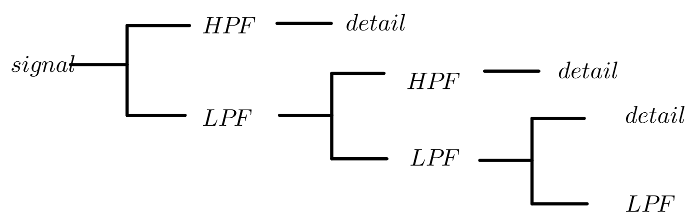
EE5585 Data Compression	April 23, 2013
Lecture 24	
Instructor: Arya Mazumdar	*Scribe: Yuanhang Wang*

Recap - Sub-band Coding

In the last class we talked about sub-band coding. In the first step of it, signal goes into two parts, high frequency component (details, local property) and low frequency component (average values, global property). Then after several LPFs we will end with the average value of this signal. The idea is that we don't need some of the details that comes from HPFs, so we throw them out and do the data compression.



One possible way to achieve this is doing the DFT in each step. The main problem of DFT is that we may not have a single matrix to achieve the whole steps. Suppose we have some changes in time domain. Then we cannot know how it changes from observing frequency domain. According to *uncertainty principle* in Fourier analysis, a function f and its Fourier transform $\mathcal{F}$ cannot be simultaneously well localized. To take care of that, we need something else called *wavelets*.

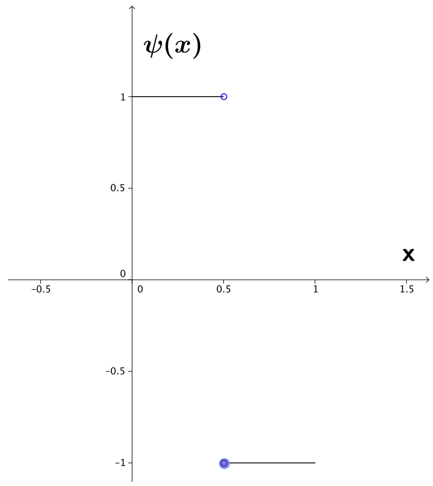
Wavelets

Wavelet is also one of data transform coding techniques. It is a kind of multi-resolution coding.

Let's look at the Haar Wavelet first.

Definition 1 (Mother function of Haar wavelet $\psi(x)$)

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1, & \frac{1}{2} \le x < 1 \\ 0, & otherwise \end{cases}$$



Now our idea is to find some basis space for any function f.

Here let's consider the space of all square-integrable real-valued functions H. A function is said to be *square-integrable* if

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty \tag{1}$$

Definition 2 (Hilbert Space) *A Hilbert space H is an inner product space that is also a complete metric space with respect to the distance function induced by the inner product.*

In space H, two square-integrable real-valued functions f and g have an *inner product*, which is defined as

$$< f, g >= \int f(x)g(x)dx \tag{2}$$

Norm is defined as $\|f\|$, where

$$\|f\|^2 = \int |f(x)|^2 dx =< f, f > \tag{3}$$

Our idea is to change the basis of $f(x)$ and represent it in other basis, say, $g(x)$. Then the expression will be:

$$f(x) = \sum_k C_k g_k(x), k = 0, 1, 2, ...$$

Also we have the concept of *Lebesgue Space*: The space of all measurable functions f on the interval [T_1, T_2] satisfying

$$\int_{T_1}^{T_2} |f(x)|^2 dx < \infty$$

is called the L^2 space, $L^2[T_1,T_2]$.

Let's define

$$\psi_{0,0}(x) = \psi(x)$$

and

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k), where\, j \in \mathbb{N}, k \in \mathbb{Z}.$$

It's clear that

$$\psi_{j,k}(x) = \begin{cases} 2^{\frac{j}{2}}, & \frac{k}{2^j} \leq x < \frac{k+\frac{1}{2}}{2^j} \\ \\ 2^{\frac{-j}{2}}, & \frac{k+\frac{1}{2}}{2^j} \leq x < \frac{k+1}{2^j} \end{cases}$$

Here the set $\{\psi_{j,k}, j \in \mathbb{N}, k \in \mathbb{Z}\}$ is an *orthonormal basis* in $L^2[\mathbb{R}]$.

Let's verify that this basis is orthonormal. We can notice that:

$$\int_{-\infty}^{+\infty} |\psi_{j,k}(x)|^2 dx = \int_{-\infty}^{+\infty} (2^{\frac{j}{2}})^2 |\psi(2^j x - k)|^2 dx = \int_{\frac{k}{2^j}}^{\frac{k+1}{2^j}} 2^j dx = 2^j \frac{1}{2^j} = 1$$

Also it's easy to prove that:

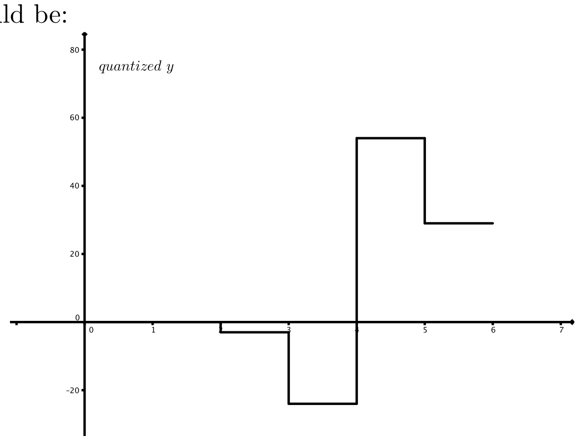
$$\int_{-\infty}^{+\infty} \psi_{j,k}(x)\psi_{j',k'}(x)dx = 0, if\, j \neq j' and\, k \neq k'$$

So $\{\psi_{j,k}, j \in \mathbb{N}, k \in \mathbb{Z}\}$ forms an orthonormal basis.

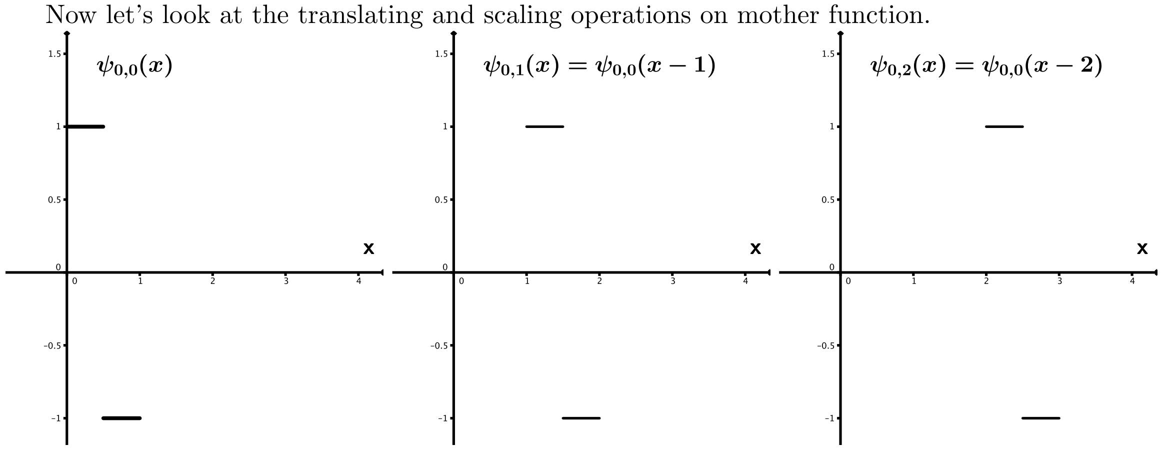
Now let's try to use this basis to represent function $f(x)$.

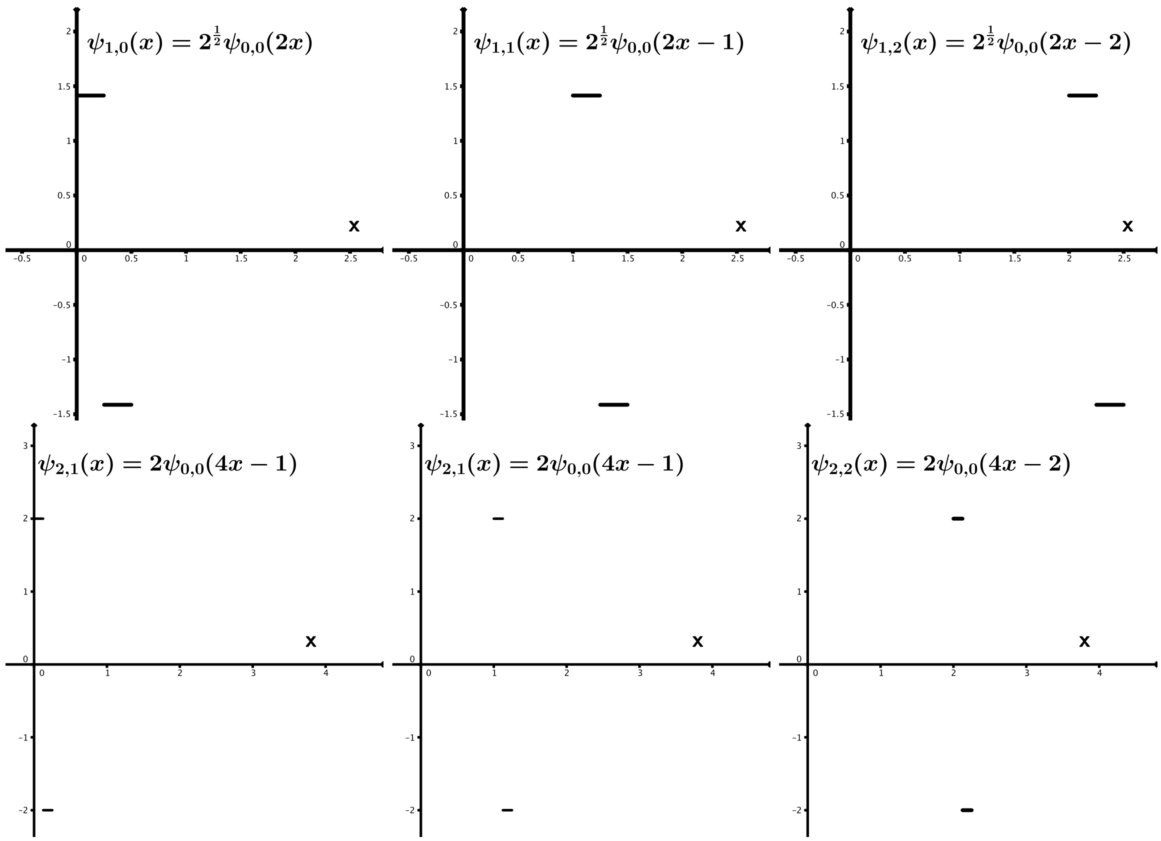
$$f(x) = \sum_{j \geq 0,\, k \in \mathbb{Z}} c_{j,k}\psi_{j,k}(x)$$

In practice, we take discrete points of $f(x)$ and use $\psi(x)$ to represent it. For example, y= [0 -3 -24 54 29 ...], then our graph would be:



Now let's look at the translating and scaling operations on mother function.





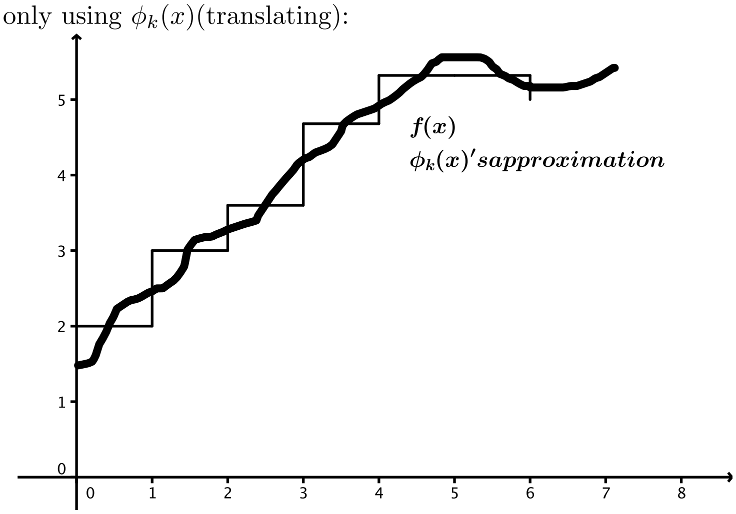
The rows show how translating works, and the columns show how scaling works.

Definition 3 (Scaling function of Haar wavelet $\phi(x)$)

$$\phi(x) = \begin{cases} 1, & 0 \leq x < 1 \\ \\ 0, & otherwise \end{cases}$$

We use $\phi_k(x)$ or $\phi_{0,k}(x)$ to denote $\phi(x-k)$, and $\phi_{1,k}(x) = \phi_{0,k}(2x)$. Here are the steps that we use to approximate an arbitrary function $f(x)$:

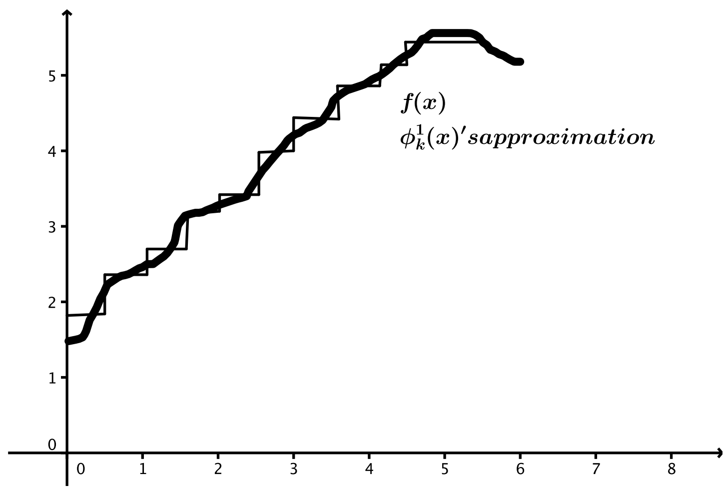
1. Approximate only using $\phi_k(x)$(translating):



The approximation $\phi_f^0(x) = \sum C_{0,k}\phi(x-k)$, where $C_{0,k} = \int f(x)\phi(x-k)dx = \int_k^{k+1} f(x)dx$. $C_{0,k}$ is the average value of $f(x)$ in each interval.

2. Based on step one, we do the scaling:

$$\phi_{1,k}(x) = \phi(2x-k) = \begin{cases} 1, & \frac{k}{2} \leq x < \frac{k+1}{2} \\ 0, & otherwise \end{cases}$$

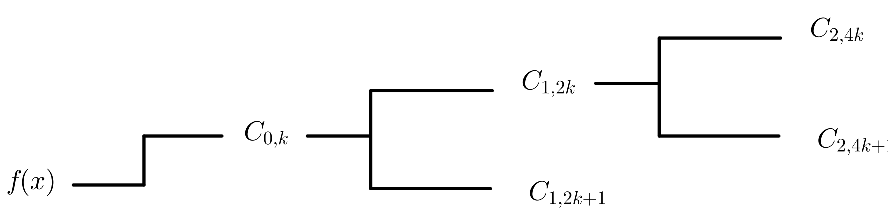


This approximation $\phi_f^1(x) = \sum C_{1,k}\phi_{1,k}(x)$, where $C_{1,k} = 2\int f(x)\phi_{1,k}(x)dx = 2\int_{\frac{k}{2}}^{\frac{k+1}{2}} f(x)dx$. $C_{1,k}$ is still the average value of $f(x)$ in each smaller interval.

We can see that:

$$C_{1,2k} + C_{1,2k+1} = 2\int_k^{k+\frac{1}{2}} f(x)dx + 2\int_{k+\frac{1}{2}}^{k+1} f(x)dx = 2\int_k^{k+1} f(x)dx = C_{0,k}$$

So we may apply this property to construct a kind of sub-band coding like:



This is a simple wavelet called Haar wavelet, and eventually it gives you a matrix of transformation.

There are also other complex wavelet like *Daubechies Wavelets*. They all belong to *MRA (Multiresolution Analysis)*, and ideas are similar (have a mother fuction $\psi(x)$ and approximate function using $\psi(2^j x - k)$).

Fixed to Fixed Almost Lossless Coding

Let's look at our lossy coding techniques:

$$lossy\,coding\,techniques \begin{cases} transform\,coding \\ rate-distortion\,theory \\ quantization \\ sub-bandcoding \end{cases}$$

In all of these cases, especially in transform coding, we are taking a set of vectors and transform it to another set of vector. So it is from fixed length message to fixed length codewords. No matter what the message is the length of codewords are the same.

But when we do lossless coding:

$$lossless\,coding\,techniques \begin{cases} Huffman\,code \\ Lempel\ Ziv \end{cases}$$

They are all fixed to variable length codes. For example, when we do Huffman coding, if $\mathcal{X} = \{a, b, c, d\}$, our codewords can be $a = 0, b = 10, c = 110, d = 111$, which are not fixed.

Let assume $\mathcal{X}$ is a source alphabet, say {0,1}, and $x \in \mathcal{X}^n$. Can we uniformly map x to a k length string? Is there a way to do lossless coding in this case, which is fixed length to fixed length?

The answer is that we cannot do exact lossless coding but we can do something almost lossless. So that requires our decoder work with high probability.

Suppose we have a binary sequence of length n:

$$x = 00...101......$$

Assume each bit has $Pr(0) = 1 - p, Pr(1) = p$ and they are i.i.d.. Then we know:

$$Pr(00111011) = f(number\,of\,1's,\ L) = p^L(1-p)^{n-L}$$

So

$$Pr(a\,sequence\,having\,L\,1's) = \binom{n}{L} p^L (1-p)^{n-L} \approx \frac{2^{nh(\frac{L}{n})}}{\sqrt{n}} 2^{L\log p + (n-L)\log(1-p)}$$

$$= \frac{1}{\sqrt{n}} 2^{n[h(\frac{L}{n}) + \frac{L}{n}\log p + (1-\frac{L}{n})\log(1-p)]}$$

$$= \frac{1}{\sqrt{n}} 2^{n[-\frac{L}{n}\log\frac{L}{np} - (1-\frac{L}{n})\log\frac{(1-\frac{L}{n})}{1-p}]}$$

$$= \frac{1}{\sqrt{n}} 2^{-nD(\frac{L}{n}||p)}$$

We know that $D(p||q) \geq 0$, and it is 0 only when $p = q$.

So when $L \neq np$, $D(\frac{L}{n}||p) > 0$, and $\frac{1}{\sqrt{n}}2^{-nD(\frac{L}{n}||p)}$ will diverge to 0 with $n \to \infty$. It is not 0 only when $L \approx np$, or $n(p-\epsilon) \leq L \leq n(p+\epsilon)$. Then in this case

$$\binom{n}{np} \sim 2^{nh(p)}$$

The size of these sequences is $2^{nh(p)}$, so we only need $nh(p)\,bits$ to keep any of them.

Suppose $k = nh(p)$, then we can have fixed length codes that use only $nh(p)$ bits. Therefore the rate of compression is

$$Rate = \frac{k}{n} = h(p) = entropy$$

Now we compress the data from fixed length to fixed length, but we have probability of error that equals to sequences that does not have np 1's. Even the probability is very small with very large n, those sequences are still there. But our decoder will work with very high probability. This connects lossless coding to lossy coding.