MLISP: Machine Learning in Signal Processing

Lecture 10

Prof. V. I. Morgenshtern

Scribe: M. Solomon

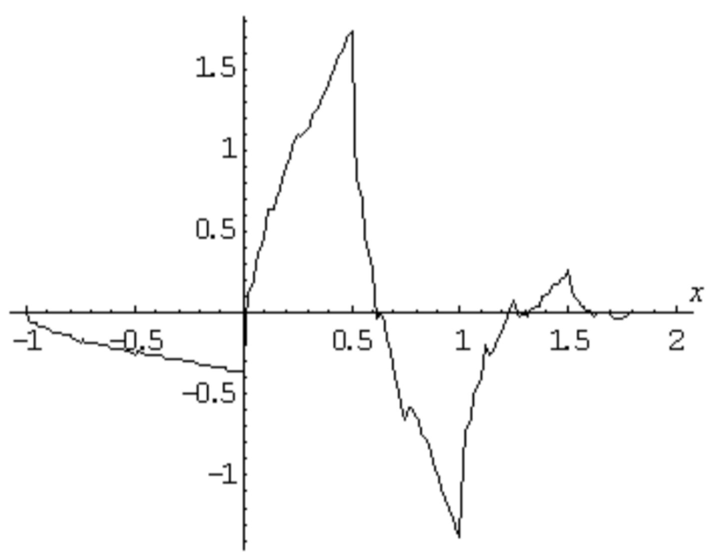
Illustrations: Lecture notes on basics of wavelets by Kon

Agenda

1. Wavelets: the basic idea
2. Haar wavelets

1 Wavelets: the basic idea

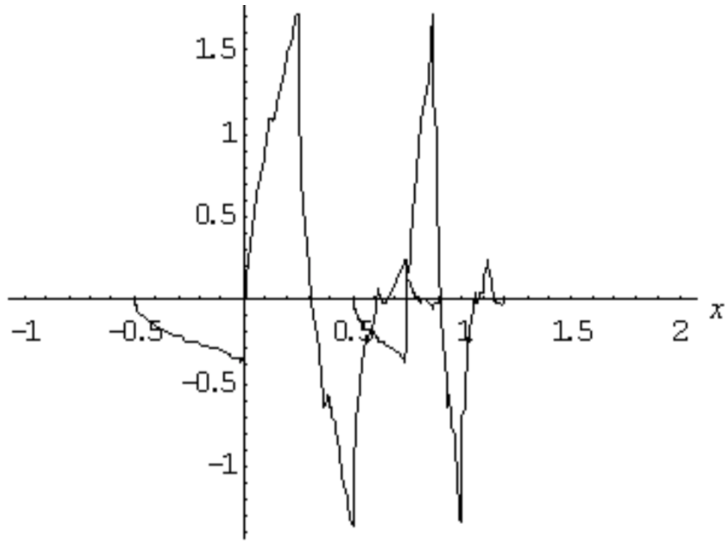
The basic idea is to start with an appropriate basic function $h(t)$ (the father wavelet), which might look like this:



Then form all possible translations by integers and all possible stretchings by powers of 2 of the father wavelet:

$$h_{nk}(t) = 2^{\frac{n}{2}} h(2^n t - k).$$

Above, $2^{\frac{n}{2}}$ is just a normalization constant. To illustrate this construction, below we display $h(2t)$ and $h(4t-3)$:



It turns out that if h is chosen properly, then $h_{nk}(t)$ are orthogonal:

$$\langle h_{nk}, h_{n'k'} \rangle = \int h_{nk}(t) h_{n'k'}(t) dt = 0$$

unless $n = n'$ and $k = k'$.

Furthermore, h_{nk} form a complete set for a wide class of functions (details later). Therefore every function of interest to us can be expected as a linear combination of these basis functions:

$$f(t) = \sum_{n,k} c_{nk} h_{nk}(t).$$

How do we find the coefficients c_{nk}? By completeness:

$$f(t) = \sum_{n,k} c_{nk} h_{nk}(t)$$

for some c_{nk}. To find the coefficients, use the orthonormality of $h_{nk}(t)$:

$$\langle f, h_{n'k'} \rangle = \sum_{n,k} c_{nk} \langle h_{nk}, h_{n'k'} \rangle = c_{n'k'}$$

which implies that

$$c_{nk} = \langle f, h_{n'k'} \rangle = \int f(t) h_{nk}(t) dt.$$

This is just like in the case of Fourier series.

There are major advantages compared to Fourier series. For high frequencies (n large), the functions $h_{nk}(t)$ have good localization (they get thinner as $n \to \infty$). The location of high frequency components can be seen from wavelet analysis, but not from Fourier series.

Next we consider the simples possible wavelet construction – Haar wavelets.

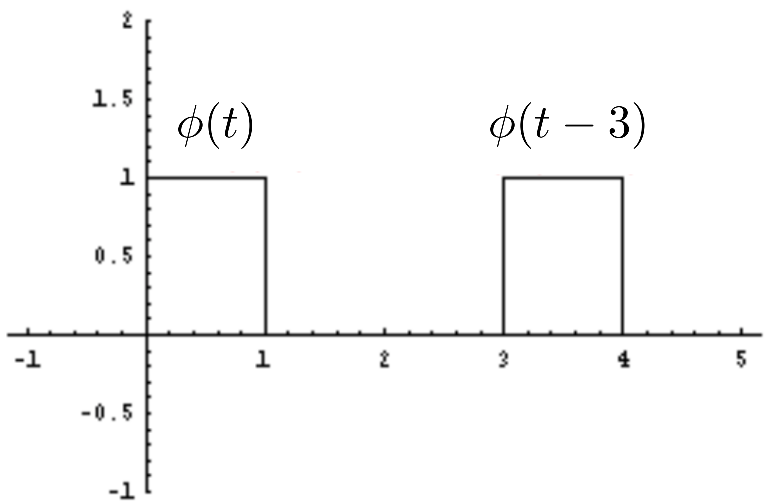
2 Haar wavelets

2.1 The 'pixel' approximation spaces

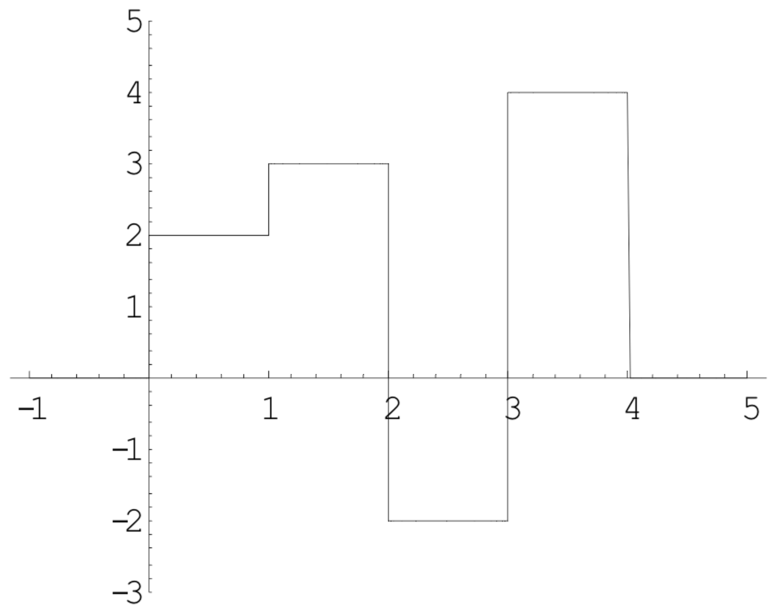
Suppose we have a *pixel* function

$$\phi(t) = \begin{cases} 1, & \text{if} \quad 0 \leqslant t \leqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

We wish to build all other functions out of $\phi(t)$ and all its integer translates $\phi(t-k)$, $k \in \mathbb{Z}$:



Note that any function that is constant on integers can be written as a linear combination of $\phi(t-k)$. For example, if $f(t)$ look like this:



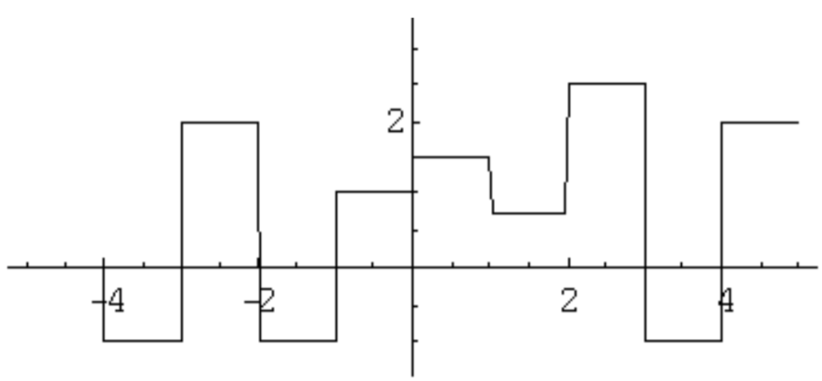
it can be written as follows:

$$f(t) = 2\phi(t) + 3\phi(t-1) - 2\phi(t-2) + 4\phi(t-3).$$

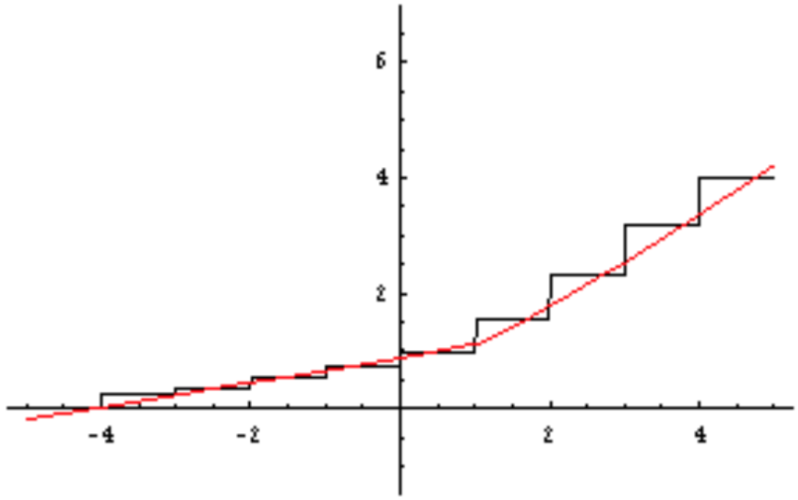
Let V_0 denote the set of all squared integrable functions that are constant on integer intervals. Every element of $g \in V_0$ can be written as a linear combination of integer translates of the pixel ϕ:

$$g(t) = \sum_{k \in \mathbb{Z}} a_k \phi(t-k).$$

The elements of V_0 looks like this:



Given a function $f(t)$ (that is not constant on integers) we can approximate it by a linear combination of $\phi(t-k),\ k \in \mathbb{Z}$, i.e., by an element of V_0:



To get better approximation shrink the basic function:

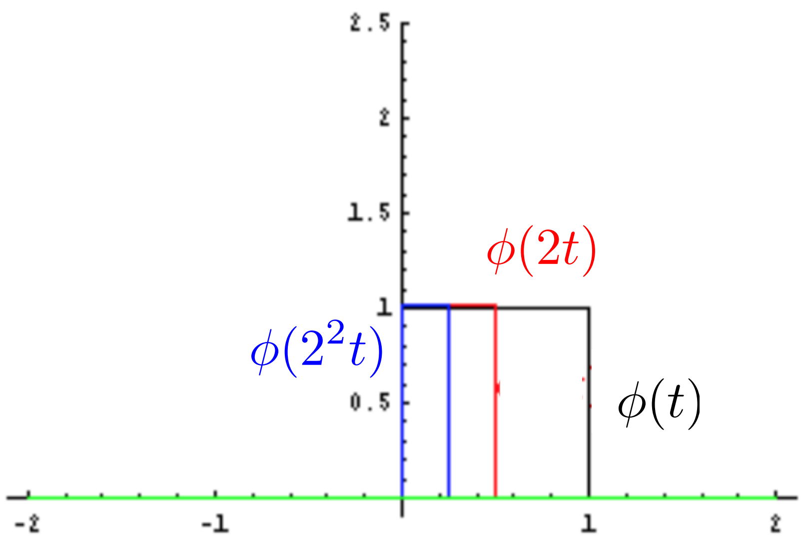
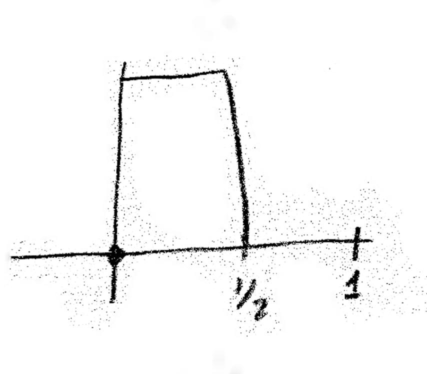
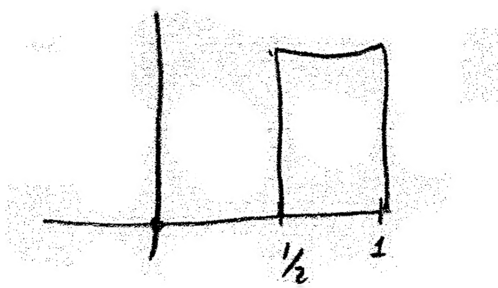


Figure 1: $\phi(t), \phi(2t), \phi(2^2t)$

In details, $\phi(2t)$ looks like this:

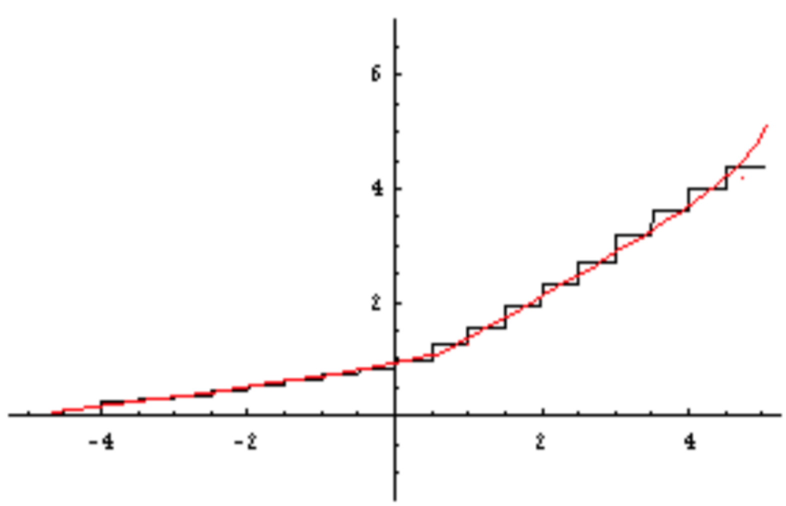


and $\phi(2t-1)$ looks like this



To see this note: $2t = 1$ iff $t = \frac{1}{2}$ and $2t - 1 = 1$ iff $t = 1$.

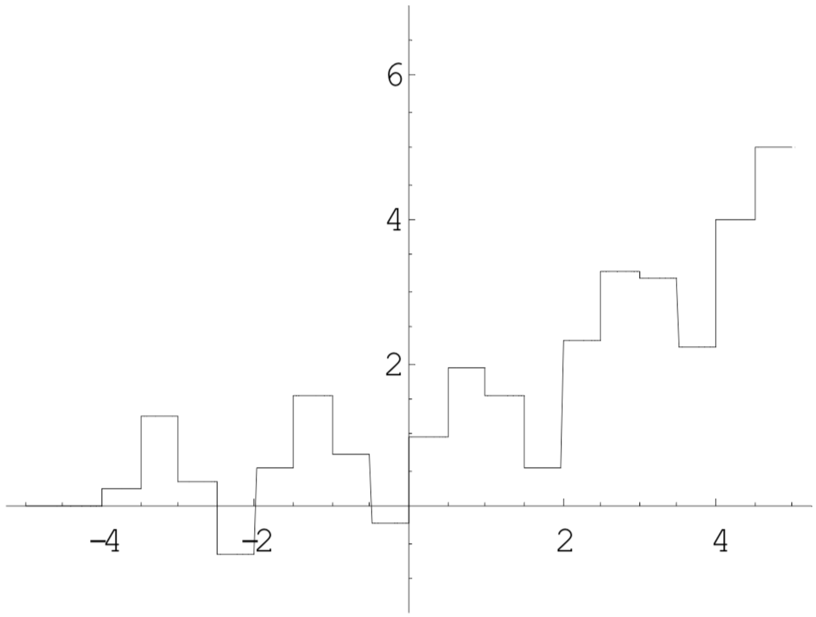
The narrower functions $\phi(2t + k)$, $k \in \mathbb{Z}$, give a better approximation to $f(t)$:



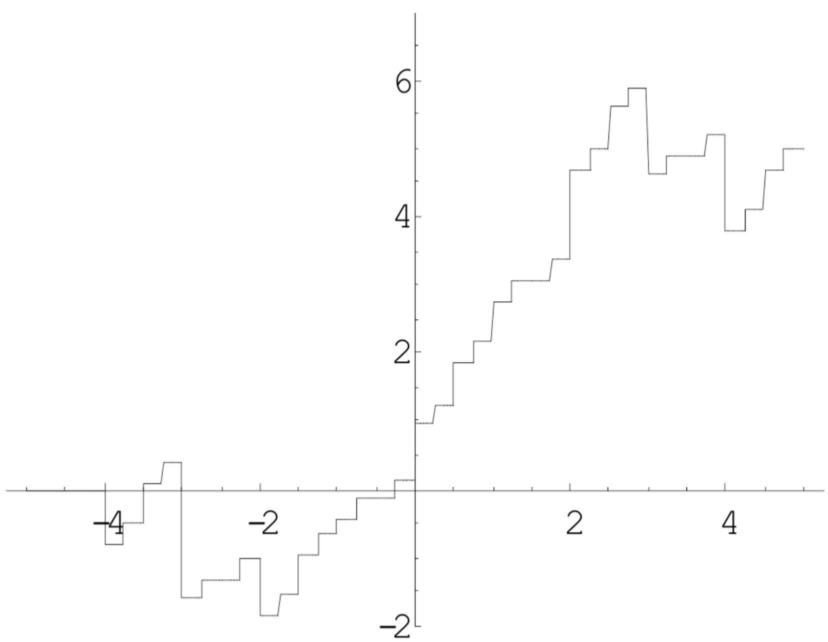
Let V_1 denote the set of all square integrable functions that are constant on all half-integers. Every element of $g \in V_1$ can be written as a linear combination of integer translates of the pixel $\phi(2t)$:

$$g(t) = \sum_{k \in \mathbb{Z}} a_k \phi(2t - k).$$

The elements of V_1 look like this:



We can continue in the analogous fashion. The elements of V_2 look like this:



In general, let V_j denote the set of all square integrable functions that are constant on 2^{-j}-length intervals. Every element of $g \in V_j$ can be written as a linear combination of integer translates of the pixel $\phi(2^j t)$:

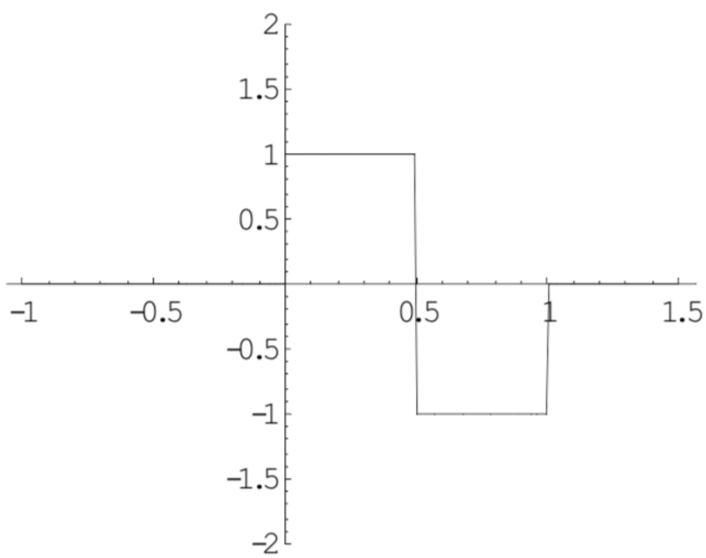
$$g(t) = \sum_k a_k \phi(2^j t - k).$$

2.2 The multi-scale spaces

The pixel function $\phi(\cdot)$ generates the approximation space we need. However its scaled versions and translates are not orthonormal to each other. Hence, we can use each space V_j separately, but we cannot use $\ldots V_{-1}, V_0, V_1, V_2, \ldots$ together at once. We do not have a *multiscale space.* We fix this problem next. Define the *father* function:

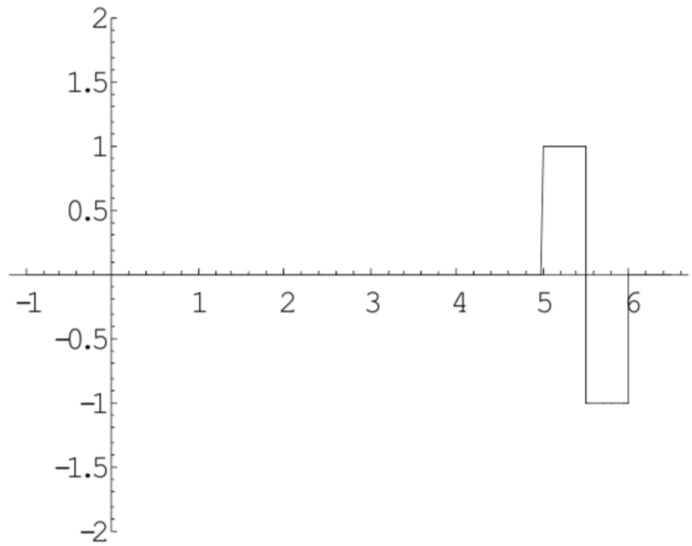
$$\psi(t) = \begin{cases} 1, & \text{if } \ 0 \leqslant t \leqslant \frac{1}{2} \\ -1, & \text{if } \ \frac{1}{2} \leqslant t \leqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

This function looks like this:

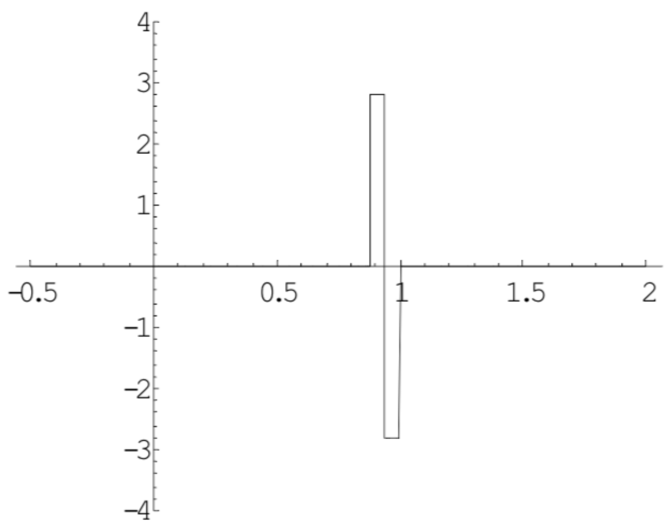


It is important not to confuse $\psi(\cdot)$ and $\phi(\cdot)$!

From the father function we generate the family of Haar wavelets by integer translation, $\psi_{0,5} = \psi(t-5)$:



and scaling, $\psi_{3,7} = 2^{\frac{3}{2}}\psi(2^3 t - 7)$:



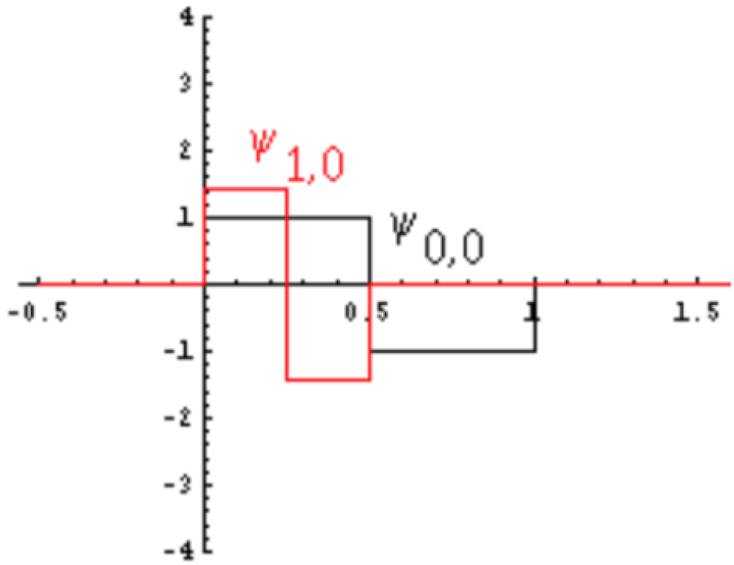
In general,

$$\psi_{jk} = 2^{\frac{j}{2}}\psi(2^j t - k), \;\; j \in \mathbb{Z}, k \in \mathbb{Z}.$$

It is not difficult to see that Haar wavelets are orthogonal across scales and within the same scale:

$$\left\langle \psi_{jk}, \psi_{j'k'} \right\rangle = \int_{-\infty}^{+\infty} \psi_{jk}(t)\psi_{j'k'}(t)dt = 0$$

if $j \neq j'$ or $k \neq k'$. Indeed, If $j = j'$ and $k \neq k'$, then $\left\langle \psi_{jk}, \psi_{j'k'} \right\rangle = 0$ because $\psi_{jk}(t) = 0$ for those t for which $\psi_{jk'}(t) \neq 0$ and vice versa. If $j \neq j'$ then $\left\langle \psi_{jk}, \psi_{j'k'} \right\rangle = \int \psi_{jk}(t)\psi_{j'k'}(t)dt = 0$, as can be seen from the figure:

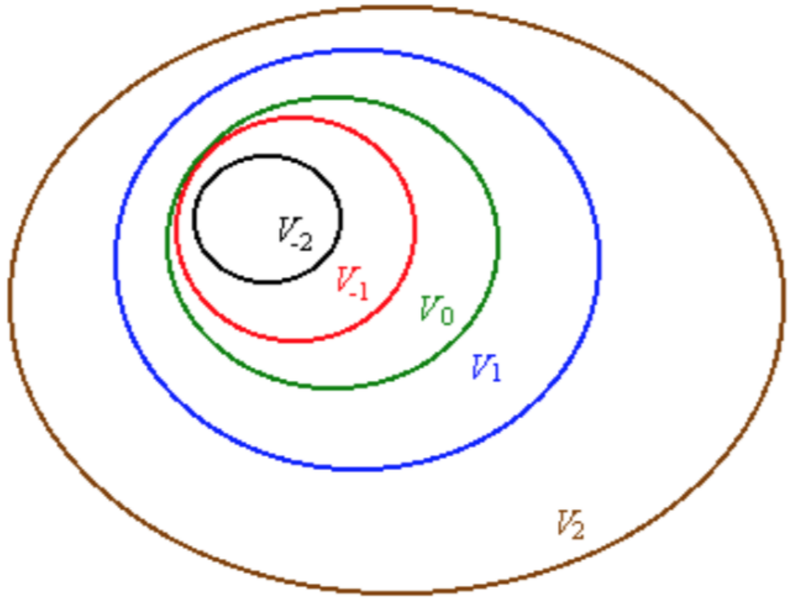


Can every function be represented as a combination of Haar wavelets?

Recall that V_j is the space of square integrable functions that are constant on dyadic intervals of length 2^{-j}. The elements of this space can be written as $\sum_{k\in\mathbb{Z}} a_k \phi(2^j t - k)$. If j is negative, the intervals are of length greater than 1. Concretely, V_{-1} contains functions constant on intervals of length 2; V_{-2} contains functions constant on intervals of length 4, etc.

We will next list the properties of $\{V_j\}$:

1. If a function is piecewise constant on integers then it is piecewise constant on half integers. Therefore, $...V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset V_3...$



2. It can be shown that $\cap_n V_n = \{0\}$ (the constant zero function).

3. Every function can be approximated by a staircase function arbitrarily well if the side of the stairs is small enough. Therefore, $\cup_n V_n$ is dense in the space of square integrable functions.

4. Take a function that is constant on all intervals of length 2^{-n}. Shrink it by a factor of 2. The result is a function that is constant on intervals of length 2^{-n-1}. Therefore, if $f(t) \in V_n$, then $f(2t) \in V_{n+1}$.

5. Translating a function by an integer does not change the fact that it is constant on integer intervals. Therefore, if $f(t) \in V_0$ then $f(t-k) \in V_0$.

6. The family of functions:
$$\phi_{0k} = \phi(t-k),\ k \in \mathbb{Z}$$
forms an orthogonal basis for V_0. The function ϕ is called the *scaling function*.

Definition 1. *A sequence of spaces $\{V_j\}_{j\in\mathbb{Z}}$ together with the scaling function ϕ that generates V_0 so that (1)-(6) are satisfied is called the* multiresolution analysis*.*

Definition 2. *Assume M_1 and M_2 are orthogonal subspaces, i.e., $w_1 \perp w_2$ for all $w_1 \in M_1$ and $w_2 \in M_2$. The subspace V is the* orthogonal direct sum *of M_1 and M_2, denoted as $V = M_1 \oplus M_2$, if every $v \in V$ can be written uniquely as*
$$v = w_1 + w_2$$
with $w_1 \in M_1$ and $w_2 \in M_2$.

Now consider again our hierarchy of subspaces
$$\ldots\ V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset V_3\ \ldots$$

Since $V_0 \subset V_1$, there is a subspace W_0 such that $V_0 \oplus W_0 = V_1$, we denote this subspace $W_0 = V_1 \ominus V_2$. Similarly define $W_1 = V_2 \ominus V_1$ and, in general, $W_{j-1} = V_j \ominus V_{j-1}$.

We have:

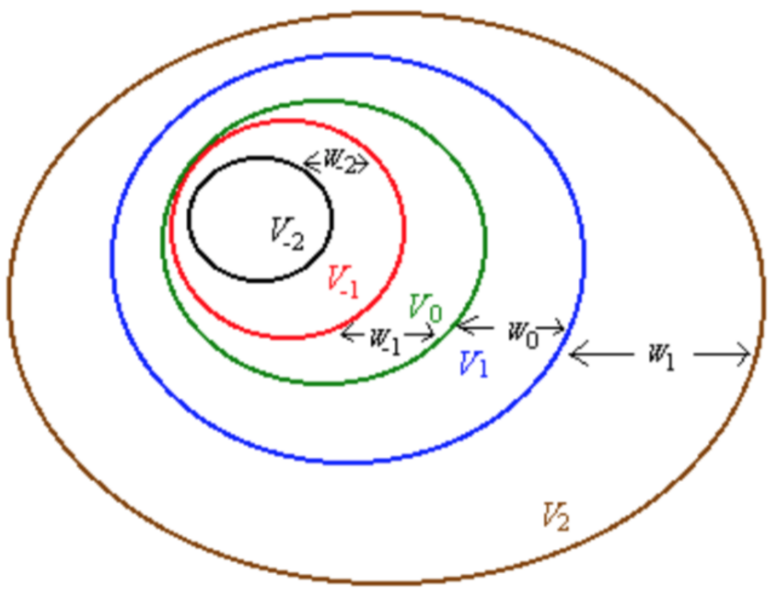
$$\begin{aligned} V_3 &= V_2 \oplus W_2 \\ &= V_1 \oplus W_1 \oplus W_2 \\ &= V_0 \oplus W_0 \oplus W_1 \oplus W_2 \\ &= V_{-1} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \\ &= \dots \end{aligned}$$

and therefore for $v_3 \in V_3$:

$$\begin{aligned} v_3 &= v_2 + w_2 \\ &= v_1 + w_1 + w_2 \\ &= v_0 + w_0 + w_1 + w_2 \\ &= v_{-1} + w_{-1} + w_0 + w_1 + w_2 \\ &= \dots \end{aligned}$$

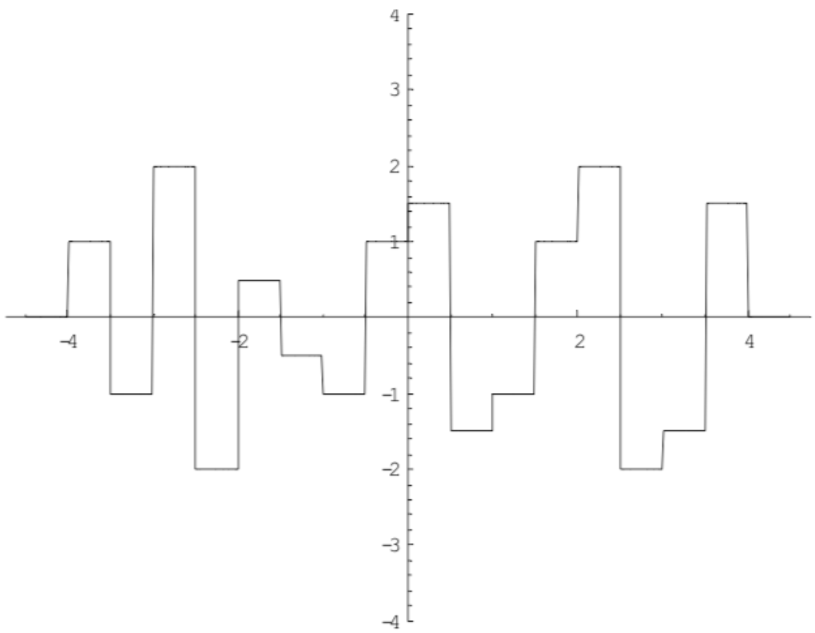
with $v_i \in V_i$ and $w_i \in W_i$.

In conclusion, the relationship between $\{V_j\}$ and $\{W_j\}$ looks like this:



How can we characterize the W_j spaces?

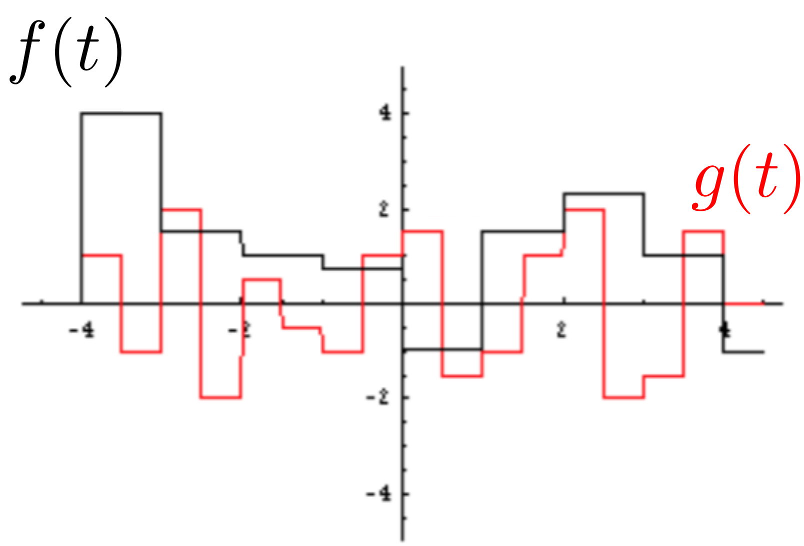
Claim 3. *W_0 is the set of functions that are constant on half integers and take equal and opposite values on half of each integer interval. For example, here is an element from W_0:*



Proof. Let A be the the set of functions that are constant on half integers and take equal and opposite values on half of each integer interval. We need to show that $W_0 = A$. First, show that V_0 and A are orthogonal. Second, show that $V_0 \oplus A = V_1$. Then it will follow that:

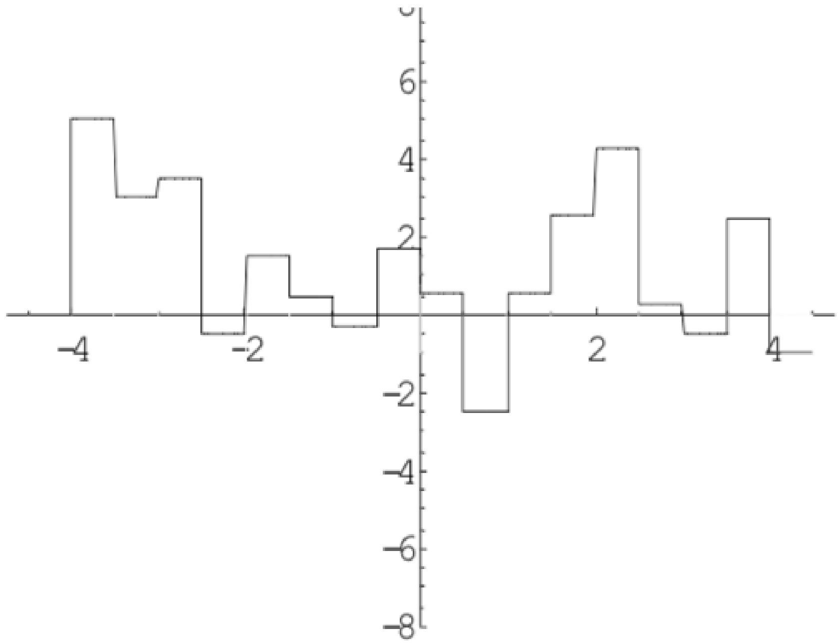
$$A = V_1 \ominus V_0 \equiv W_0.$$

- First, take $f \in V_0$ and $g \in A$, these functions look like this:

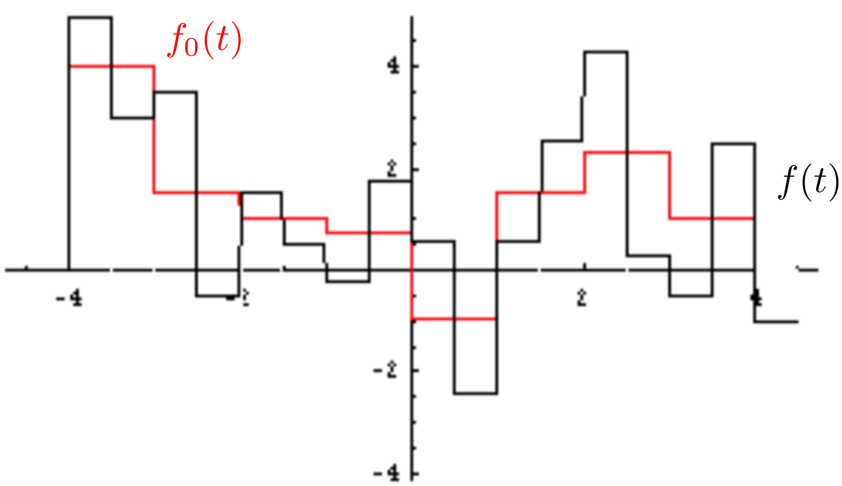


Thus it can be directly verified that $\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t)g(t)dt = 0$ because $f(t)g(t)$ takes on equal and opposite values on each half of every integer interval, and so integrates to 0 on each integer interval.

- Second, for every $f \in V_1$ there exist $f_0 \in V_0$ and $g_0 \in A$ such that $f = f_0 + g_0$. Indeed, take $f \in V_1$. It is constant on the half integer intervals:



Define f_0 to be the function that is constant on each integer interval and whose values is the average of two values of f on that interval:



Then, f_0 is constant on integer intervals, and so $f_0 \in V_0$. Now define $g_0(t) = f(t) - f_0(t)$. Clearly g_0 takes on equal and opposite values on each half of every integer interval, and so $g_0(x) \in A$.

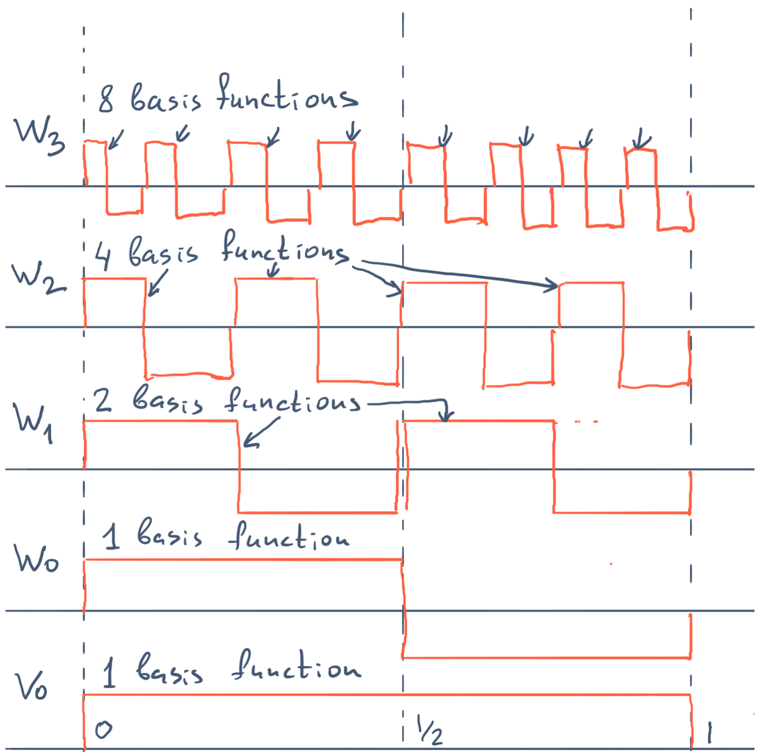
To summarize we have

$$f(x) = f_0(x) + g_0(x)$$

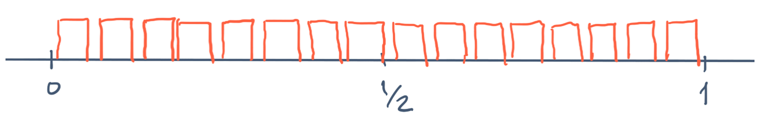
where $f_0 \in V_0$ and $g_0 \in A$. Therefore, $V_1 = V_0 \oplus A \Rightarrow A = W_0$. $\square$

Similarly we can show that W_j is the space of square integrable functions that take on equal and opposite values on each half of the dyadic interval of length 2^{-j}.

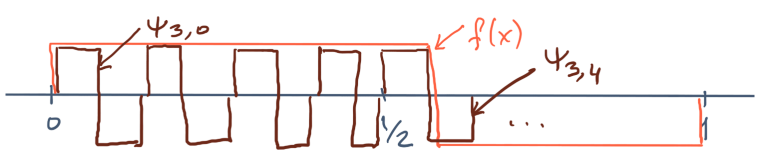
To summarize we designed the following structure:



Every function with $\frac{1}{16}$ quantization:



can be decomposed into 8 functions in W_3, plus $8 = 4 + 2 + 1 + 1$ functions in the lower layers. Typically the representation will be sparse. For example, consider the function we started with:



The representation is sparse because most of the coefficients are zero at high level of details. Concretely, consider $\psi_{3,i} \in W_3$. Note $\langle f, \psi_{3,i} \rangle = 0$ for all $i \neq 3$. Only $\langle f, \psi_{3,i} \rangle \neq 0$. Therefore, one singularity in the function only affects a few selected coefficients, unlike in the Fourier transform case.