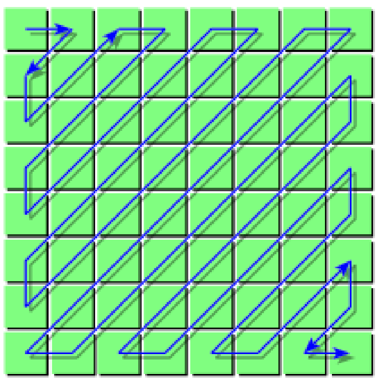
The basic idea is that while the energy within an 8×8 block of pixels tends to be more or less evenly distributed, the DCT concentrates this energy onto a relatively small number of transform coefficients. Moreover, the significant coefficients tend to be at the same place in the transform domain (low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



8×8 block



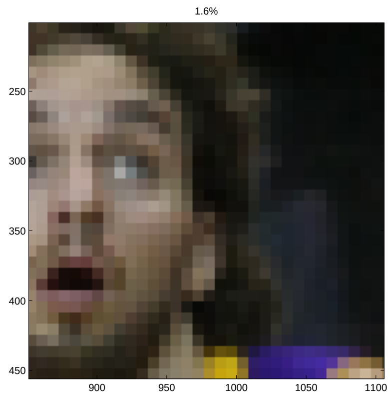
2D DCT coeffs



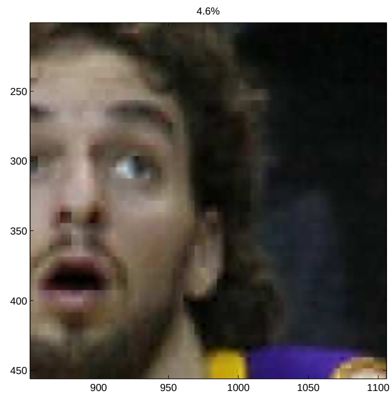
ordering

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:

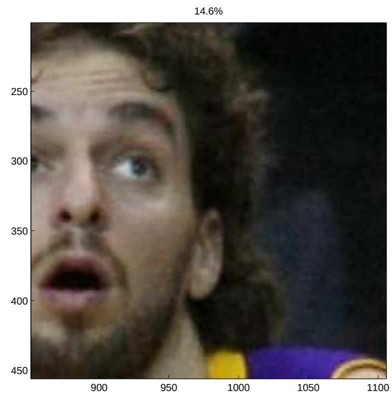
Here is the same piece after using 1 of the 64 coefficients per block ($1/64 \approx 1.6\%$), $3/64 \approx 4.6\%$ of the coefficients, and $10/64 \approx 15/62\%$:



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 $\alpha_{j,k} \rightarrow \tilde{\alpha}_{j,k}$ 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 x_b using the synthesis formula

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

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

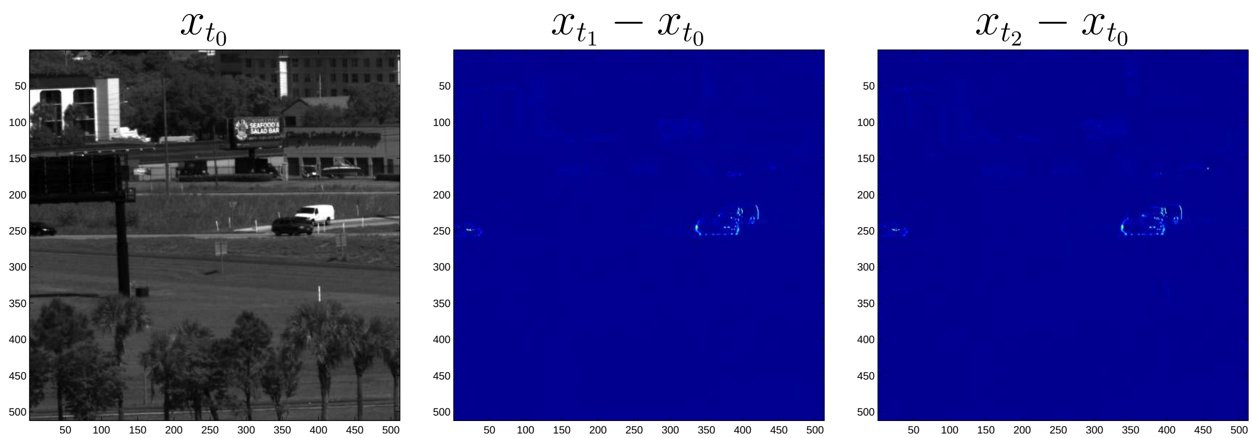
$$\|x_b - \tilde{x}_b\|_2^2 = \|\alpha - \tilde{\alpha}\|_2^2 = \sum_{k=0}^{7}\sum_{\ell=0}^{7} |\alpha_{k,\ell} - \tilde{\alpha}_{k,\ell}|^2.$$

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.

The Lapped Orthogonal Transform

The Lapped Orthogonal Transform is a *time-frequency* decomposition which is also an orthobasis. It has been used extensively in audio CODECS.

The essential idea is to divide up the real line into intervals with endpoints

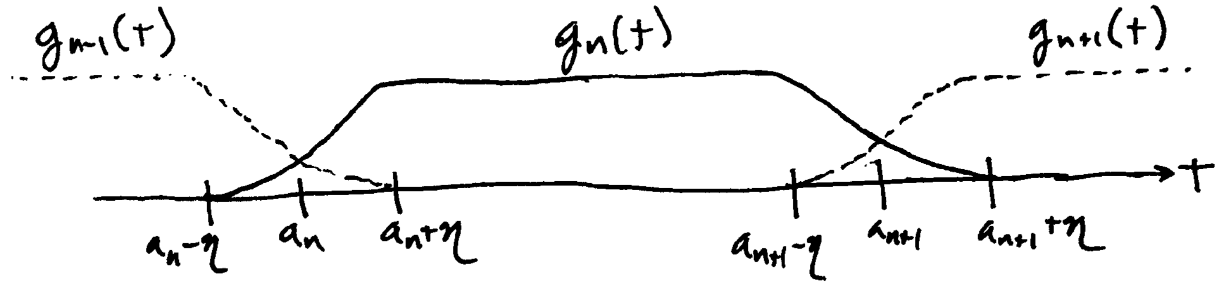
$$\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, a_3, \ldots$$

And then inside each of these intervals take a windowed cosine transform.

In its most general form, the collection of LOT orthobasis functions is

$$\left\{ g_n(t)\tilde{\phi}\left(\frac{t-a_n}{a_{n+1}-a_n}\right) \right\}$$

where $g_n(t)$ is a window that is "flat" in the middle, monotonic on its ends:

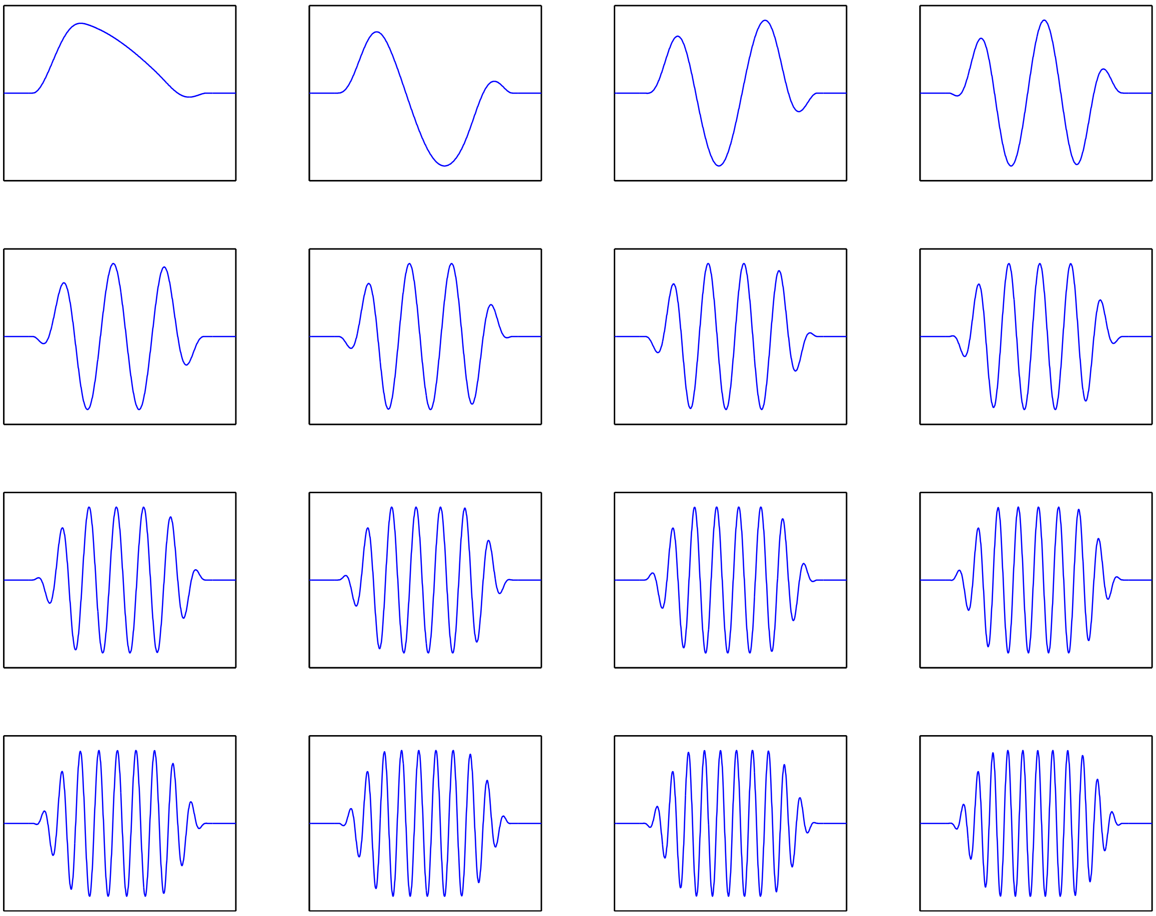


and obeys

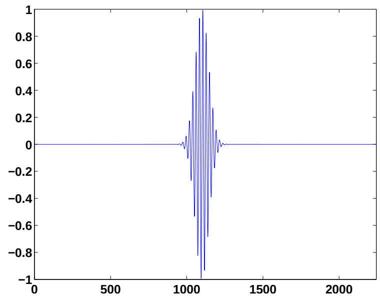
$$\sum_n |g_n(t)|^2 = 1 \quad \text{for all } t$$

The $\tilde{\phi}$ above must be symmetric around a_n and anti-symmetric around a_{n+1} — just like a cosine-IV function.

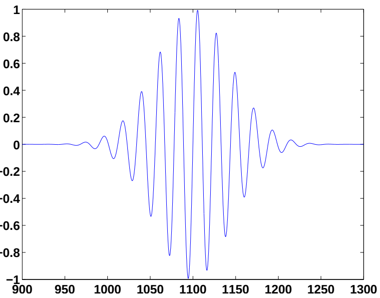
Plots of the LOT basis functions, single window, first 16 frequencies:



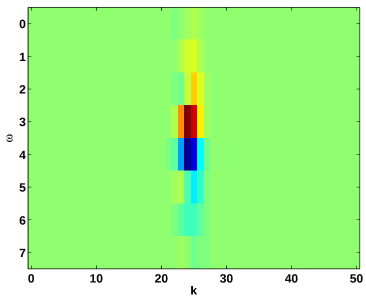
LOT of a modulated pulse:



pulse



zoom



grid of LOT coefficients

Non-orthogonal bases in $\mathbb{R}^N$

When $x \in \mathbb{R}^N$, basis representations fall squarely into the realm of linear algebra. Let $\psi_0, \psi_1, \ldots, \psi_{N-1}$ be a set of N linearly independent vectors in $\mathbb{R}^N$. Since the ψ_k are linearly independent, then every $x \in \mathbb{R}^N$ produces a unique sequence of inner products against the $\{\psi_k\}$. That is, we can recover x from the sequence of inner products

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} = \begin{bmatrix} \langle x, \psi_0 \rangle \\ \langle x, \psi_1 \rangle \\ \vdots \\ \langle x, \psi_{N-1} \rangle \end{bmatrix}.$$

Stacking up the (transposed) ψ_k as rows in an $N \times N$ matrix Ψ^*,

$$\Psi^* = \begin{bmatrix} — & \psi_0^* & — \\ — & \psi_1^* & — \\ \vdots & \vdots & \vdots \\ — & \psi_{N-1}^* & — \end{bmatrix},$$

we have the straightforward relationships

$$\alpha = \Psi^* x, \quad \text{and} \quad x = \Psi^{*-1}\alpha.$$

(In this case we know that Ψ^* is invertible since it is square and its rows are linearly independent.) Let $\tilde{\psi}_0, \tilde{\psi}_1, \ldots, \tilde{\psi}_{N-1}$ be the columns of Ψ^{*-1}:

$$\Psi^{*-1} = \begin{bmatrix} | & | & \cdots & | \\ \tilde{\psi}_0 & \tilde{\psi}_1 & \cdots & \tilde{\psi}_{N-1} \\ | & | & \cdots & | \end{bmatrix}.$$

Then the straightforward relation

$$x = \Psi^{*-1}\Psi^* x,$$

can be rewritten as the **reproducing formula**

$$x[n] = \sum_{k=0}^{N-1} \langle x, \psi_k \rangle \tilde{\psi}_k[n].$$

For the non-orthogonal case, we are using different families of basis functions for the analysis and the synthesis. The analysis operator that maps x to the $\alpha(k) = \langle x, \psi_k \rangle$ is the $N \times N$ matrix Ψ^*. The synthesis operator, which uses the vector α to build up x, is the $N \times N$ matrix Ψ^{*-1} which we could conveniently re-label as $\Psi^{*-1} = \tilde{\Psi}^*$. When the ψ_k are orthonormal, we have $\Psi = \Psi^{*-1}$, and so $\tilde{\Psi}^* = \Psi^*$, meaning that the analysis and synthesis basis functions are the same ($\tilde{\psi}_k = \psi_k$). In the orthonormal case, the analysis operator is Ψ^* and the synthesis operator is Ψ, matching our previous notation.

For non-orthogonal $\{\psi_k\}_k$, the Parseval theorem does not hold. However, we can put bounds on the energy of the expansion coefficients in relation to the energy of the signal x. In particular,

$$\begin{array}{c} \sigma_1^2 \|x\|_2^2 \ \leq \ \displaystyle\sum_{k=0}^{N-1} |\langle \psi_k, x \rangle|^2 \ \leq \ \sigma_N^2 \|x\|_2^2 \\ \Updownarrow \\ \sigma_1^2 \|x\|_2^2 \ \leq \ \|\alpha\|_2^2 \ \leq \ \sigma_N^2 \|x\|_2^2, \end{array}$$

where σ_1 is the smallest *singular value* of the analysis operator matrix Ψ and σ_N is its largest singular value.

To extend these ideas to infinite dimensions, we need to use the language of linear operators in place of matrices (which introduces a few interesting complications). Before doing this, we will take a first look at *overcomplete* expansions.

Overcomplete frames in $\mathbb{R}^N$

A sequence of vector $\psi_0, \psi_1, \ldots, \psi_M$ in $\mathbb{R}^N$ are a *frame* if there is no $x \in \mathbb{R}^N,\ x \neq 0$ that is orthogonal to all of the ψ_k. This means that the sequence of inner products

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{M-1} \end{bmatrix} = \begin{bmatrix} \langle x, \psi_0 \rangle \\ \langle x, \psi_1 \rangle \\ \vdots \\ \langle x, \psi_{M-1} \rangle \end{bmatrix}.$$

will be unique for every different x. The difference between a basis and a frame is that we allow $M \geq N$, and so the number of inner product coefficients in α can exceed the number of entries in x. If we again stack up the (transposed) ψ_k as rows in an $M \times N$ matrix Ψ^*,

$$\Psi^* = \begin{bmatrix} — & \psi_0^* & — \\ — & \psi_1^* & — \\ \vdots & \vdots & \vdots \\ — & \psi_{M-1}^* & — \end{bmatrix},$$

this means that Ψ^* is overdetermined and has no null space (and hence has full column-rank). Of course, Ψ^* does not have an inverse, so we must take a little more caution with the reproducing formula.

Since the $M \times N$ matrix Ψ^* has full column rank, we know that the $N \times N$ matrix $\Psi\Psi^*$ is invertible. The reproducing formula can then comes from

$$x = (\Psi\Psi^*)^{-1}\Psi\Psi^* x.$$

Now define the synthesis basis vectors $\tilde{\psi}_k$ as the columns of the pseudo-inverse $(\Psi\Psi^*)^{-1}\Psi$:

$$\tilde{\psi}_k = (\Psi\Psi^*)^{-1}\psi_k.$$

Then the reproducing formula is almost identical as the above (except now we are using $M \geq N$ vectors to build up x):

$$x[n] = \sum_{k=0}^{M-1} \langle x, \psi_k \rangle \tilde{\psi}_k[n].$$

We have the same relationship as above between the energy in the coefficients $\alpha = \Psi^* x$ and the signal x:

$$\begin{array}{ccccc} \sigma_1^2 \|x\|_2^2 & \leq & \displaystyle\sum_{k=0}^{M-1} |\langle \psi_k, x \rangle|^2 & \leq & \sigma_N^2 \|x\|_2^2 \\ & & \Updownarrow & & \\ \sigma_1^2 \|x\|_2^2 & \leq & \|\alpha\|_2^2 & \leq & \sigma_N^2 \|x\|_2^2, \end{array}$$

where now σ_1 is the smallest *singular value* of the analysis operator matrix Ψ^* and σ_N is its largest singular value (i.e. σ_N^2 is the largest eigenvalue of the symmetric positive-definite matrix $\Psi\Psi^*$). If the rows of Ψ^* are orthogonal and all have the same energy A, then $\Psi\Psi^* = A \cdot \text{Identity}$ and we have a Parseval relation

$$\langle \Psi^* x, \Psi^* y \rangle = \langle x, \Psi\Psi^* y \rangle = A \langle x, y \rangle$$

and so

$$\sum_{k=0}^{M-1} |\langle x, \psi_k \rangle|^2 = \|\Psi^* x\|_2^2 = A \|x\|_2^2.$$

Moral: A frame can be overcomplete and still obey a Parseval relation.

Example: Mercedes-Benz frame in $\mathbb{R}^2$

Let's start with the simplest possible example of a tight frame for $H = \mathbb{R}^2$:

$$\psi_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \psi_2 = \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix}, \qquad \psi_3 = \begin{bmatrix} -\sqrt{3}/2 \\ -1/2 \end{bmatrix}.$$

Sketch it here:

The associated frame operator is the 3×2 matrix

$$\Psi^* = \begin{bmatrix} 0 & 1 \\ \sqrt{3}/2 & -1/2 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}.$$

Thus

$$\Psi\Psi^* =$$

and so

$$\_\_\_\_\_\_\_\_ \;\leq\; \|\Psi^* x\|_2^2 \;\leq\; \_\_\_\_\_\_\_\_$$

and

$$A = B = \_\_\_\_\_\_\_\_$$

Example: Unions of orthobases in $\mathbb{R}^N$ Suppose our sequence $\{\psi_\gamma\}$ is a union of sequences, each of which is an orthobasis:

$$\{\psi^1_{\gamma_1}\}_{\gamma_1\in\Gamma_1} \ \cup\ \{\psi^2_{\gamma_2}\}_{\gamma_2\in\Gamma_2} \ \cup\ \cdots\ \cup \{\psi^L_{\gamma_L}\}_{\gamma_L\in\Gamma_K}$$

Then

$$\begin{aligned}\|\Psi x\|_2^2 &= \sum_{\gamma_1\in\Gamma_1} |\langle x, \psi^1_{\gamma_1}\rangle|^2 \ +\ \sum_{\gamma_2\in\Gamma_2} |\langle x, \psi^2_{\gamma_2}\rangle|^2 \ +\ \cdots\ + \sum_{\gamma_L\in\Gamma_L} |\langle x, \psi^L_{\gamma_L}\rangle|^2 \\ &= L\|x\|_2^2\end{aligned}$$