8.5 WAVELET-BASED IMAGE CODING

- We begin with 1D wavelets.
- Recall the Fourier Transform:
  - ▶ Our signal $x(t)$ comes from a vector space of allowable signals.
  - ▶ The set of signals $e^{j\omega t}$ (for all real ω) is a basis for the vector space.
  - ▶ The objective is to write $x(t)$ as an (uncountable) linear combination of the basis signals.
  - ▶ To find the "Fourier coefficients" in this linear combination, we take dot products between our signal $x(t)$ and the basis signals:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt.$$

- Usually with wavelets, the signals $x(t)$ we are interested in come from a space called $L^2(\mathbb{R})$; it is the space of all square-integrable signals.
- For suitable "mother wavelet" signals $\psi(t)$, a basis for this space can be generated by dilating and translating $\psi(t)$.

- The signal $\psi(t-u)$ is a shifted version of $\psi(t)$. It is called a TRANSLATE of $\psi(t)$.
- The signal $\psi_a(t) = \sqrt{a}\psi(ax)$ is a stretched version of $\psi(t)$; it is called a DILATE of $\psi(t)$.
- The signal $\psi_a(t - \frac{u}{a}) = \sqrt{a}\psi(ax - u)$ has BOTH dilation and translation.
- A suitably constructed set of translates and dilates of a suitable mother wavelet will form a basis for $L^2(\mathbb{R})$.
- Then, we can write a signal $x(t)$ from $L^2(\mathbb{R})$ as a linear combination of the wavelets in the basis.
- To find the coefficients in the linear combination, we take dot products between $x(t)$ and the wavelets:

$$Wx(a, a^{-1}u) = \int_{-\infty}^{\infty} x(t)\psi_a^*(t - a^{-1}u)dt.$$

- The set of coefficients $Wx(a, a^{-1}u)$ is called the WAVELET TRANSFORM of $x(t)$.

- For the wavelet transform to be invertible, it is required that

$$\int_{-\infty}^{\infty} |\Psi(\omega)|^2 \, \frac{d\omega}{\omega} < \infty,$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$.

  - This implies that $\Psi(0) = 0$ and that, as $\omega \to 0$, $|\Psi(\omega)|^2 \to 0$ faster than $1/\omega \to \infty$.
  - In other words, $\psi(t)$ is like the impulse response of a high-pass filter (it oscillates).

- If $\psi(t)$ is a good mother wavelet, then it is only nonzero on a set that is closed and bounded (a "compact set").

  - This means that $\psi(t)$ is localized in time.

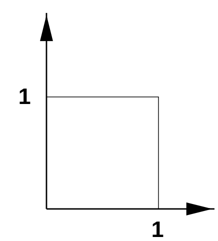
- Thus, a LARGE coefficient in the wavelet transform tells us that there is a great deal of similarity between $x(t)$ and a certain dilated, translated wavelet.

  - This tells us that a certain oscillation was present in $x(t)$, and it also tells us WHEN that oscillation was present.

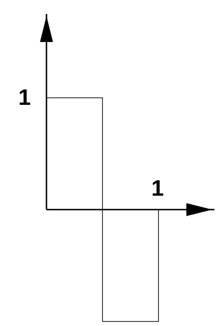
- NOTE: the Fourier Transform NEVER tells you anything about WHEN a certain frequency was present.

The Scaling Function

- Associated with the mother wavelet $\psi(t)$, there is a SCALING FUNCTION $\phi(t)$.
  - Actually, in rigorous mathematical theory, you start with $\phi(t)$ and derive $\psi(t)$ from it.
- Whereas the wavelet $\psi(t)$ would be the impulse response of a HIGH PASS filter, the scaling function $\phi(t)$ would be the impulse response of a LOW PASS filter.
- Example: Haar wavelet (also, the simplest Daubechies wavelet).
  - Scaling function $\phi(t)$:



  - Wavelet $\psi(t)$:



- The scaling function satisfies a "two-scale dilation equation":

$$\phi(t) = \sum_k h(k)\phi(2t + k),$$

where $h(k)$, $0 \leq k \leq M - 1$, is the length-M unit pulse response of a low-pass digital filter.

  ▶ This says that the scaling function at one scale can be written as a linear combination of the TRANSLATES of the scaling function at the next higher scale.

- The mother wavelet $\psi(t)$ also satisfies a similar equation:

$$\psi(t) = 2\sum_k g(k)\phi(2t - k),$$

where $g(k) = (-1)^k h(M - k - 1)$, $0 \leq k \leq M - 1$, is the length-M unit pulse response of a high-pass digital filter.

- The filters H and G with unit pulse responses $h(k)$ and $g(k)$ have a special relationship. They are called QUADRATURE MIRROR FILTERS.

- Let $\phi_{2^j}(t) = 2^j \phi(2^j t)$ be a dilation of $\phi(t)$ at the scale 2^j.

- If $\phi(t)$ is a "good" scaling function, then the set of translates $\left\{2^{-\frac{j}{2}}\phi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$ is an orthonormal basis for a signal SUBSPACE V_{2^j}.

- Loosely, you might think of V_{2^j} as the space of all signals in $L^2(\mathbb{R})$ that can be exactly represented with 2^j samples per unit in time.

- Then $V_{2^{j-1}}$ is the space of all signals that can be represented with half as many samples.

- Clearly, if $x(t) \in V_{2^{j-1}}$, then $x(t) \in V_{2^j}$, so $V_{2^{j-1}} \subset V_{2^j}$.

- A signal $x(t)$ in V_{2^j} can be placed into one of two categories:
  - It might vary slowly enough to be in $V_{2^{j-1}}$,
  - or it might not lie in $V_{2^{j-1}}$.

- The signals $x(t) \in V_{2^j}$ that are not in $V_{2^{j-1}}$ form a signal subspace called $O_{2^{j-1}}$.

- Thus, every signal in V_{2^j} is either in $V_{2^{j-1}}$ or in $O_{2^{j-1}}$:

$$V_{2^j} = V_{2^{j-1}} \oplus O_{2^{j-1}}.$$

- This is like a subband decomposition. A signal $x(t) \in V_{2^j}$ can be decomposed into a part that lies in $V_{2^{j-1}}$ and a part that lies in $O_{2^{j-1}}$. The latter is called the "detail signal".

- FACT: the set of translates $\left\{2^{-\frac{j}{2}}\psi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$ of $\psi(t)$ at scale 2^j form an orthonormal basis for the subspace O_{2^j}.

- For a signal $x(t) \in V_{2^{j+1}}$, this gives us a way to perform a subband decomposition of $x(t)$.
  - We break it into a low-pass part that lies in V_{2^j} by taking dot products with the basis $\left\{2^{-\frac{j}{2}}\phi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$,
  - and a high-pass part that lies in O_{2^j} by taking dot products with the basis $\left\{2^{-\frac{j}{2}}\psi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$.

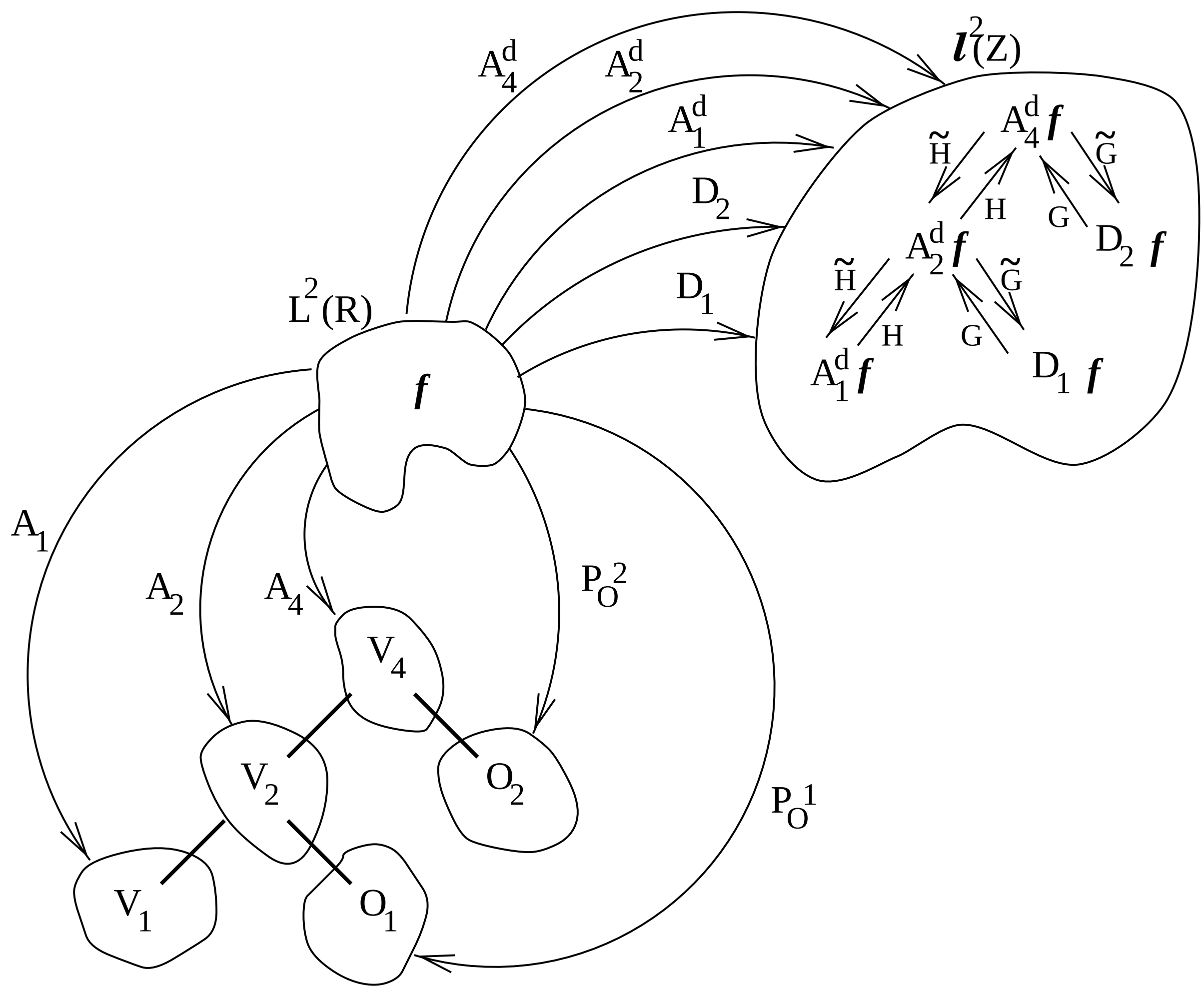
- FACT: for all integer values of j, the subspaces O_{2^j} are all disjoint. Their union is exactly $L^2(\mathbb{R})$.

- Thus, any signal $x(t)$ in $L^2(\mathbb{R})$ can be represented in terms of its projections into all the detail spaces O_{2^j}.

- Together, the set of wavelet basis functions for all the spaces O_{2^j} are an orthonormal basis for $L^2(\mathbb{R})$.

- The dot product of a signal with all of these wavelets is the DYADIC WAVELET TRANSFORM.

- Pictorially,



Mallat Algorithm

- FACT: if $x(t) \in V_{2^j}$ and the coefficients of $x(t)$ with respect to the basis $\left\{2^{-\frac{j}{2}}\phi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$ are known, then
  - ▶ The coefficients of the projection of $x(t)$ into $V_{2^{j-1}}$ can be found by filtering the coefficients in V_{2^j} with a low-pass filter $\widetilde{h}(k) = h(M-k-1)$ and dropping every other sample from the result (downsampling).
  - ▶ The coefficients of the projection of $x(t)$ into $O_{2^{j-1}}$ can be found by filtering the coefficients in V_{2^j} with a high-pass filter $\widetilde{g}(k) = g(M-k-1)$ and dropping every other sample from the result (downsampling).
- FACT: the coefficients of $x(t)$ in V_{2^j} can be recovered from the coefficients in $V_{2^{j-1}}$ and in $O_{2^{j-1}}$:
  1. Insert zeros between each coefficient in $V_{2^{j-1}}$ and each coefficient in $O_{2^{j-1}}$ (upsample).
  2. Filter the upsampled coefficients from $V_{2^{j-1}}$ with the filter $h(k)$ and filter the coefficients from $O_{2^{j-1}}$ with $g(k)$. Add the resulting sequences. This gives the coefficients of $x(t)$ in V_{2^j}.

How it Works

- To apply the algorithm, begin with a discrete signal $y(k)$.
- For some signal $x(t)$ in V_{2^j}, ASSUME that the signal $y(k)$ contains the coefficients of $x(t)$ in V_{2^j} with respect to the basis $\left\{2^{-\frac{j}{2}}\phi_{2^j}(t-2^{-j}k)\right\}_{k\in\mathbb{Z}}$.
- Repeatedly apply the filters $\widetilde{h}(k)$ and $\widetilde{g}(k)$ and downsample each result to project $x(t)$ down some number of scales.
- If we apply the filters three times, this gives us a subband decomposition of our discrete signal $y(k)$ as the wavelet coefficients of $x(t)$ in $O_{2^{j-1}}$, $O_{2^{j-2}}$, and $O_{2^{j-3}}$ plus the scaling function coefficients in $V_{2^{j-3}}$.
  - ▶ Because of the downsampling at each stage, the number of samples in this representation is the same as the number of samples in the original signal $y(k)$.
- The original signal $y(k)$ can be recovered from the representation by repeatedly upsampling, applying the filters $h(k)$ and $g(k)$, and adding the results.

What is it Good For?

- Often, many of the wavelet coefficients will be negligibly small or zero.
- Then, if the wavelet coefficients are quantized and coded (by entropy coding, for example), we can COMPRESS our original discrete signal $y(k)$.
  - ▶ If there is no quantization, then this gives a lossless code.
- The original signal can be recovered (approximately) from the quantized and entropy coded representation.

Example: Haar Wavelet

- $h(k) = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\}$. $(M = 2)$.
- $g(k) = \left\{ \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right\}$.
- $\widetilde{h}(0) = h(2 - 0 - 1) = h(1)$.
- $\widetilde{h}(1) = h(2 - 1 - 1) = h(0)$.
- $\widetilde{g}(0) = g(2 - 0 - 1) = g(1)$.
- $\widetilde{g}(1) = g(2 - 1 - 1) = g(0)$.

Doing it in 2D

- We will consider the SEPARABLE case only.
- Suppose that $\phi(t)$ is a 1D scaling function and $\psi(t)$ is the associated 1D mother wavelet.
- We define the 2D scaling function by

$$\phi(x, y) = \phi(x)\phi(y).$$

- THREE 2D wavelets are associated with this scaling function:

$$\begin{aligned} {}_1\psi(x, y) &= \phi(x)\psi(y) \\ {}_2\psi(x, y) &= \psi(x)\phi(y) \\ {}_3\psi(x, y) &= \psi(x)\psi(y) \end{aligned}$$

- 2D dilation is defined by applying the SAME scaling factor to the x and y coordinates.
- 2D translation is defined by applying INDEPENDENT translations to the x and y coordinates.

2D Mallat Algorithm

- Begin with an $N \times N$ image $I(i,j)$, and assume that it represents scaling function coefficients (dot products with $\phi(x,y)$) for some function $J(x,y) \in L^2(\mathbb{R}^2)$.

- Apply the filters $\widetilde{h}(k)$ and $\widetilde{g}(k)$ to the rows of $I(i,j)$ and discard every other sample (horizontal downsampling). This gives two horizontally downsampled images.

- Apply $\widetilde{h}(k)$ and $\widetilde{g}(k)$ to the columns of each horizontally downsampled images. This gives four result images. discard every other sample from each column of these four images (vertical downsampling).

- The result is four $\frac{N}{2} \times \frac{N}{2}$ images; we will call them LL, LH, HL, and HH.

- Repeat by applying the procedure to LL.

* Pictorially,