Information Extraction and Data Compression using **Wavelets**

What are they?

How do I use them?

Why would I want to?

Fourier analysis

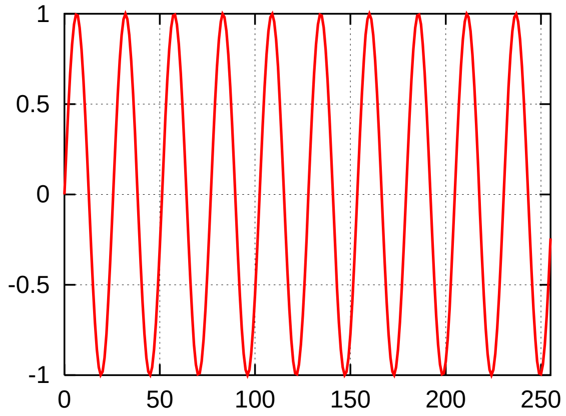
$\sin(kx)$ and $\cos(kx)$ for a set of **basis functions** from which any function can be built.

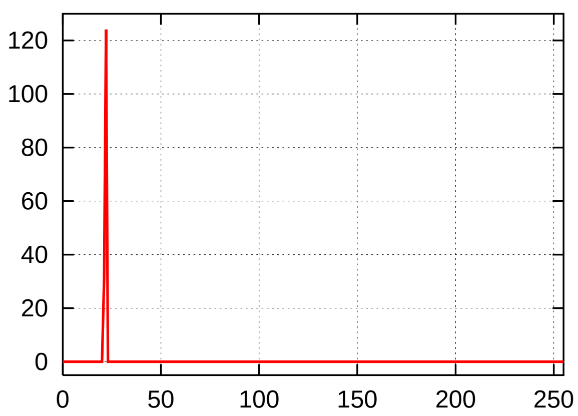
$$\boxed{f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)}$$

where the Fourier coefficients a_k and b_k are calculated by

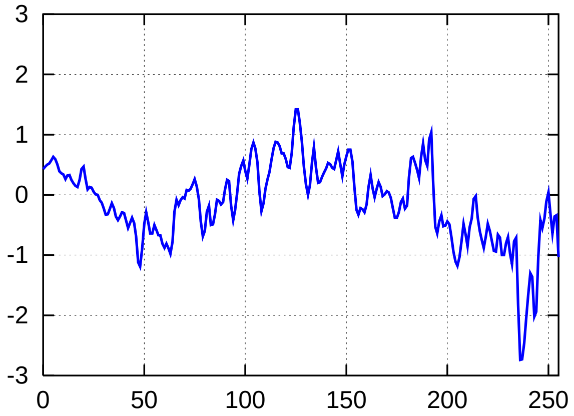
$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \\ a_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx dx \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx dx \end{aligned}$$

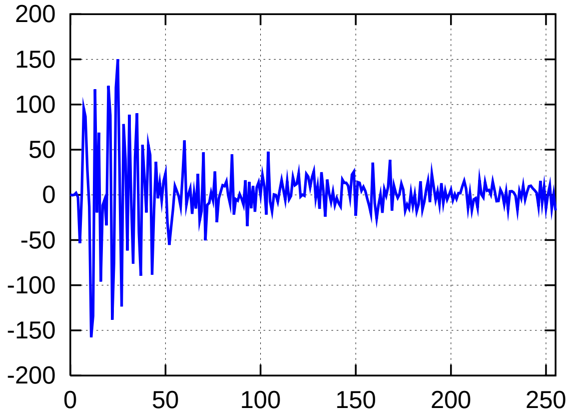
Fourier transform of a sine curve...



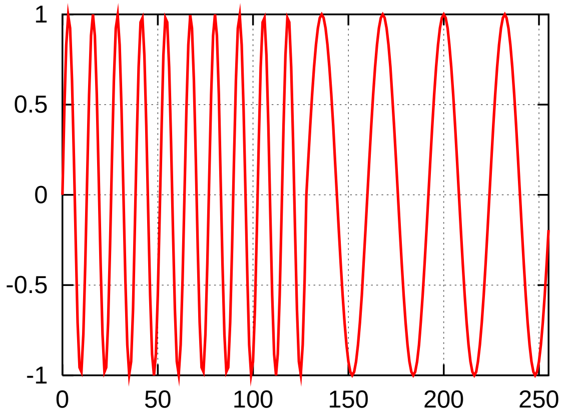


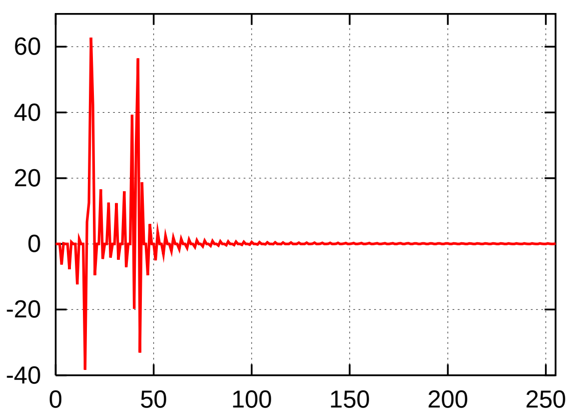
And a more complex signal...



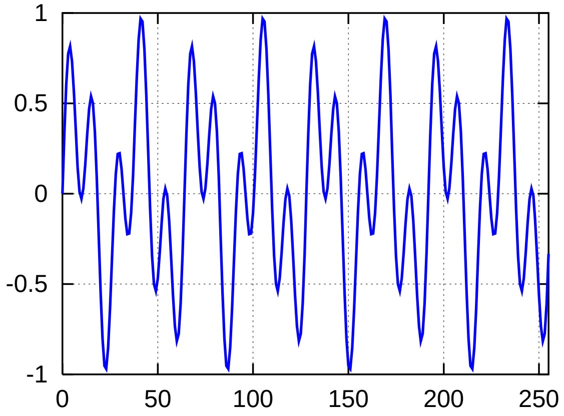


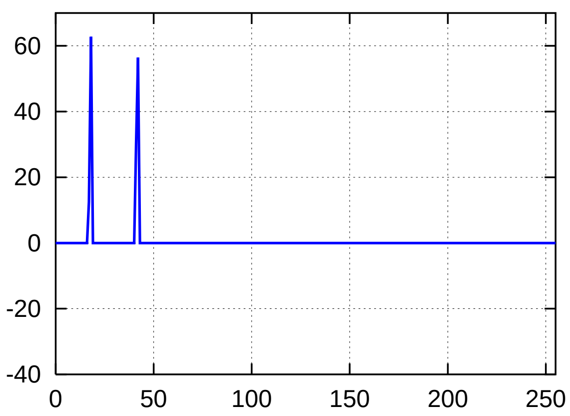
Case 1: f_1 for time t, followed by f_2





Case 2: $f_1 + f_2$ for the entire time

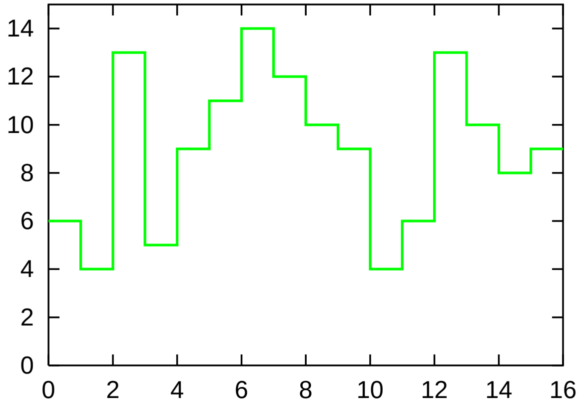




Multiresolution decomposition of a function $f(x)$

(example from Dremin et al. hep-ph/0101182)

$$f(x) = \sum_{k=0}^{15} s_{4,k}\phi_{4,k}(x)$$



where

$$s_{4,k} = f\left(\frac{k}{16}\right)/4$$

and $\phi_{4,k}$ is a block (height 4, width $\frac{1}{16}$) centered on the k-th bin.

If we impose a normalisation condition (for arbitrary j)

$$\int |\phi_{j,k}|^2 dx = 1 \qquad \left(\Delta x_j = \frac{1}{2^j}\right)$$

then $\phi_{jk}(x)$ have the following form

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$$

Decompose $f(x)$ by considering the sums and differences of neighbouring bins at increasingly coarse resolution levels. The normalised sums $s_{j-1,k}$ and differences $d_{j-1,k}$ for an arbitrary resolution level j are given by

$$s_{j-1,k} = \frac{1}{\sqrt{2}}\left[s_{j,2k} + s_{j,2k+1}\right] \quad d_{j-1,k} = \frac{1}{\sqrt{2}}\left[s_{j,2k} - s_{j,2k+1}\right]$$

Introduce $\psi(x)$ which is $+1$ and -1 in neighbouring bins and satisfies

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$$

which allows us to write $f(x)$ as

$$f(x) = \sum_{k=0}^{7} s_{3,k}\phi_{3,k}(x) + \sum_{k=0}^{7} d_{3,k}\psi_{3,k}(x)$$

which is equivalent to

$$f(x) = s_{0,0}\phi_{0,0}(x) + d_{0,0}\psi_{0,0}(x) + \sum_{k=0}^{1} d_{1,k}\psi_{1,k}(x) + \sum_{k=0}^{3} d_{2,k}\psi_{2,k}(x) + \sum_{k=0}^{7} d_{3,k}\psi_{3,k}(x)$$

Wavelet Analysis

ψ_{jk} are the **basis functions** from which any function can be built.

$$\boxed{f(x) = \sum_j \sum_k d_{jk} \psi_{jk}(x)}$$

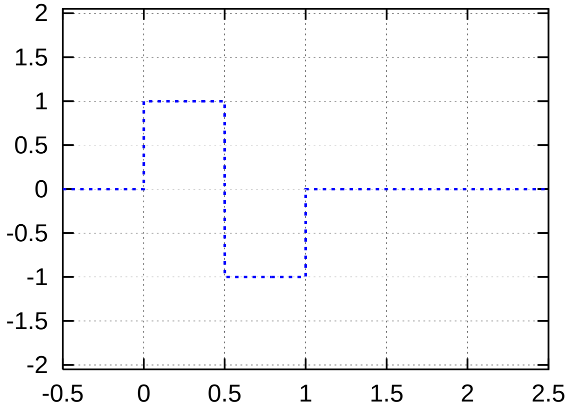
where d_{jk} are the wavelet coefficients.

Each basis function $\psi_{jk}(x)$ is obtained from one mother function $\psi(x)$ by shifting and scaling:

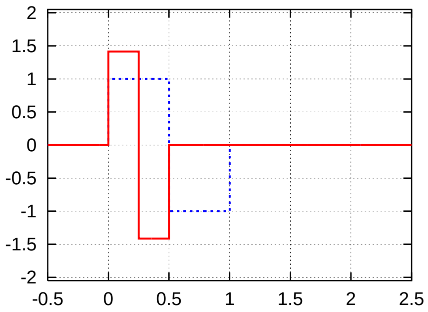
$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$$

- **j** is the **scaling** parameter
- **k** is the **shifting** parameter

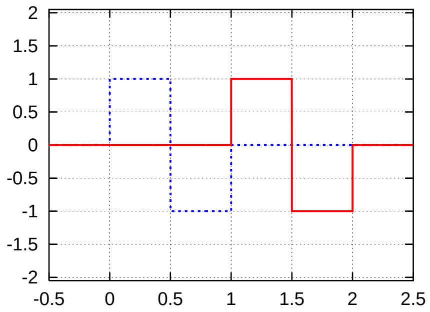
The **Haar** function



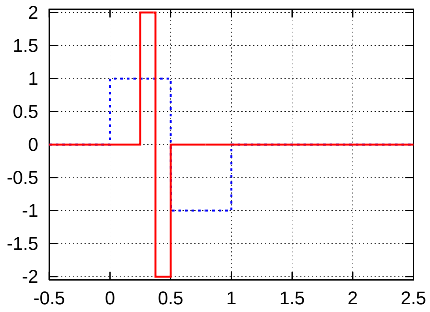
$$\psi_H(x) = \begin{cases} 1 & : \quad 0 \le x < \frac{1}{2} \\ -1 & : \quad \frac{1}{2} \le x < 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$



$\psi_{10}(x)$



$\psi_{01}(x)$



$\psi_{21}(x)$

How do you get the mother function in the first place?

$$\psi(x) = ?$$

The **Daubechies mother function**

First we relate a function $\phi(x)$ to its dilated and translated version:

$$\phi(x) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \phi(2x - k) \tag{1}$$

The solution to this equation, $\phi(x)$ is known as the scaling function. Once we know the scaling function we can derive the mother function $\psi(x)$ using:

$$\psi(x) = \sqrt{2} \sum_{k=0}^{2M-1} g_k \phi(2x - k) \tag{2}$$

The orthogonality of the scaling functions

$$\int \phi(x)\phi(x-m)dx = 0 \tag{3}$$

implies

$$\sum_k h_k h_{k+2m} = \delta_{0m} \tag{4}$$

The orthogonality of the wavelets to the scaling functions

$$\int \psi(x)\phi(x-m)dx = 0 \tag{5}$$

implies

$$\sum_k h_k g_{k+2m} = 0 \tag{6}$$

which has the solution

$$g_k = (-1)^k h_{2M-k-1} \tag{7}$$

We also have the condition that the wavelets are orthogonal to all polynomials up to the power $(M-1)$

$$\int x^n \psi(x)dx = 0, \ n = 0, \ldots, (M-1) \tag{8}$$

which implies

$$\sum_k k^n g_k = 0 \tag{9}$$

Substituting Eqn. 7 into Eqn. 9 gives

$$\sum_k (-1)^k k^n h_k = 0 \tag{10}$$

The final condition is that the wavelets are normalized:

$$\int \phi(x)dx = 1 \tag{11}$$

which gives

$$\sum_k h_k = \sqrt{2} \tag{12}$$

If we now write Eqns. 4, 10 and 12 explicitly for the case $M = 2$ we get

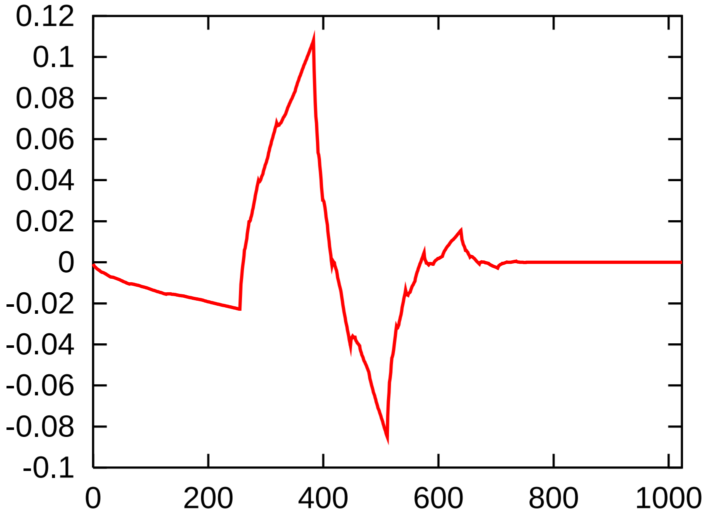
$$\begin{aligned} h_0 h_2 + h_1 h_3 &= 0 \\ h_0 - h_1 + h_2 - h_3 &= 0 \\ -h_1 + 2h_2 - 3h_3 &= 0 \\ h_0 + h_1 + h_2 + h_3 &= \sqrt{2} \end{aligned}$$

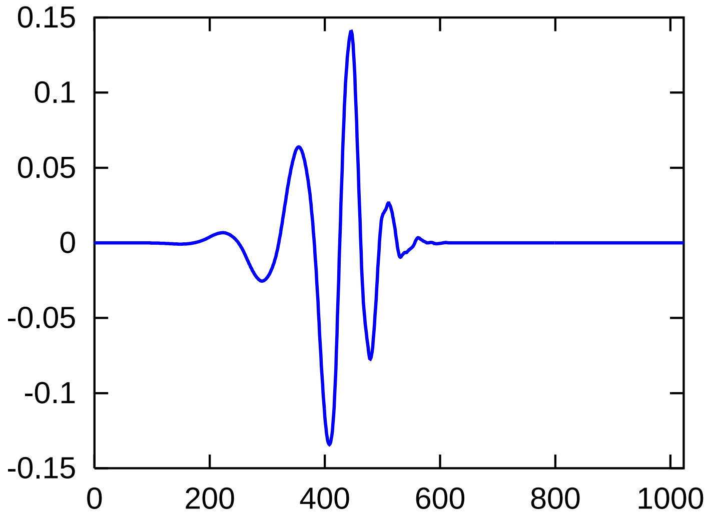
The solution of this system is

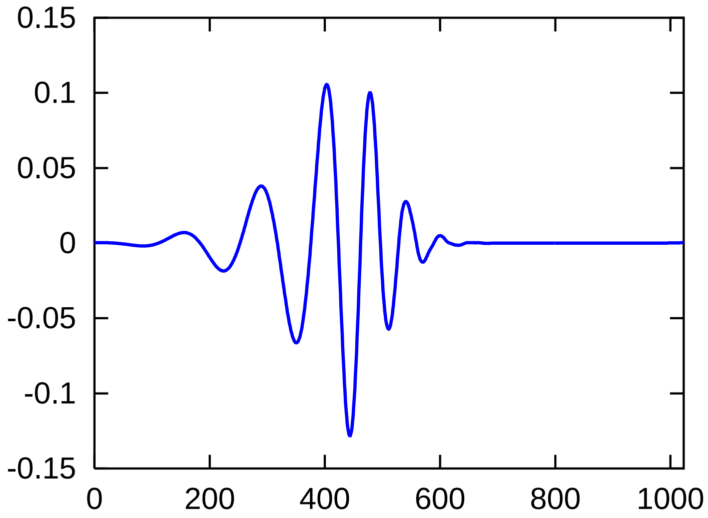
$$\begin{aligned} h_0 &= \frac{1}{4\sqrt{2}}(1+\sqrt{3}) \\ h_1 &= \frac{1}{4\sqrt{2}}(3+\sqrt{3}) \\ h_2 &= \frac{1}{4\sqrt{2}}(3-\sqrt{3}) \\ h_3 &= \frac{1}{4\sqrt{2}}(1-\sqrt{3}) \end{aligned}$$

which define the $\mathbf{D^4}$ mother wavelet,
the simplest member of the **Daubechies** family.

The **Daubechies** family







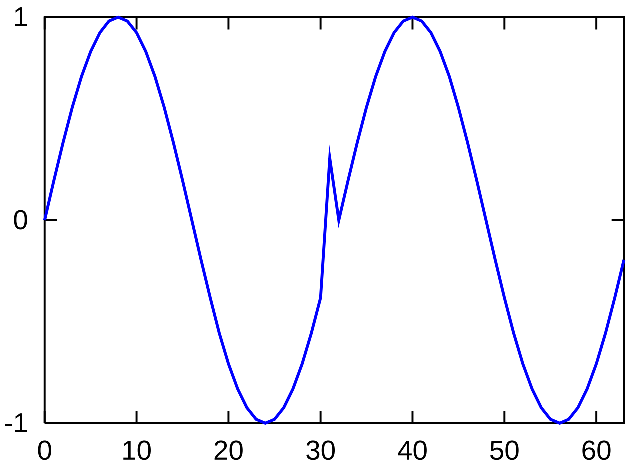
The Discrete Wavelet Transform

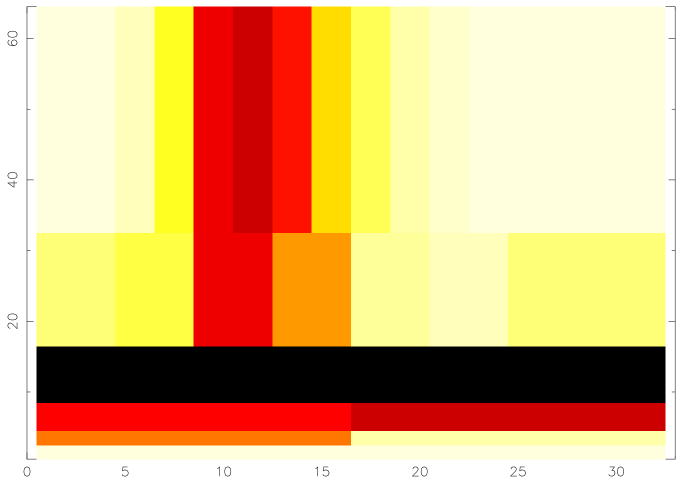
$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & & \\ c_3 & -c_2 & c_1 & -c_0 & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & \\ & & c_3 & -c_2 & c_1 & -c_0 & & \\ & & & & c_0 & c_1 & c_2 & c_3 \\ & & & & c_3 & -c_2 & c_1 & -c_0 \\ c_2 & c_3 & & & & & c_0 & c_1 \\ c_1 & -c_0 & & & & & c_3 & -c_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \end{bmatrix}$$

- (c_0, c_1, c_2, c_3) – **smooth** filter
- (c_3, $-c_2$, c_1, -c_0) – **detail** filter

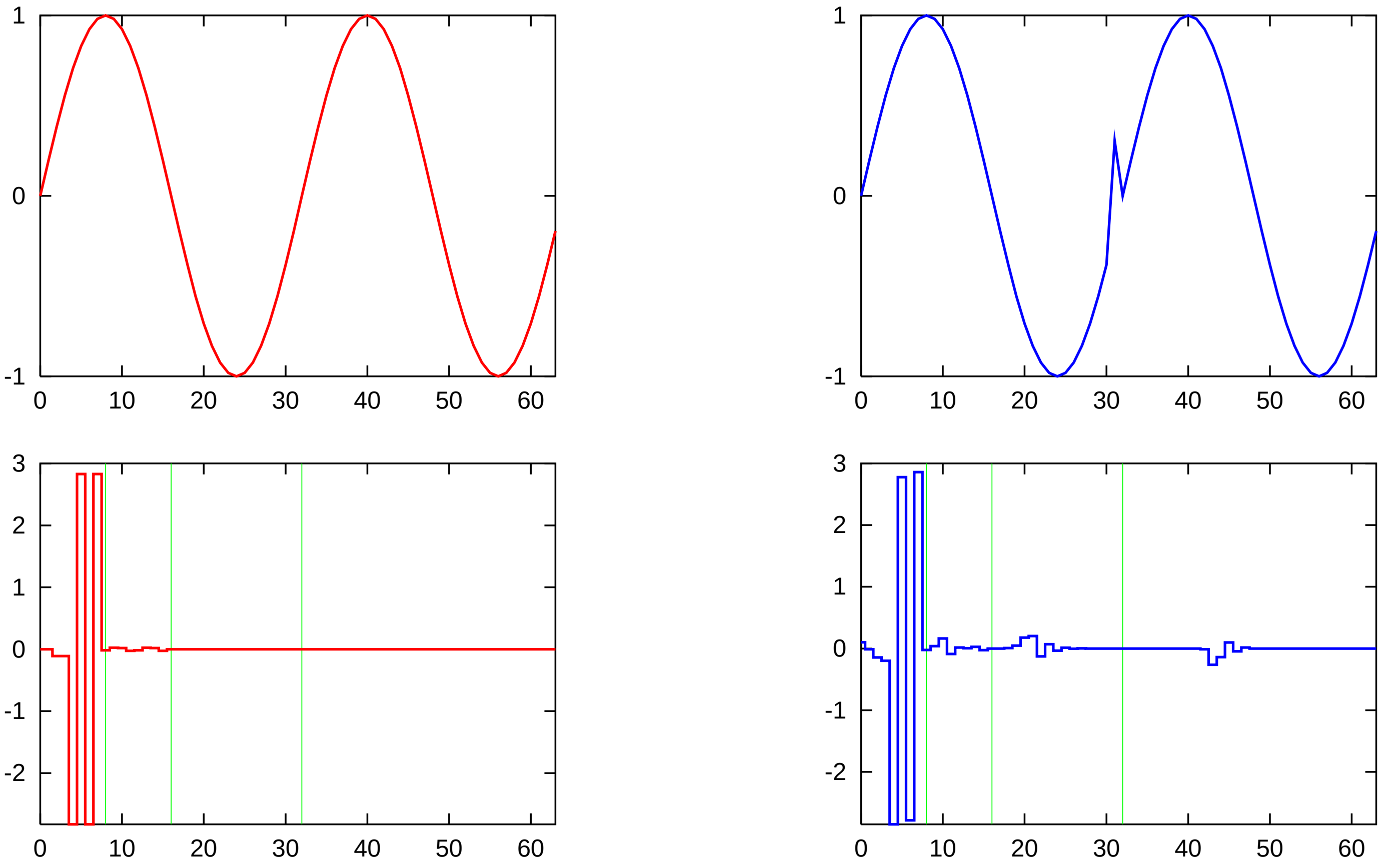
Demonstration of the pyramid algorithm

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix}
\xrightarrow{\text{apply 8x8}}
\begin{bmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \end{bmatrix}
\xrightarrow{\text{rearrange}}
\left[\begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ \hline d_1 \\ d_2 \\ d_3 \\ d_4 \end{array}\right]
\xrightarrow{\text{apply 4x4}}
\left[\begin{array}{c} S_1 \\ D_1 \\ S_2 \\ D_2 \\ \hline d_1 \\ d_2 \\ d_3 \\ d_4 \end{array}\right]
\xrightarrow{\text{rearrange}}
\left[\begin{array}{c} S_1 \\ S_2 \\ \hline D_1 \\ D_2 \\ \hline d_1 \\ d_2 \\ d_3 \\ d_4 \end{array}\right]
$$

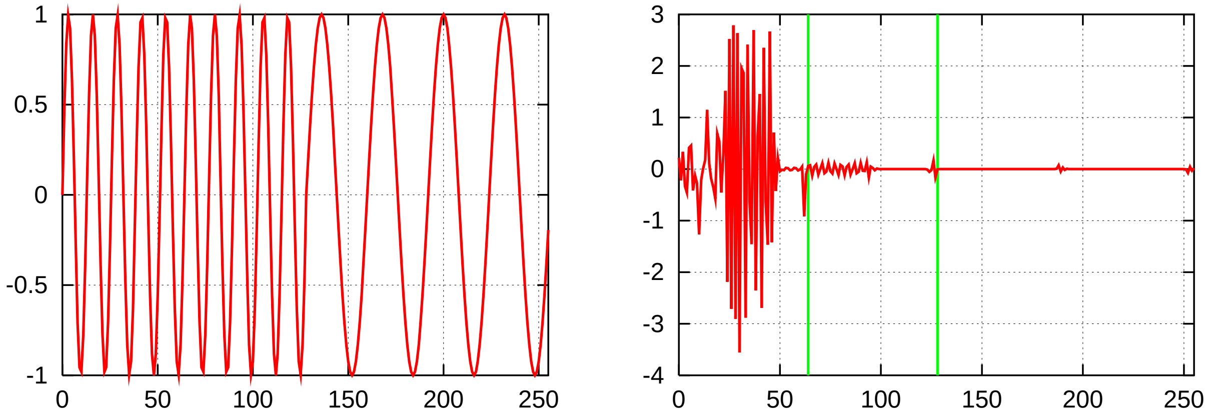




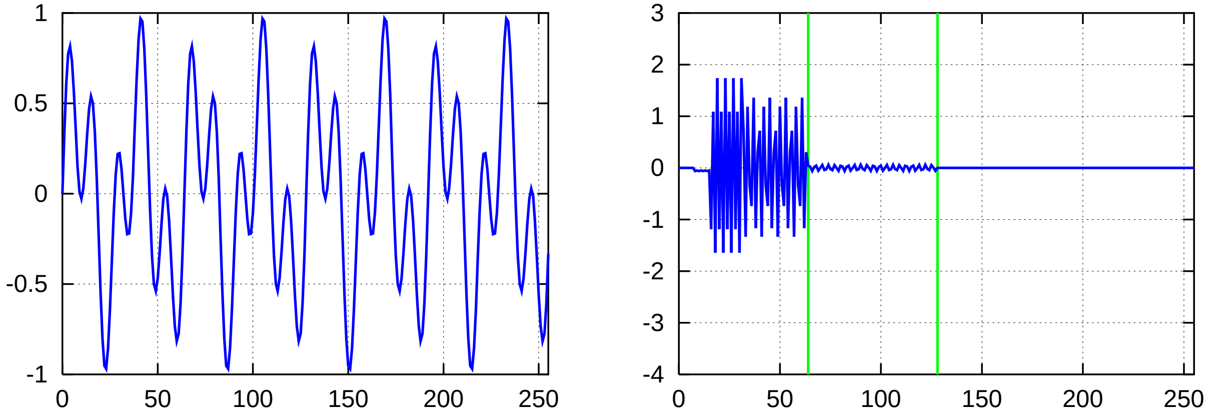
Example: wavelet transform of a sine curve



Case 1: f_1 for time t, followed by f_2



Case 2: $f_1 + f_2$ for the entire time



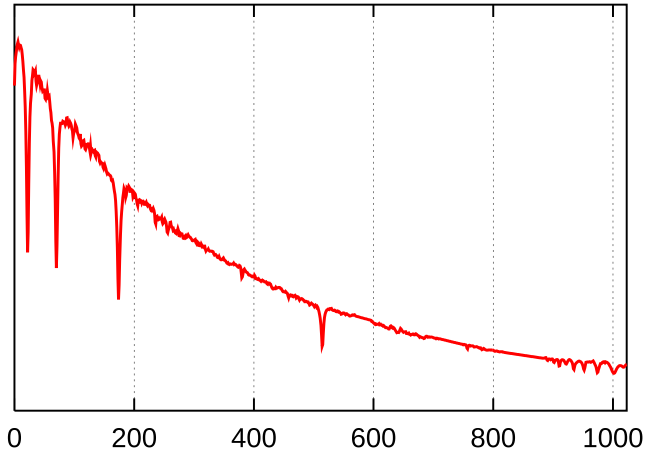
Information Extraction

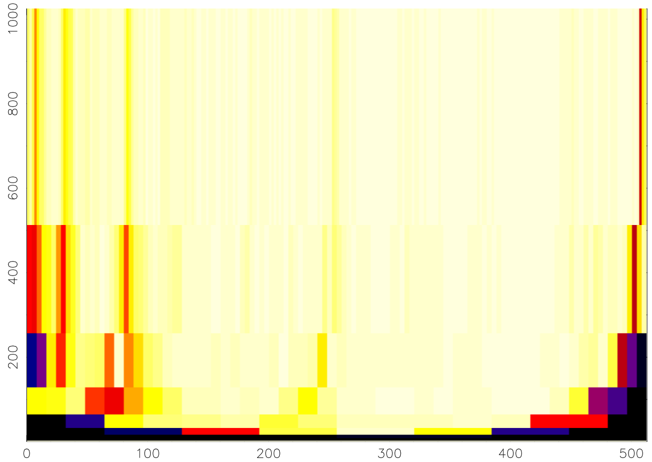
On what scale in my data/image do important features occur?

This is important for applications such as:

- Pattern recognition (eg. heartbeat intervals in ECGs)
- Source detection (eg. from CMB maps)
- Analysis of time-varying phenomena (eg. solar bursts)
- Denoising (eg. helioseismic data)
- Automatic classification (eg. blood cells)

Example: extracting important features from a stellar spectrum





Want to automatically determine which parameters will be the most important:

1. take a set of objects and wavelet transform each
2. determine which coefficients have greatest variance over the entire set
3. use these coefficients to represent each spectrum
4. classify spectrum based on these coefficients

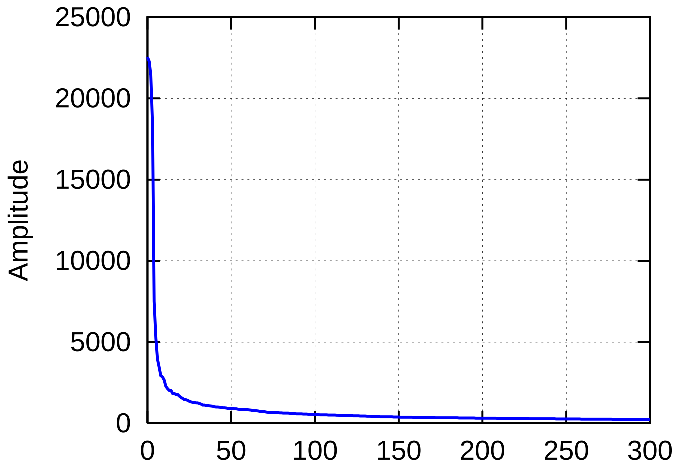
Data compression

There are several different requirements of data compression

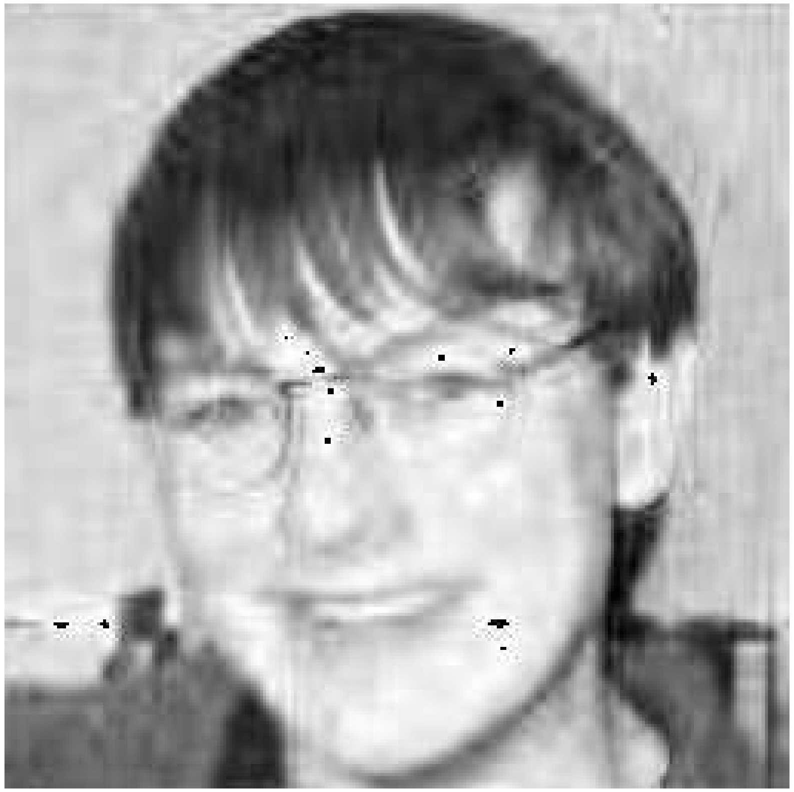
- To keep quality as good as the original
- To maximise compression ratio
- To maximise speed of compression

The compression method you choose depends on which of these is more important to you.

Example: image compression ($256 \times 256 = 65536$ pix)



Out of 65536 pixels, very few are needed for a reasonable reconstruction...



2%

5%

10%

20%

Conclusion:

Wavelet analysis proves that at least **80%** of data in Ian's head in uneccessary!