Introduction to Wavelets Analysis

A. Christen

INSTITUTO DE ESTADÍSTICA
PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

based on a presentation from Prof. Bebis

Introduction

- What are wavelets?
- What are they for?
- What are its advantages and disadvantages compared to Fourier analysis?
- What can they be used for, in time series(processes dependent of time)?

Introduction

- What are wavelets?

They are bases of functions in a certain space of functions.

- What are they for?

To approximate functions: a time series, an image, a probability density function, a regression function, etc.

Introduction

- What are its advantages and disadvantages compared to Fourier analysis?

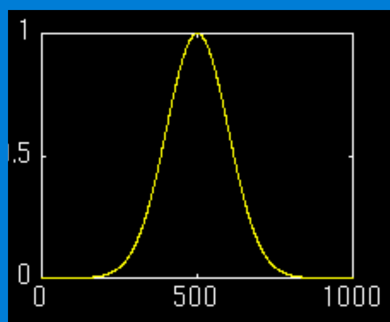
They allow local changes to be detected more efficiently. They are more flexible but do not approximate so well defined sine waves in all the real domain.

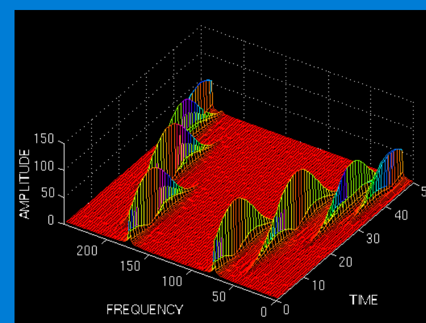
- What can they be used for, in time series (processes dependent of time)?

To estimate the time series, to denoise the time series, to detect change points, among other applications.

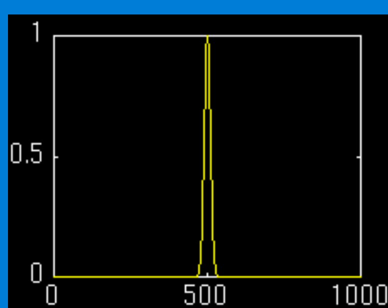
Short Time Fourier Transform - revisited

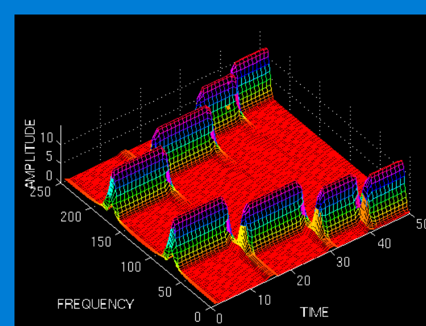
- Time - Frequency localization depends on window size.
  - Wide window → good frequency localization, poor time localization.





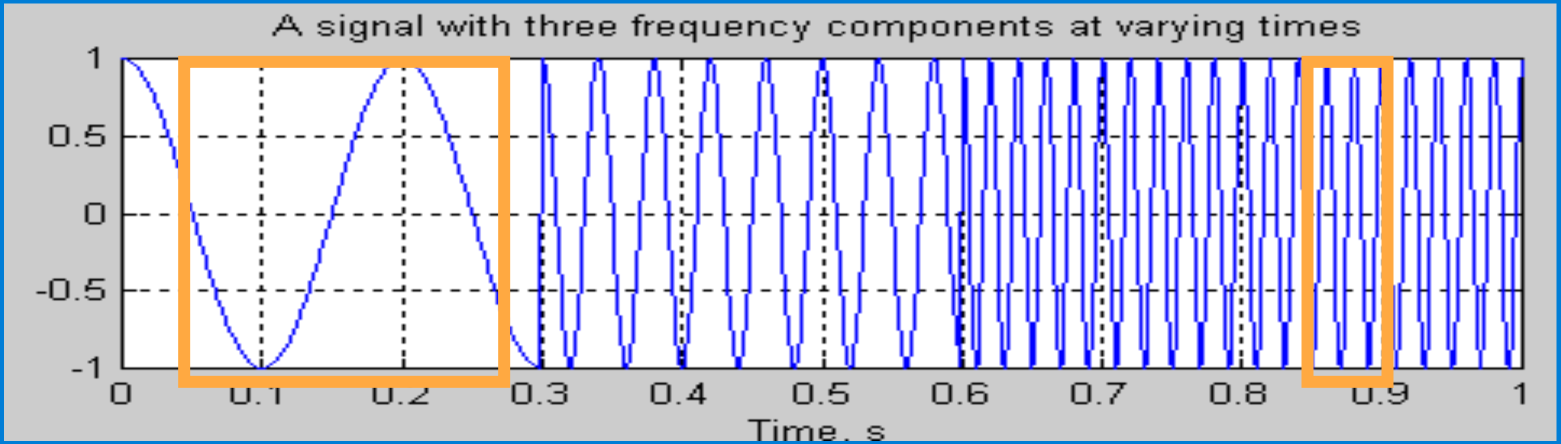
  - Narrow window → good time localization, poor frequency localization.





Wavelet Transform

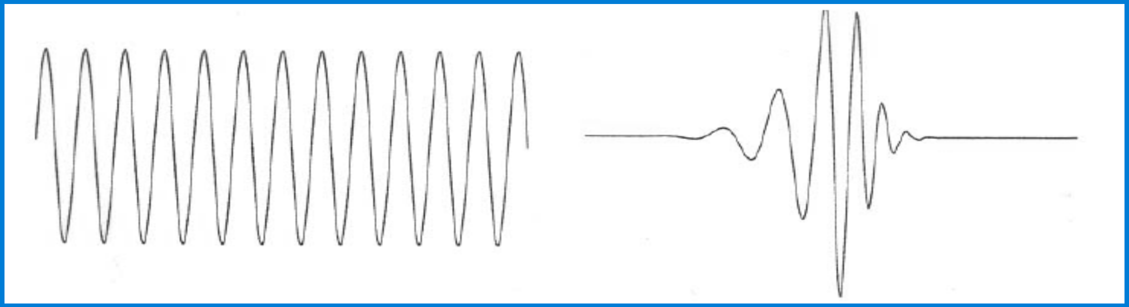
- Uses a variable length window, e.g.:
  - Narrower windows are more appropriate at high frequencies
  - Wider windows are more appropriate at low frequencies



What is a wavelet?

- A function that "waves" above and below the x-axis with the following properties:
  - Varying frequency
  - Limited duration
  - Zero average value
- This is in contrast to sinusoids, used by FT, which have infinite duration and constant frequency.

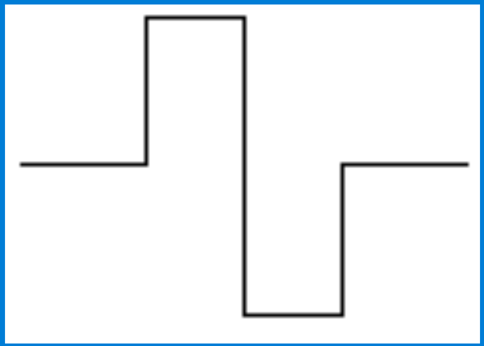
Sinusoid Wavelet



Types of Wavelets

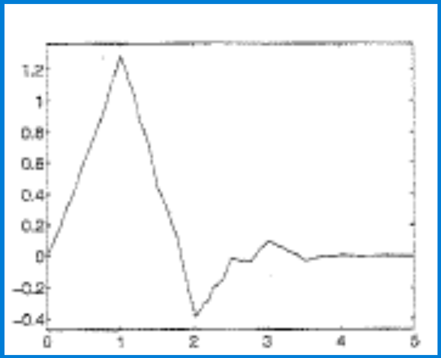
- There are many different families of wavelets, for example:

Haar



Morlet

Daubechies



Basis Functions Using Wavelets

- Like sin( ) and cos( ) functions in the Fourier Transform, wavelets can define a set of basis functions $\psi_k(t)$:

$$f(t) = \sum_k a_k \psi_k(t)$$

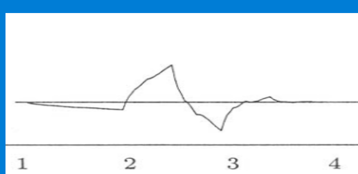
- Span of $\psi_k(t)$: vector space S containing all functions f(t) that can be represented by $\psi_k(t)$.

Basis Construction – "Mother" Wavelet

The basis can be constructed by applying translations and scalings (stretch/compress) on the "mother" wavelet ψ(t):

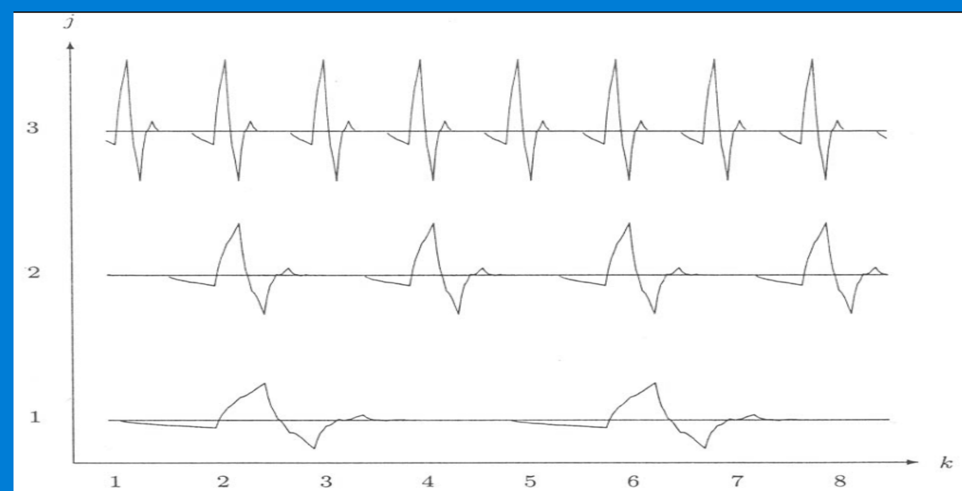
$$\psi(s,\tau,t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right)$$

Example:



ψ(t)

scale



translate

Basis Construction - Mother Wavelet

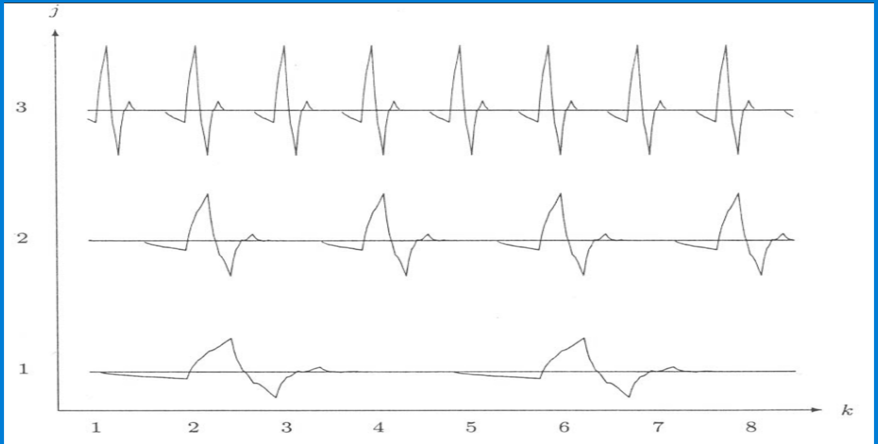
- It is convenient to take special values for s and τ in defining the wavelet basis: $\mathbf{s = 2^{-j}}$ and $\boldsymbol{\tau} \mathbf{= k \cdot 2^{-j}}$

(dyadic/octave grid)

$$\psi(s,\tau,t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) = \frac{1}{\sqrt{2^{-j}}}\psi\left(\frac{t-k.2^{-j}}{2^{-j}}\right) = 2^{\frac{j}{2}}\psi(2^j t - \mathrm{k}) = \psi_{jk}(t)$$

scale $=1/2^j$
(1/frequency)

j ↑



k ⟶

Continuous Wavelet Transform (CWT)

translation parameter
(measure of time)

scale parameter
(measure of frequency)

scale =1/2^{j}
(1/frequency)

Forward CWT:

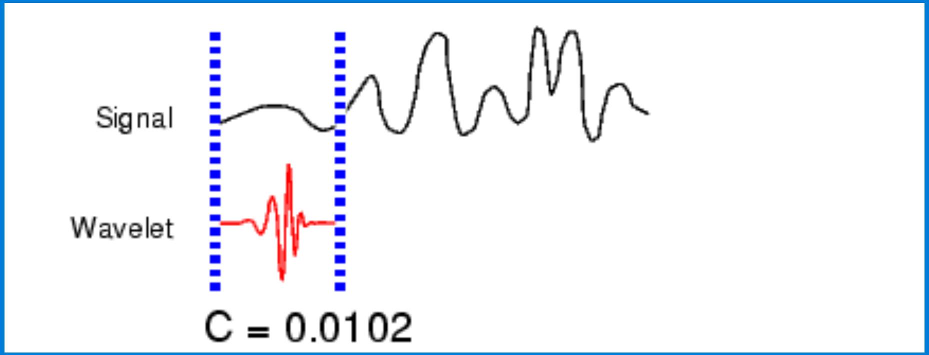
$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t)\,\psi^*\left(\frac{t-\tau}{s}\right) dt$$

normalization constant

mother wavelet (i.e., window function)

Illustrating CWT

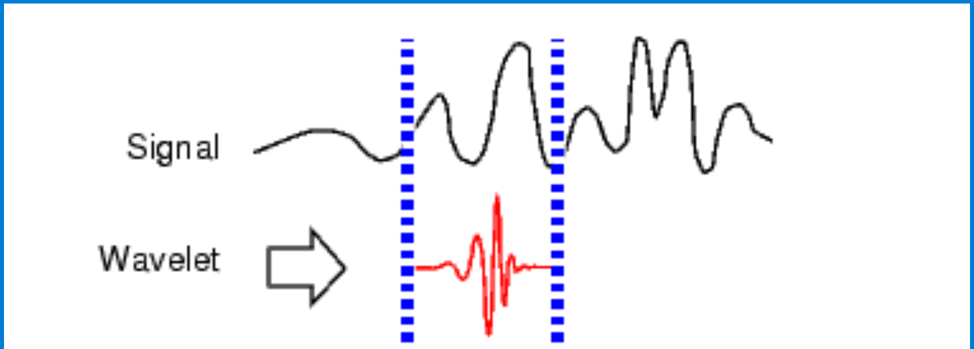
1. Take a wavelet and compare it to a section at the start of the original signal.
2. Calculate a number, C, that represents how closely correlated the wavelet is with this section of the signal. The higher C is, the more the similarity.



$$C(\tau,s)=\frac{1}{\sqrt{s}}\int_t f(t)\,\psi^*\left(\frac{t-\tau}{s}\right)dt$$

Illustrating CWT (cont'd)

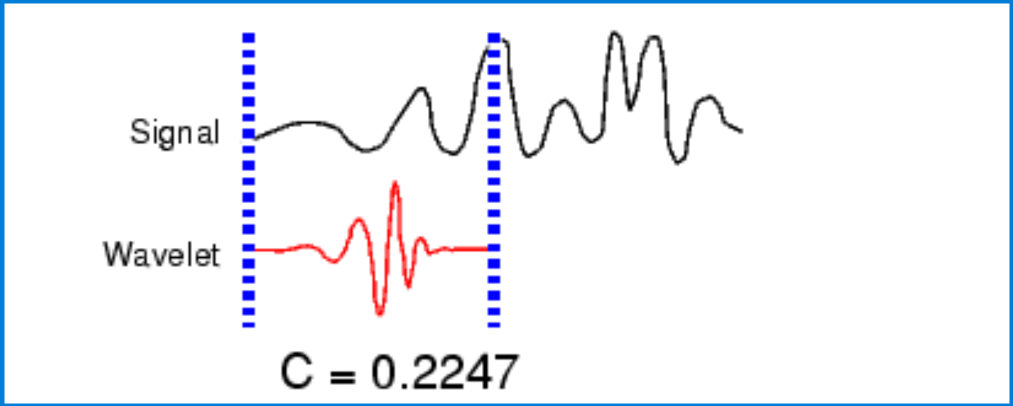
3. Shift the wavelet to the right and repeat step 2 until you've covered the whole signal.



$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt$$

Illustrating CWT (cont'd)

4. Scale the wavelet and go to step 1.

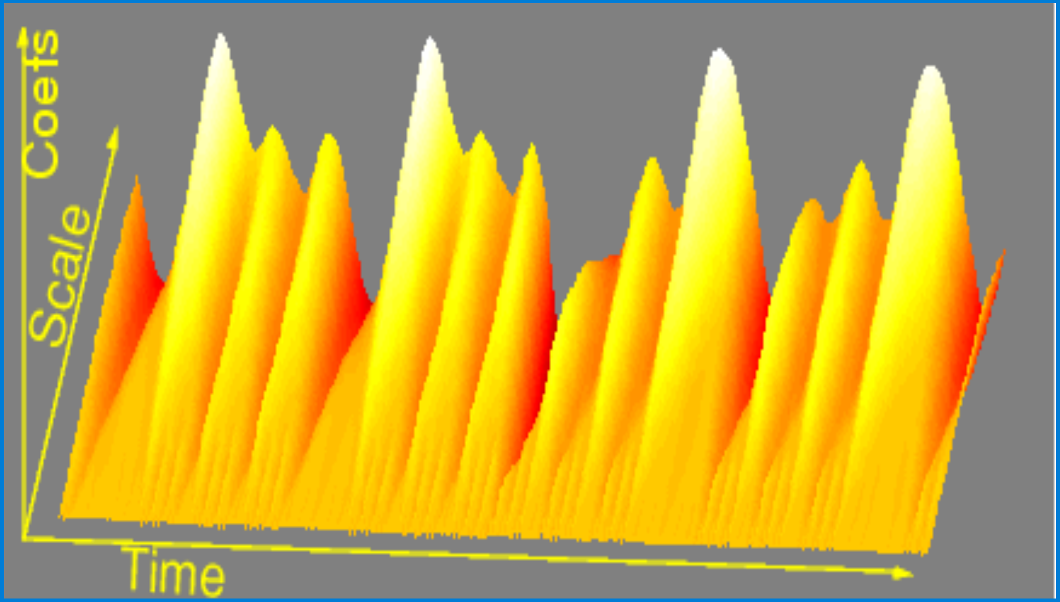


$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t-\tau}{s} \right) dt$$

5. Repeat steps 1 through 4 for all scales.

Visualize CTW Transform

- Wavelet analysis produces a time-scale view of the input signal or image.



$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t-\tau}{s} \right) dt$$

Continuous Wavelet Transform (cont'd)

Forward CWT:

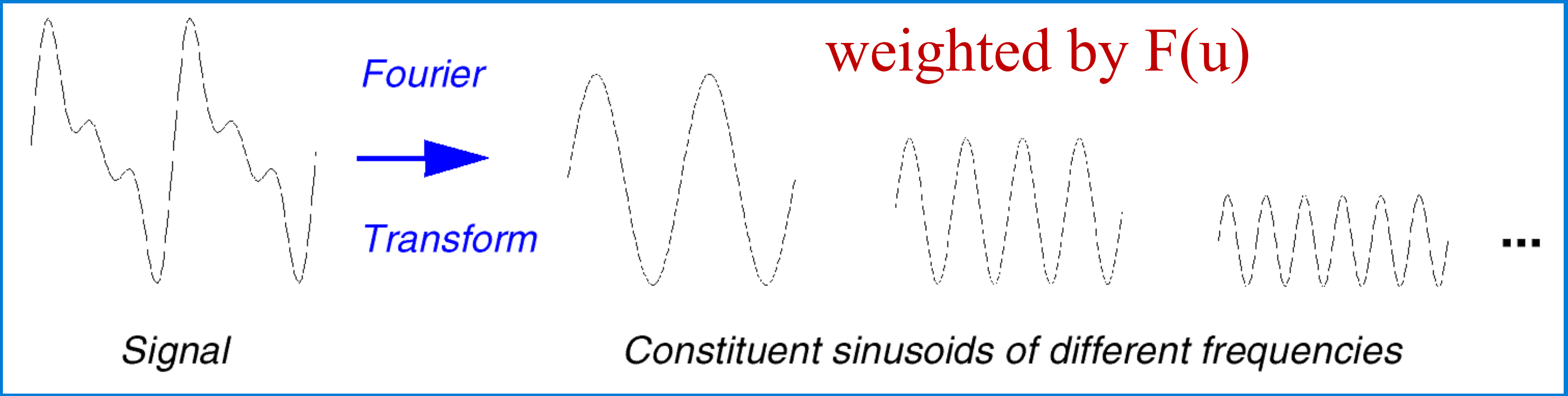
$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t)\, \psi^*\left(\frac{t-\tau}{s}\right) dt$$

Inverse CWT:

$$f(t) = \frac{1}{\sqrt{s}} \int_\tau \int_s C(\tau, s)\, \psi\left(\frac{t-\tau}{s}\right) d\tau\, ds$$

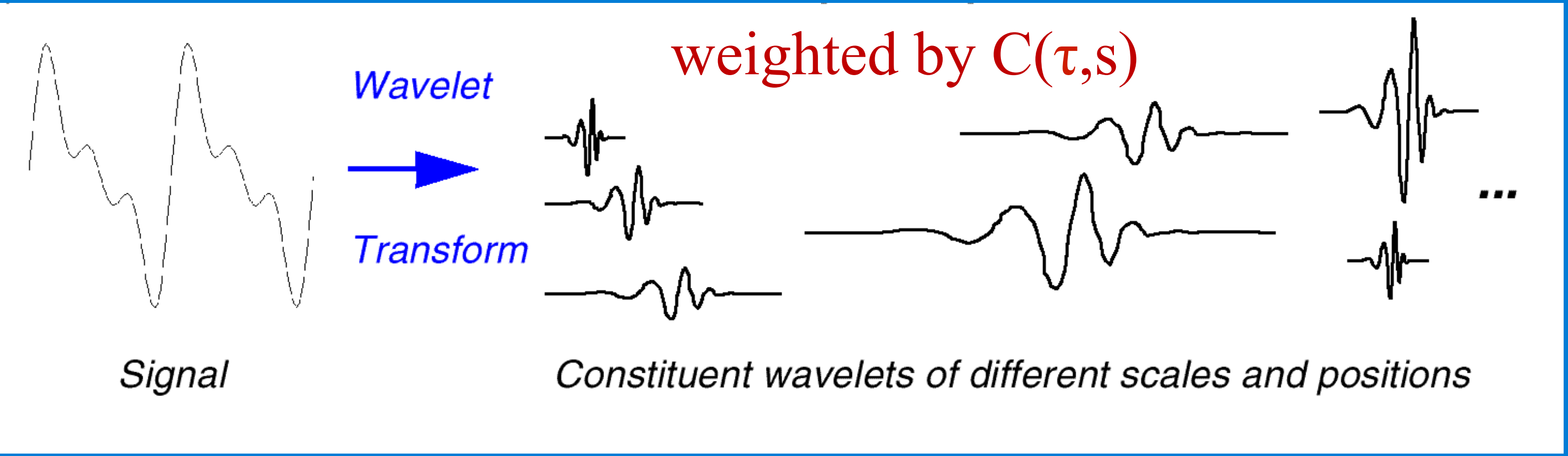
Note the double integral!

Fourier Transform vs Wavelet Transform



$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du$$

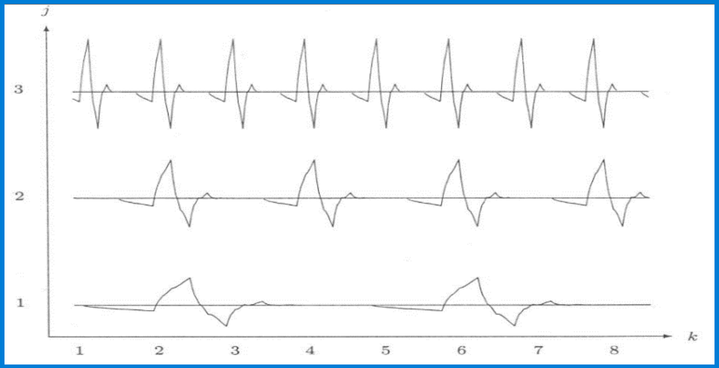
Fourier Transform vs Wavelet Transform

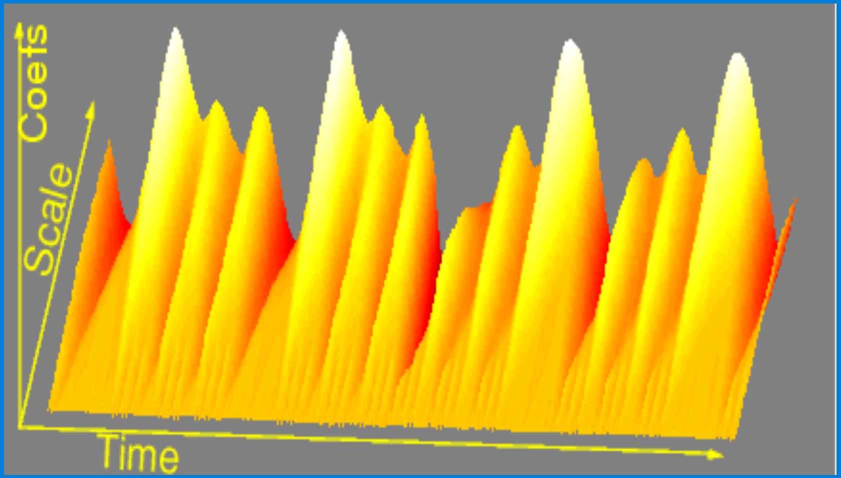


$$f(t)=\frac{1}{\sqrt{s}}\int_{\tau}\int_{s}C(\tau,s)\psi\left(\frac{t-\tau}{s}\right)d\tau ds$$

Properties of Wavelets

- Simultaneous localization in time and scale
  - The location of the wavelet allows to explicitly represent the location of events in time.
  - The shape of the wavelet allows to represent different detail or resolution.





Properties of Wavelets (cont'd)

- Sparsity: for functions typically found in practice, many of the coefficients in a wavelet representation are either zero or very small.

$$f(t) = \frac{1}{\sqrt{s}} \int_{\tau} \int_{s} C(\tau, s) \psi(\frac{t-\tau}{s}) d\tau ds$$

Properties of Wavelets (cont'd)

$$f(t) = \frac{1}{\sqrt{s}} \int_{\tau} \int_{s} C(\tau, s) \psi\left(\frac{t-\tau}{s}\right) d\tau ds$$

- Adaptability: Can represent functions with discontinuities or corners more efficiently.
- Linear-time complexity: many wavelet transformations can be accomplished in O(N) time.

Discrete Wavelet Transform (DWT)

$$a_{jk} = \sum_t f(t)\psi_{jk}^*(t)$$ (forward DWT)

$$f(t) = \sum_k \sum_j a_{jk}\psi_{jk}(t)$$ (inverse DWT)

where $$\psi_{jk}(t) = 2^{j/2}\psi\left(2^j t - k\right)$$

DFT vs DWT

- DFT expansion:

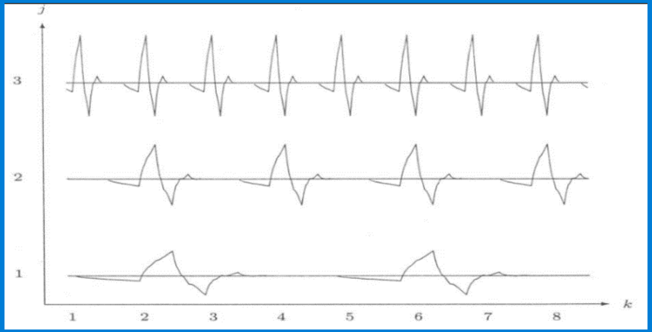
one parameter basis

$$f(x) = \sum_{u=0}^{N-1} F(u) e^{\frac{j2\pi ux}{N}}, \quad \text{or} \quad f(t) = \sum_{l} a_l \psi_l(t)$$

- DWT expansion

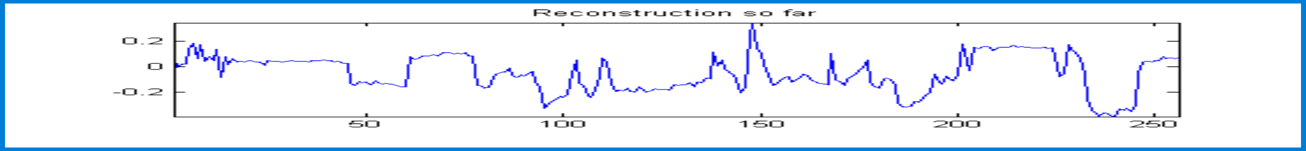
two parameter basis

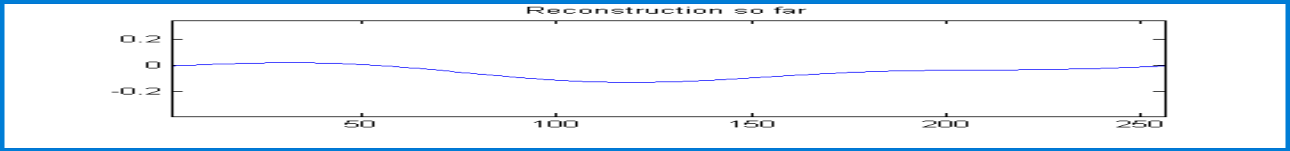
$$f(t) = \sum_{k} \sum_{j} a_{jk} \psi_{jk}(t)$$



Multiresolution Representation Using Wavelets

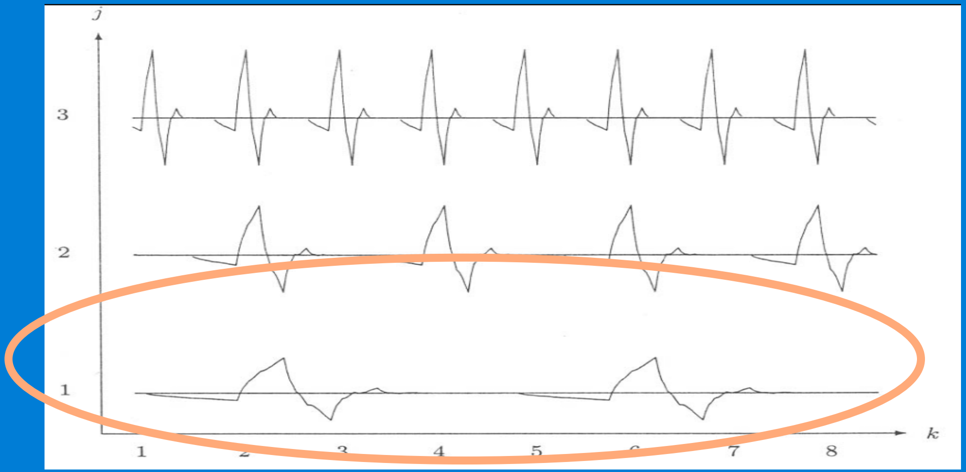
$f(t)$





wider, large translations

$$f(t)=\sum_{k}\sum_{j}a_{jk}\psi_{jk}(t)$$



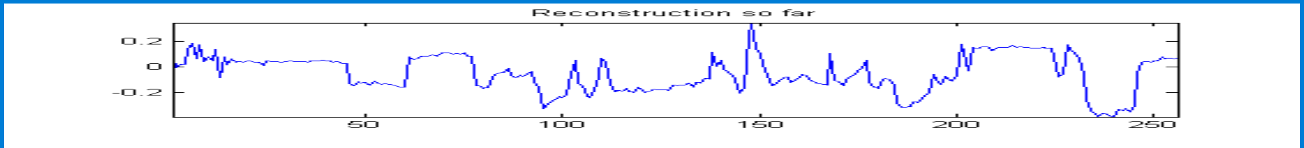
fine details

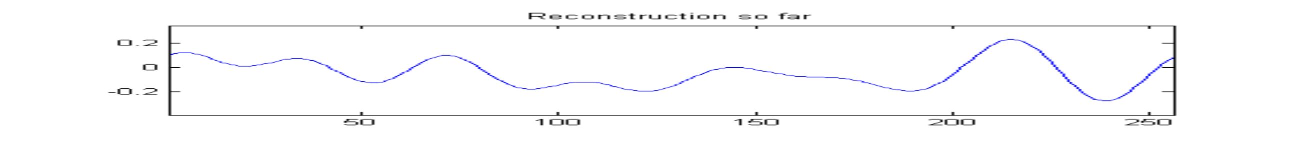
j

coarse details

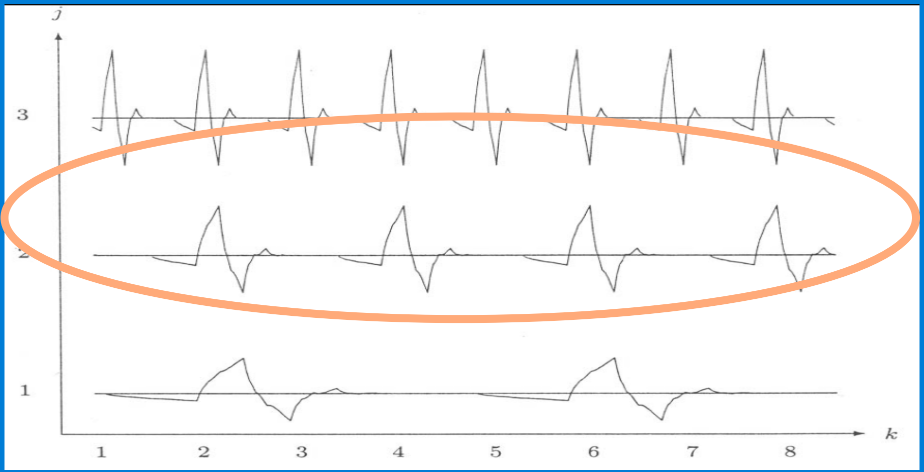
Multiresolution Representation Using Wavelets

$f(t)$





$$f(t) = \sum_{k} \sum_{j} a_{jk} \psi_{jk}(t)$$



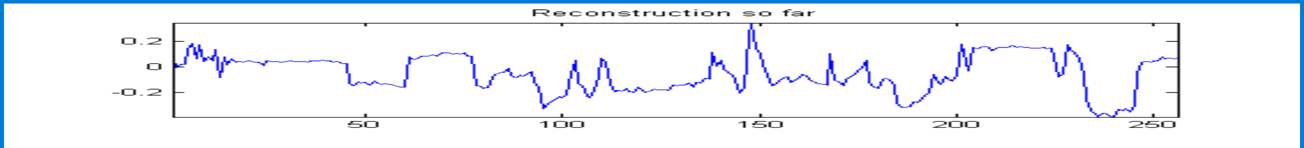
fine details

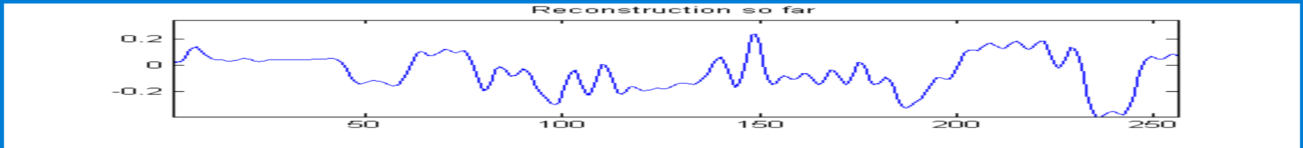
j

coarse details

Multiresolution Representation Using Wavelets

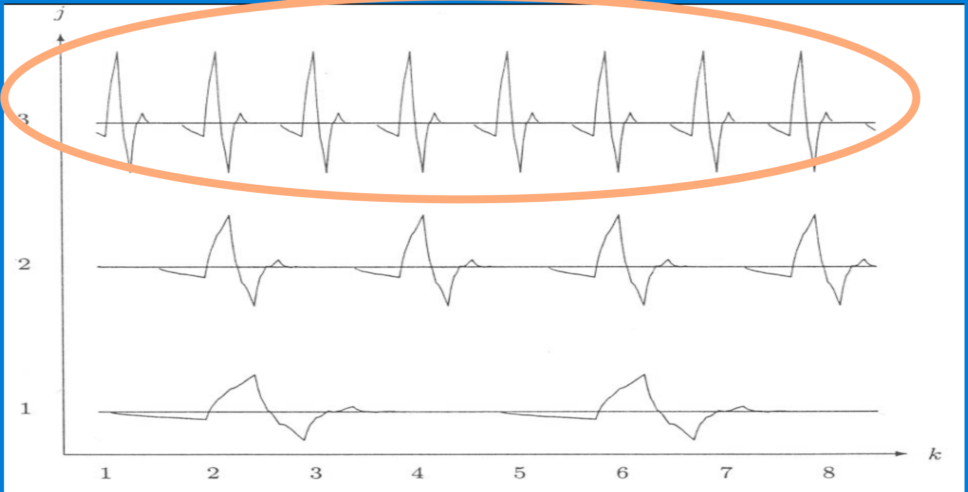
$f(t)$





narrower, small translations

$$f(t) = \sum_{k} \sum_{j} a_{jk} \psi_{jk}(t)$$



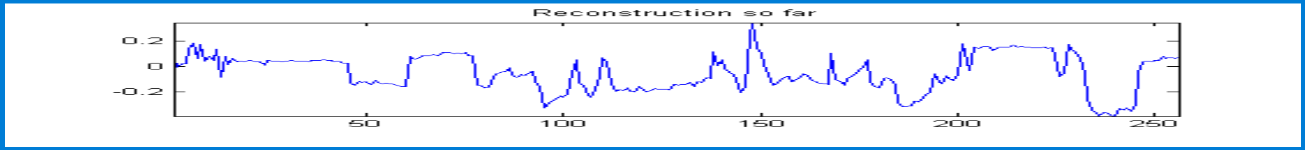
fine details

j

coarse details

Multiresolution Representation Using Wavelets

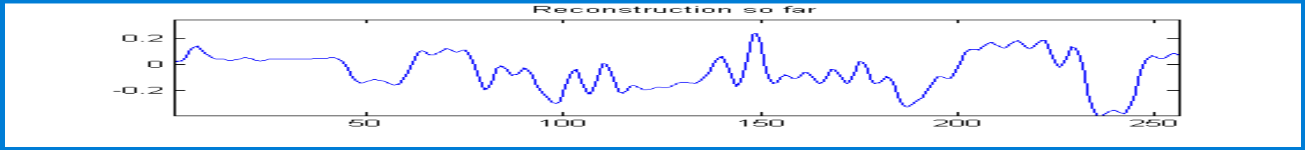
$f(t)$



high resolution

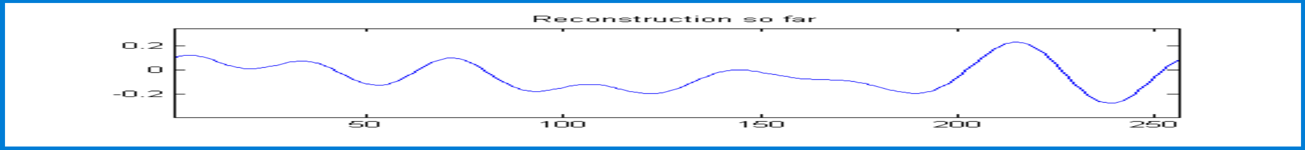
(more details)

$\hat{f}_1(t)$



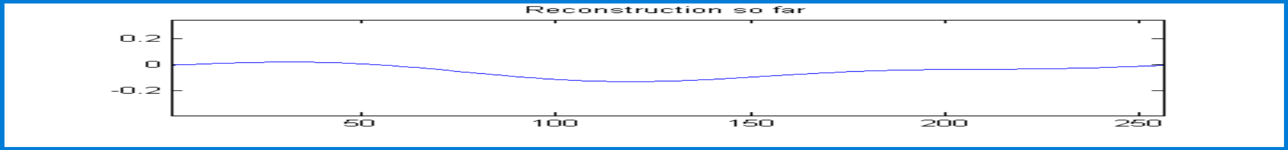
j

$\hat{f}_2(t)$



• • •

$\hat{f}_s(t)$



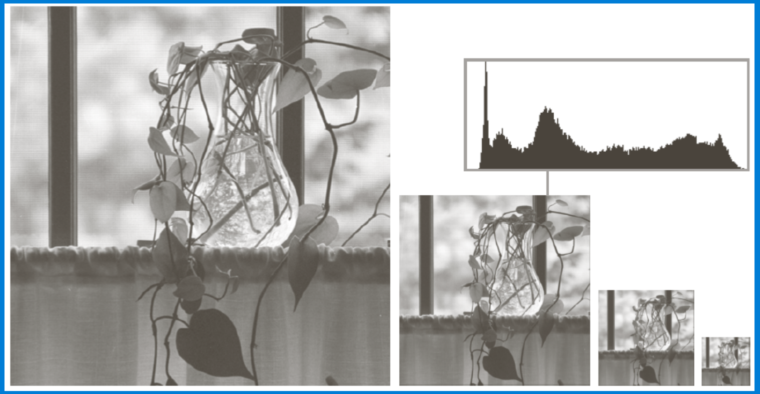
low resolution

(less details)

$$f(t) = \sum_k \sum_j a_{jk} \psi_{jk}(t)$$

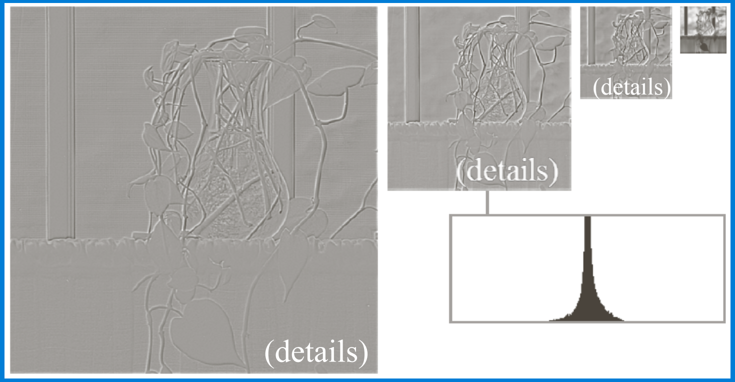
Pyramidal Coding - Revisited

Approximation Pyramid



(with sub-sampling)

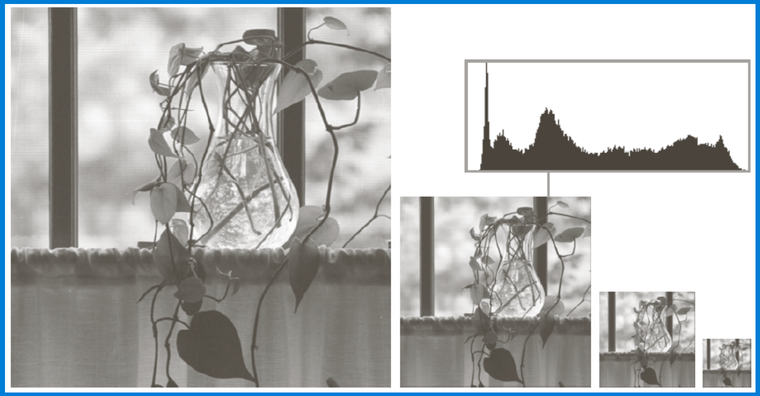
Pyramidal Coding - Revisited



Prediction Residual Pyramid

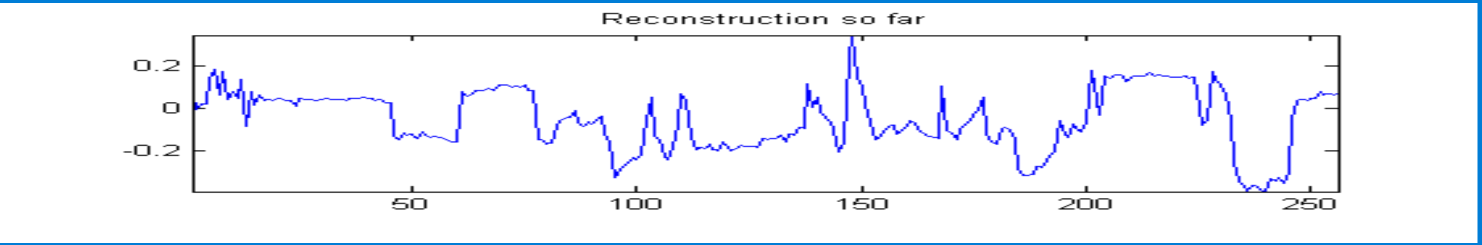
(with sub-sampling)

reconstruct



Approximation Pyramid

Efficient Representation Using Details (cont'd)

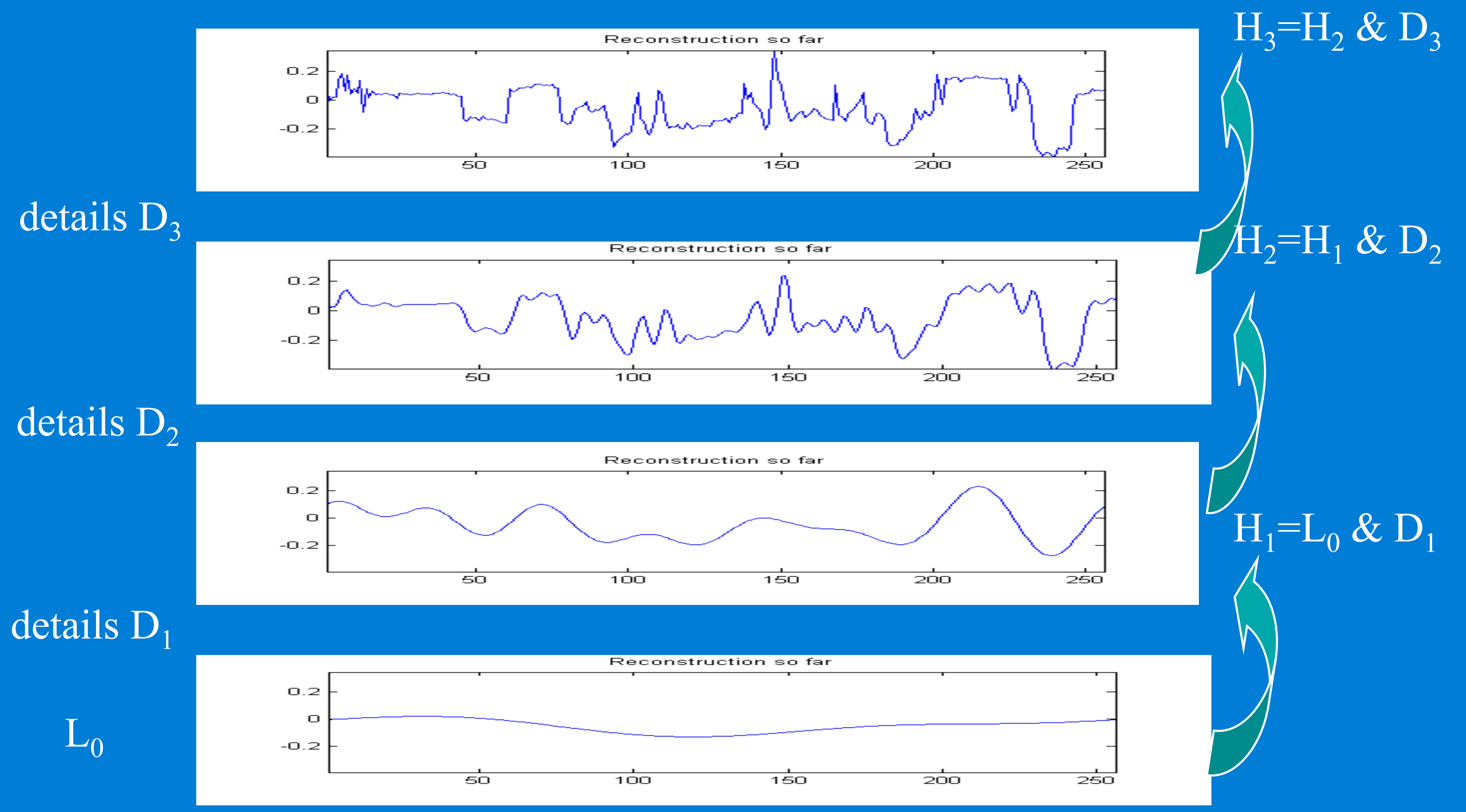


representation: $L_0\ D_1\ D_2\ D_3$

in general: $L_0\ D_1\ D_2\ D_3 \ldots D_J$

A wavelet representation of a function consists of
(1) a coarse overall approximation
(2) detail coefficients that influence the function at various scales

Reconstruction (synthesis)



(without sub-sampling)

Example - Haar Wavelets

- Suppose we are given a 1D "image" with a resolution of 4 pixels:

$$[9\ 7\ 3\ 5]$$

- The Haar wavelet transform is the following:

$$[6\quad 2\quad 1\quad -1]$$ (with sub-sampling)

$L_0\ D_1\ D_2\ D_3$

Example - Haar Wavelets (cont'd)

- Start by averaging and subsampling the pixels together (pairwise) to get a new lower resolution image:

[9 7 3 5] → [8 4]

- To recover the original four pixels from the two averaged pixels, store some *detail coefficients*.

Resolution	*Averages*	*Detail Coefficients*
1	[9 7 3 5]	[]
2	[8 4]	[1 −1]

Example - Haar Wavelets (cont'd)

- Repeating this process on the averages (i.e., low resolution image) gives the full decomposition:

Resolution	*Averages*	*Detail Coefficients*
1	$[9\ \ 7\ \ 3\ \ 5]$	$[]$
2	$[8\ \ 4]$	$[1\ \ -1]$
4	$[6]$	$[2]$

Haar decomposition: $[6\ \ 2\ \ 1\ \ -1]$

Example - Haar Wavelets (cont'd)

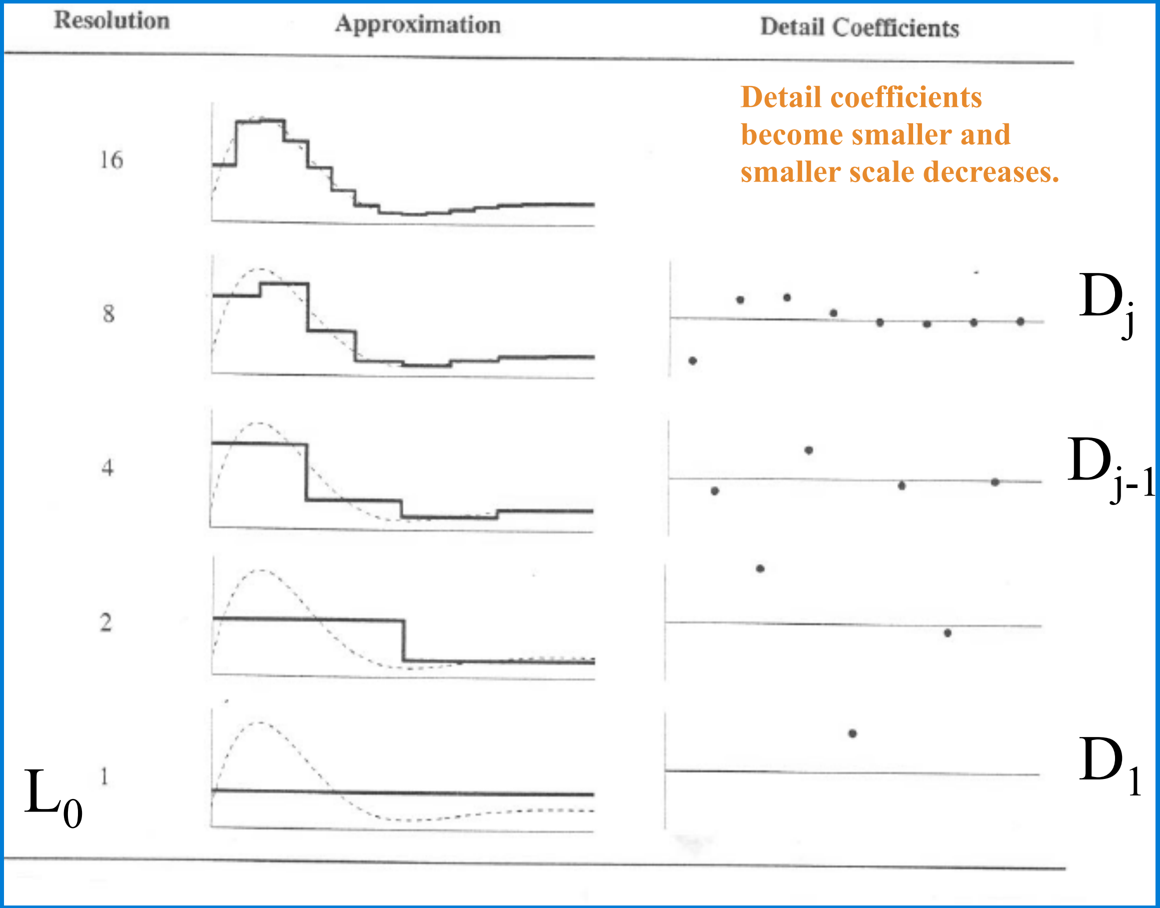
- The original image can be reconstructed by adding or subtracting the detail coefficients from the lower-resolution representations.

$$[6 \;\; 2 \;\; 1 \;\; -1]$$

$$L_0 \; D_1 \; D_2 \; D_3$$

$$[6] \xrightarrow{2} [8 \;\; 4] \xrightarrow{1 \;\; -1} [9 \;\; 7 \;\; 3 \;\; 5]$$

Example - Haar Wavelets (cont'd)



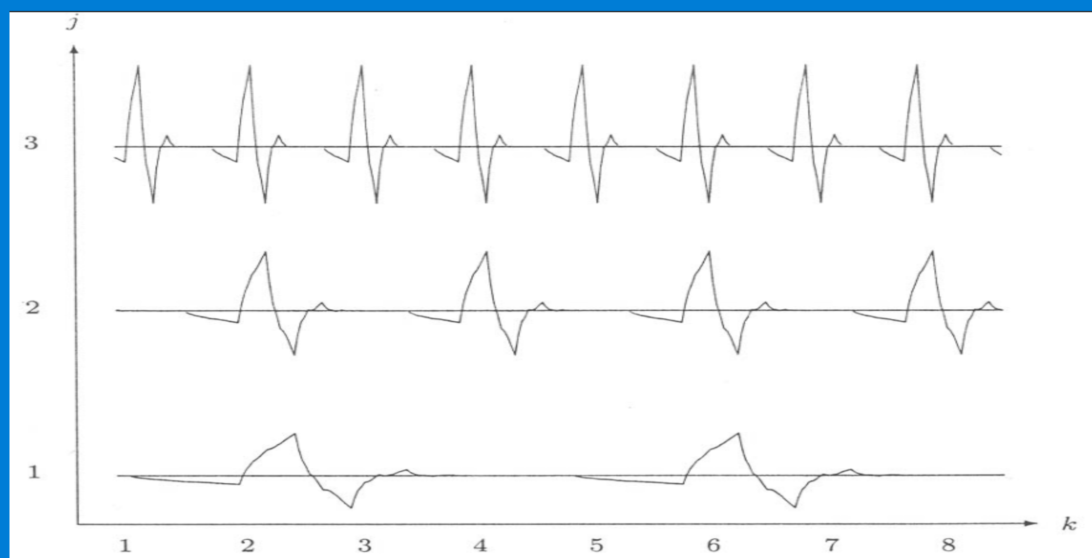
How should we compute the detail coefficients D_j ?

Multiresolution Conditions

- If a set of functions V can be represented by a weighted sum of $\psi(2^j t - k)$, then a **larger set**, including V, can be represented by a weighted sum of $\psi(2^{j+1} t - k)$.

$\psi(2^{j+1} t - k)$

$\psi(2^j t - k)$



high resolution

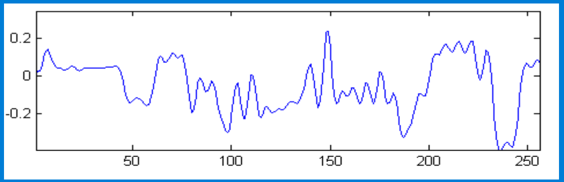
j

low resolution

Multiresolution Conditions (cont'd)

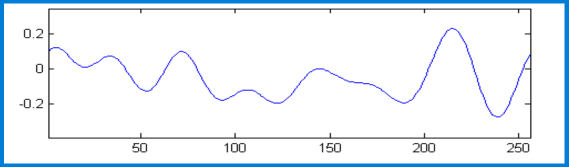
V_{j+1}: span of $\psi(2^{j+1}t - k)$:

$$f_{j+1}(t) = \sum_k b_k \psi_{(j+1)k}(t)$$



V_j: span of $\psi(2^{j}t - k)$:

$$f_j(t) = \sum_k a_k \psi_{jk}(t)$$



$$V_j \subseteq V_{j+1}$$

Multiresolution Conditions (cont'd)

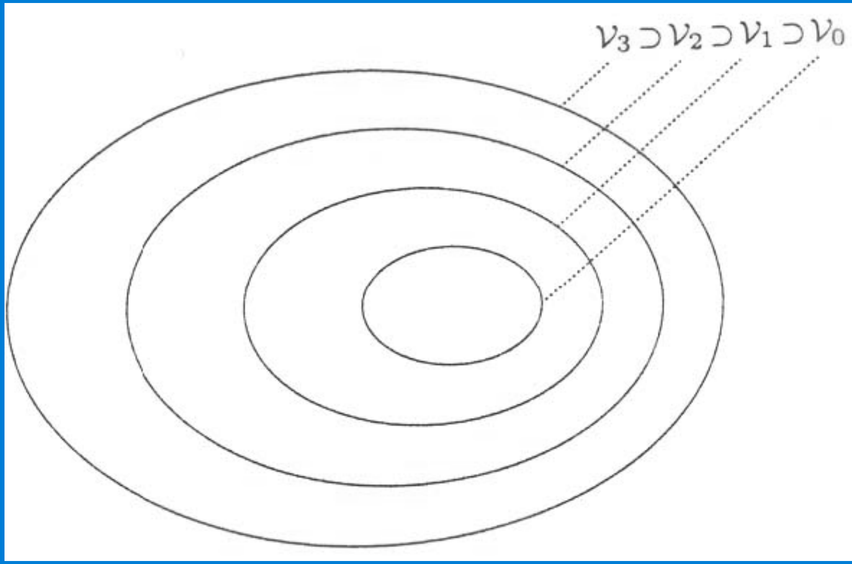
j=0 $\psi(t - k) \longrightarrow V_0$

j=1 $\psi(2t - k) \longrightarrow V_1$

…

j $\psi(2^j t - k) \longrightarrow V_j$

Nested Spaces



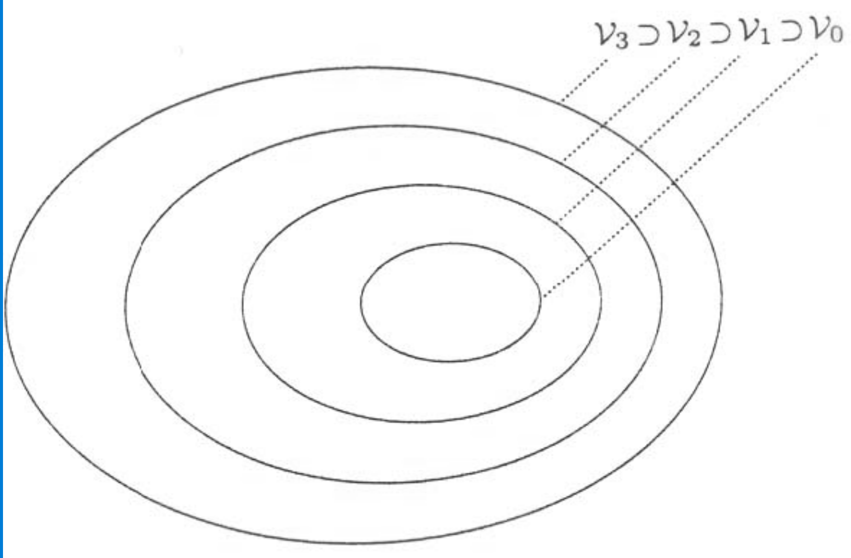
$$V_j \subset V_{j+1}$$

if f(t) $\in V_j$ then f(t) $\in V_{j+1}$

How to compute D_j ?

IDEA:

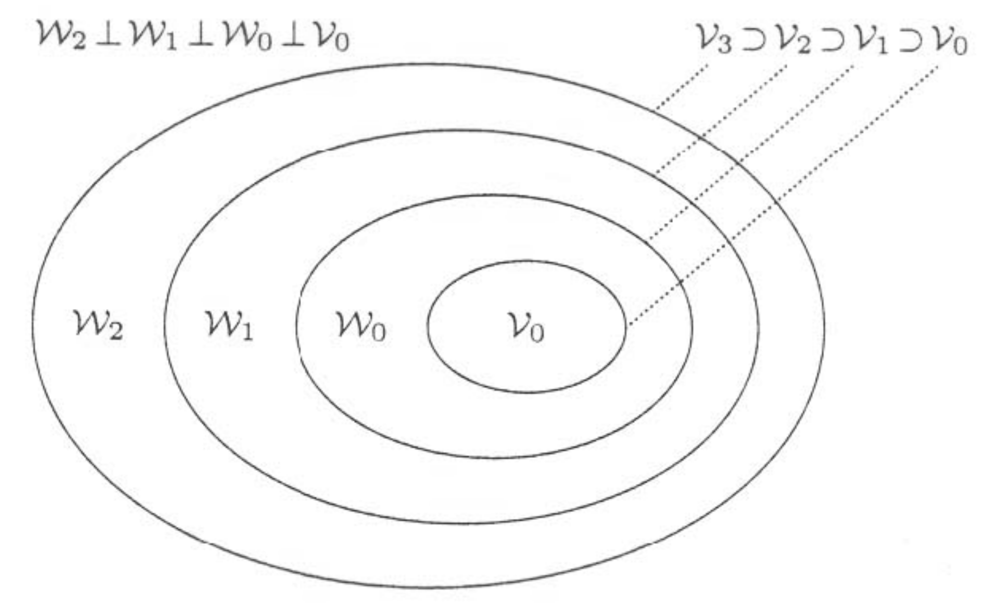
Define a set of basis functions that span the difference between V_{j+1} and V_j



How to compute D_j ? (cont'd)

- Let W_j be the orthogonal complement of V_j in V_{j+1}

$$V_{j+1} = V_j + W_j$$

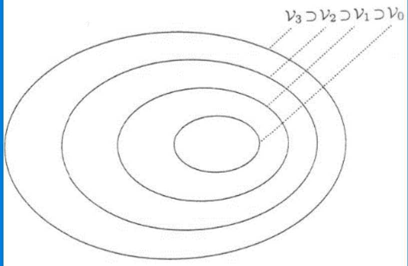


How to compute D_j ? (cont'd)

If $f(t) \in V_{j+1}$, then $f(t)$ can be represented using basis functions $\varphi(t)$ from V_{j+1}:

V_{j+1}

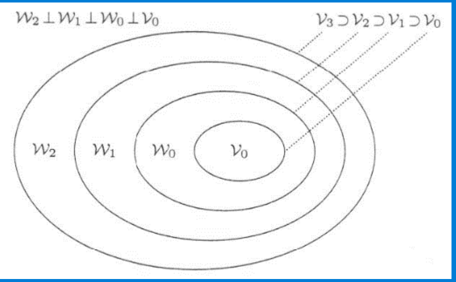
$$f(t) = \sum_k c_k \varphi(2^{j+1} t - k)$$



Alternatively, $f(t)$ can be represented using **two** sets of basis functions, $\varphi(t)$ from V_j and $\psi(t)$ from W_j:

$$V_{j+1} = V_j + W_j$$

$$f(t) = \sum_k c_k \varphi(2^j t - k) + \sum_k d_{jk} \psi(2^j t - k)$$



How to compute D_j? (cont'd)

Think of W_j as a means to represent the parts of a function in V_{j+1} that cannot be represented in V_j

$$f(t) = \sum_k c_k \varphi(2^{j+1} t - k)$$

V_{j+1}

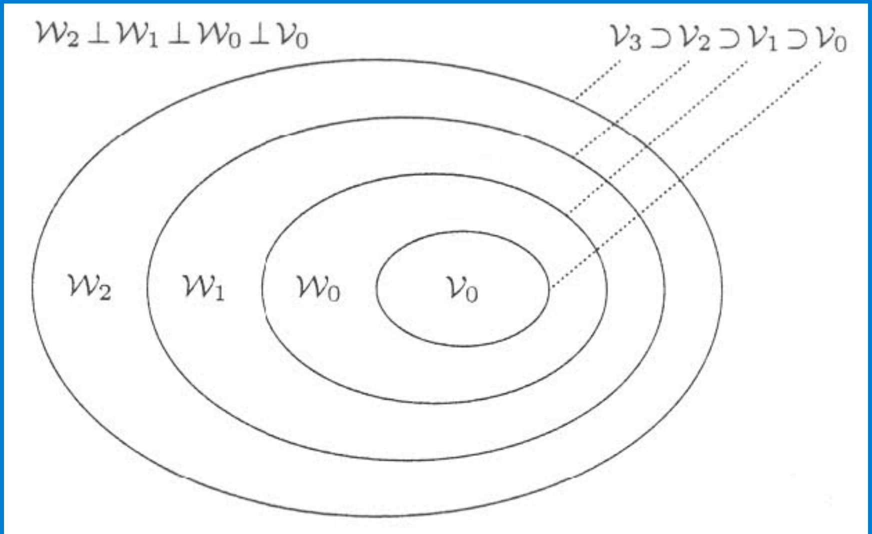
$$f(t) = \underbrace{\sum_k c_k \varphi(2^j t - k)}_{V_j} + \underbrace{\sum_k d_{jk} \psi(2^j t - k)}_{W_j}$$

differences between V_j and V_{j+1}

How to compute D_j ? (cont'd)

- $V_{j+1} = V_j + W_j$ → using recursion on V_j:

$$V_{j+1} = V_{j-1} + W_{j-1} + W_j = \ldots = V_0 + W_0 + W_1 + W_2 + \ldots + W_j$$



if $f(t) \in V_{j+1}$, then:

$$f(t) = \sum_k c_k \varphi(t-k) + \sum_k \sum_j d_{jk} \psi(2^j t - k)$$

V_0 basis functions

$W_0, W_1, W_2, \ldots$ basis functions

Summary: wavelet expansion (Section 7.2)

- Wavelet decompositions involve a pair of waveforms:

encodes low resolution info $\varphi(t)$

$\psi(t)$ encodes details or high resolution info

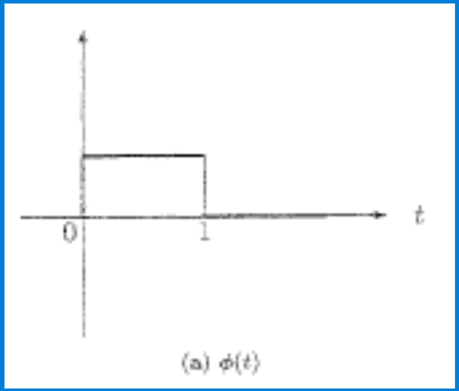
$$f(t) = \sum_k c_k \varphi(t-k) + \sum_k \sum_j d_{jk} \psi(2^j t - k)$$

Terminology: scaling function wavelet function

1D Haar Wavelets

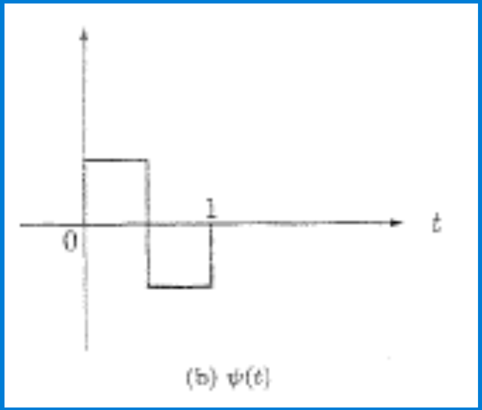
- Haar scaling and wavelet functions:

φ(t)



computes average
(low pass)

ψ(t)



computes details
(high pass)

1D Haar Wavelets (cont'd)

Let's consider the spaces corresponding to different resolution 1D images:

V_0	•	1-pixel	(j=0)
V_1	• •	2-pixel	(j=1)
V_2	• • • •	4-pixel	(j=2)

etc.

1D Haar Wavelets (cont'd)

j=0

- V_0 represents the space of 1-pixel (2^0-pixel) images

- Think of a 1-pixel image as a function that is constant over [0,1)

Example:

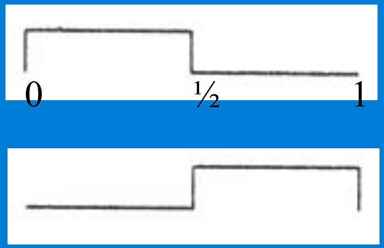
width: 1

1D Haar Wavelets (cont'd)

j=1

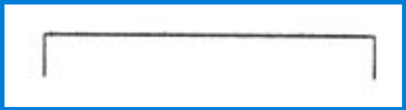
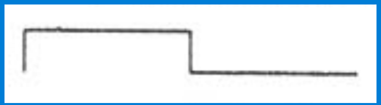
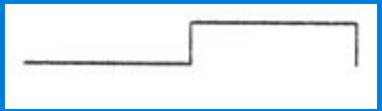
- V_1 represents the space of all 2-pixel (2^1-pixel) images
- Think of a 2-pixel image as a function having 2^1 equal-sized constant pieces over the interval [0, 1).

Example:



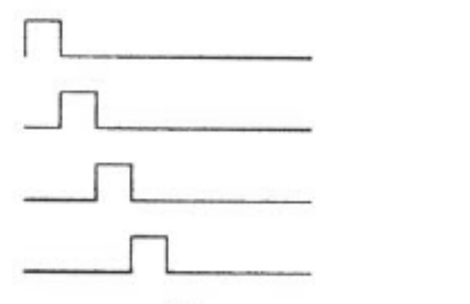
width: 1/2

Note that: $\boxed{V_0 \subset V_1}$

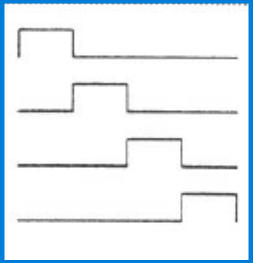
e.g.,  =  + 

1D Haar Wavelets (cont'd)

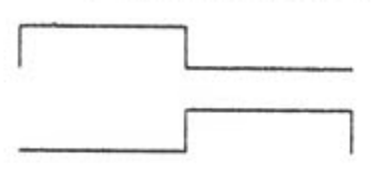
- V_j represents all the 2^j-pixel images
- Functions having 2^j equal-sized constant pieces over interval $[0,1)$.

Example:  width: $1/2^j$

Note that:

V_{j-1}  $\in V_j$

width: $1/2^{j-1}$

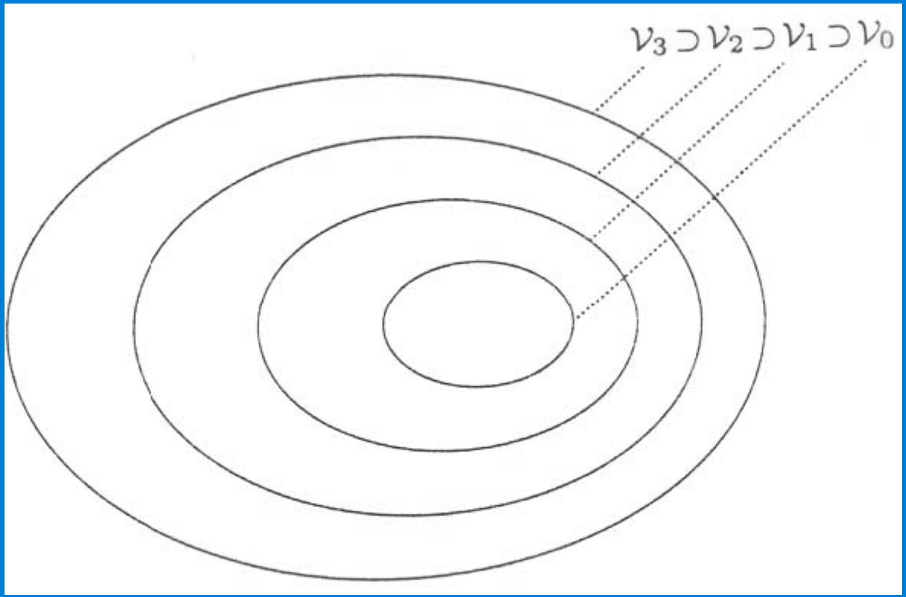
V_1  $\in V_j$

width: $1/2$

1D Haar Wavelets (cont'd)

$V_0, V_1, \ldots, V_j$ are nested

i.e., $\boxed{V_j \subset V_{j+1}}$



V_j fine details

…

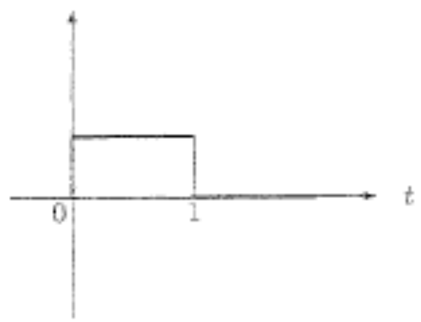
V_1

V_0 coarse info

Define a basis for V_j

- Scaling function:

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1 \\ 0 & otherwise \end{cases}$$



(a) $\phi(t)$

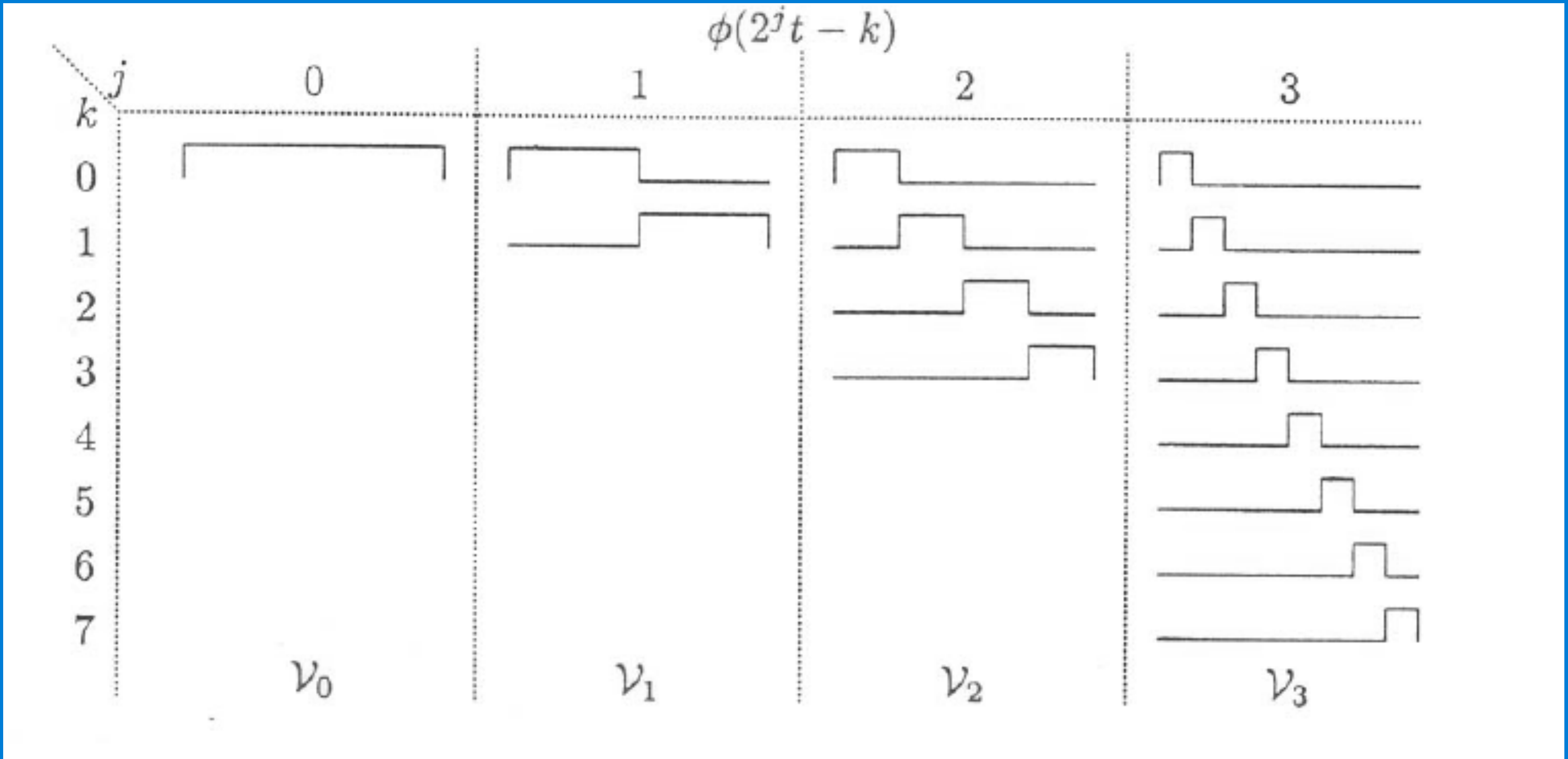
- Let's define a basis for V_j :

$$\phi_i^j(x) := \phi(2^j x - i), \quad i = 0, 1, \ldots, 2^j - 1$$

(scaled and translated versions of the box function below)

Note new notation: $\varphi_i^j(x) \equiv \varphi_{ji}(x)$

Define a basis for V_j (cont'd)

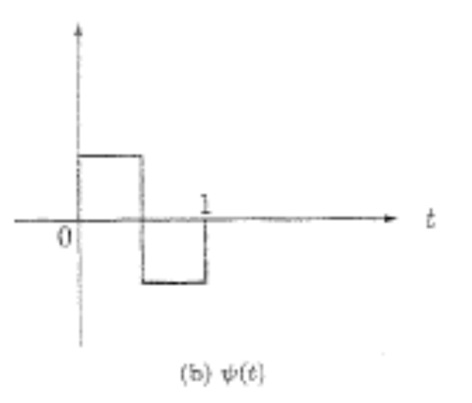


width: $1/2^0$ width: $1/2$ width: $1/2^2$ width: $1/2^3$

Define a basis for W_j

- Wavelet function:

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2 \\ -1 & \text{if } 1/2 \le x < 1 \\ 0 & \textit{otherwise} \end{cases}$$

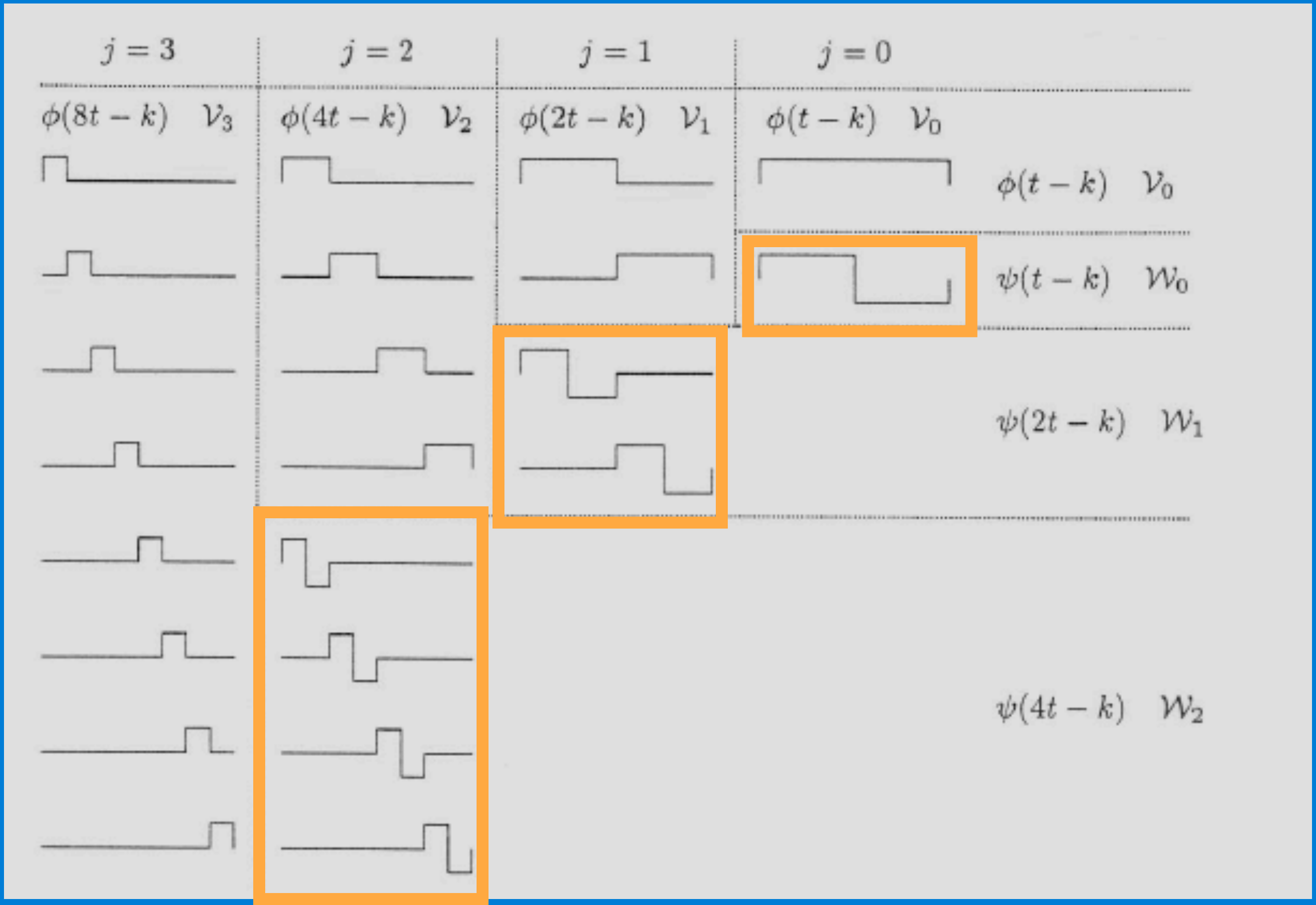


- Let's define a basis ψ^j_i for W_j :

$$\psi_i^j(x) := \psi(2^j x - i), \quad i = 0, 1, \ldots, 2^j - 1$$

Note new notation: $\psi_i^j(x) \equiv \psi_{ji}(x)$

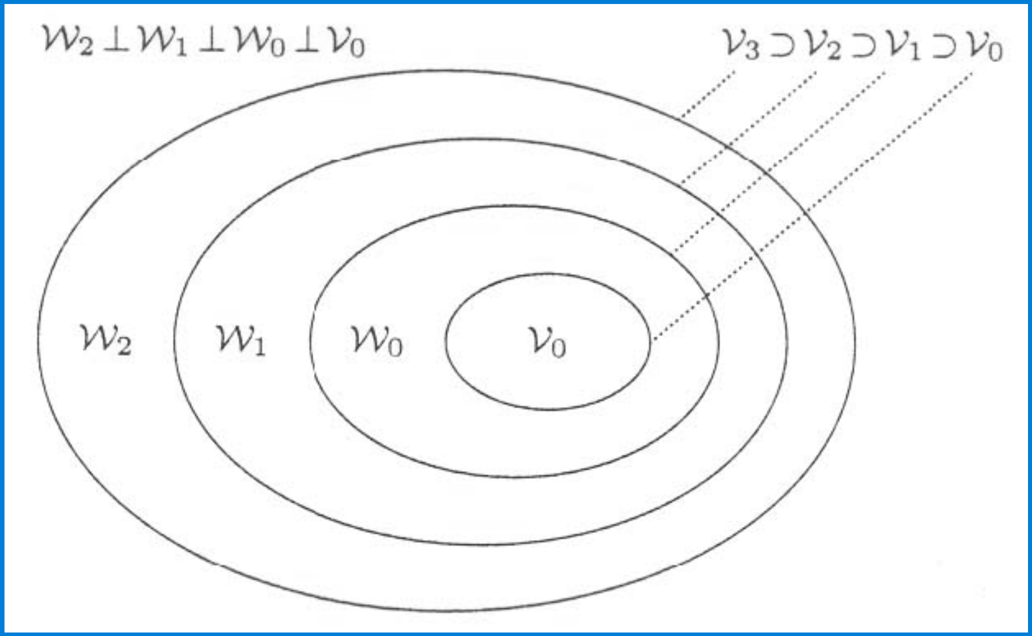
Define basis for W_j (cont'd)



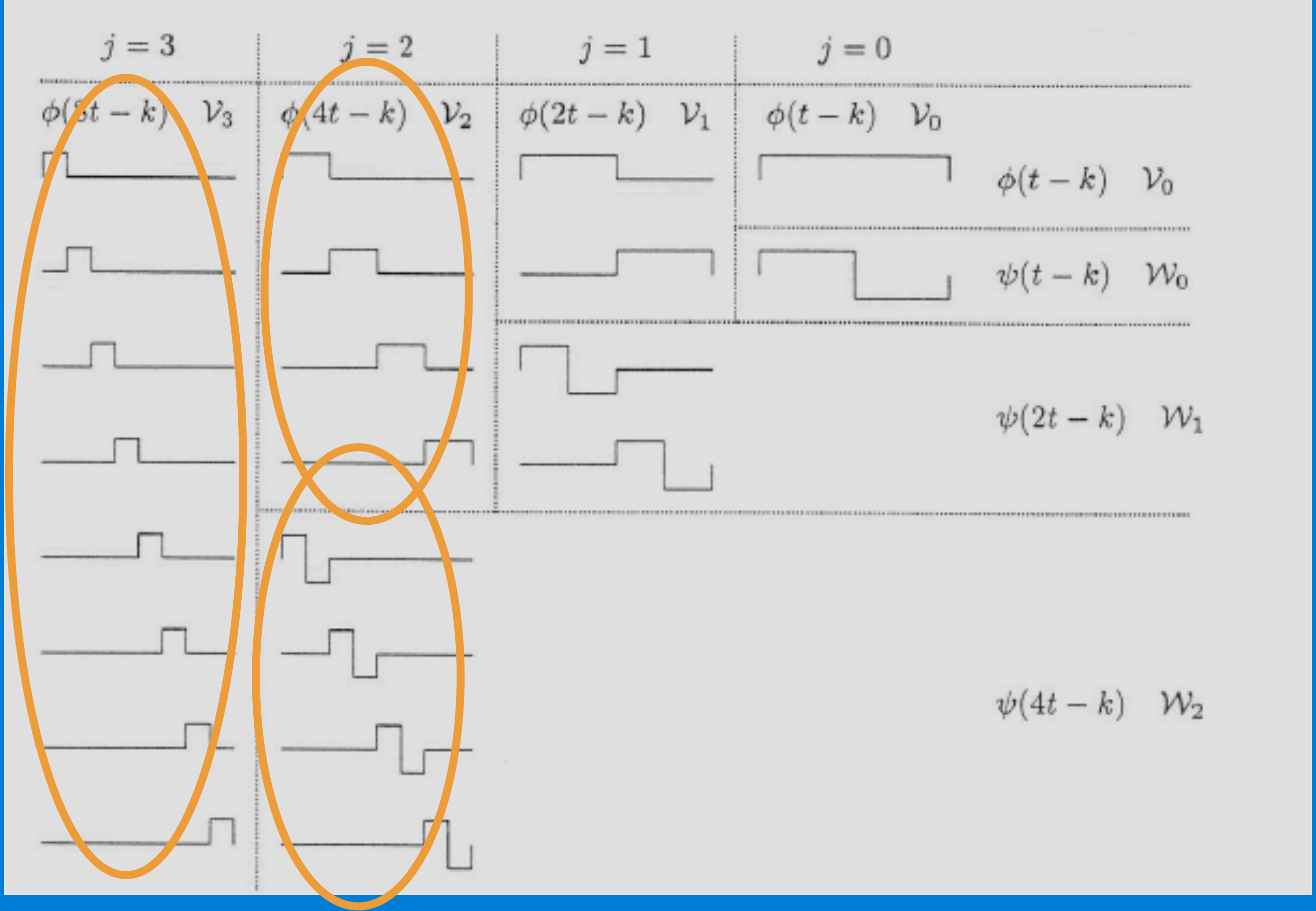
Note that the dot product between basis functions in V_j and W_j is zero!

Basis for V_{j+1}

Basis functions ψ^{j}_{i} of W_j
Basis functions φ^{j}_{i} of V_j
form a basis in V_{j+1}

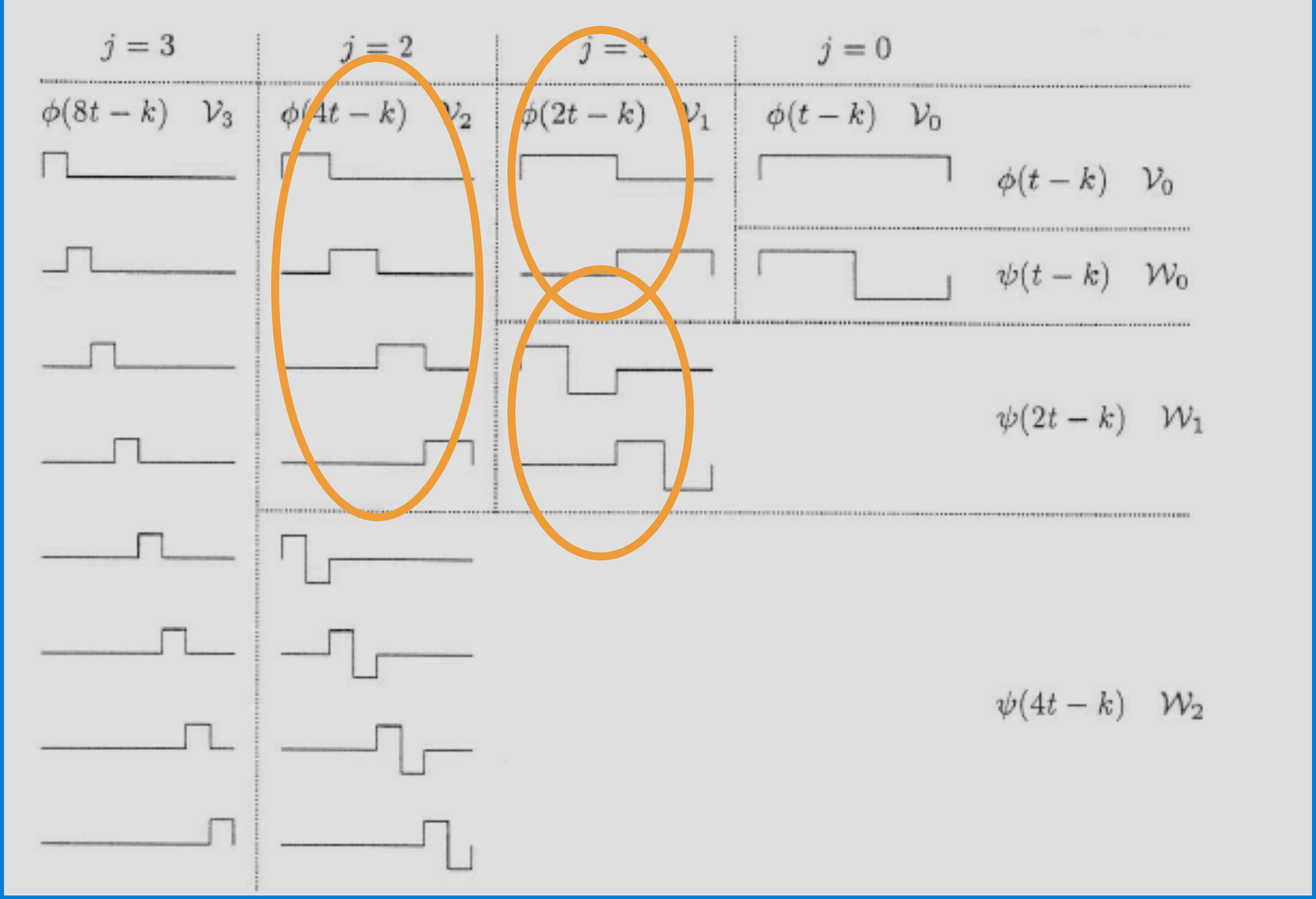


Define a basis for W_j (cont'd)



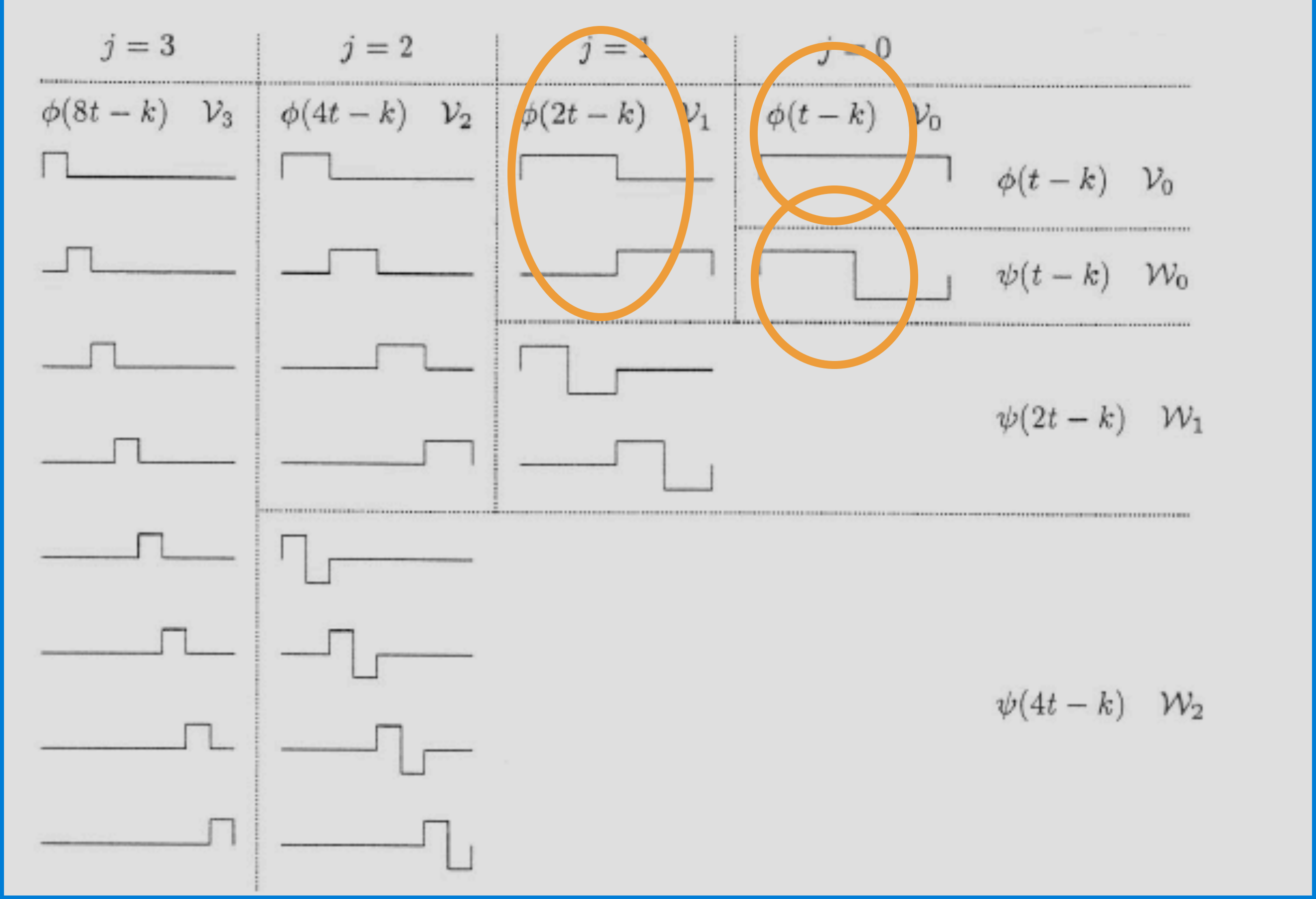
$V_3 = V_2 + W_2$

Define a basis for W_j (cont'd)



$V_2 = V_1 + W_1$

Define a basis for W_j (cont'd)



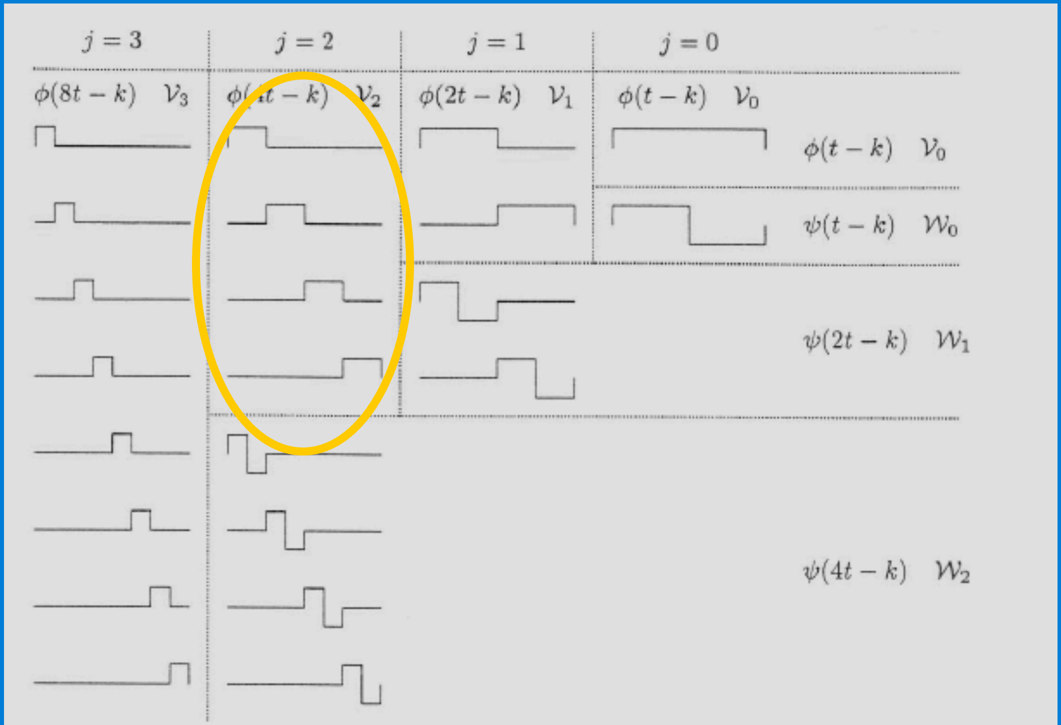
$$V_1 = V_0 + W_0$$

Example - Revisited

Resolution	*Averages*	*Detail Coefficients*
4	[9 7 3 5]	[]
2	[8 4]	[1 −1]
4	[6]	[2]

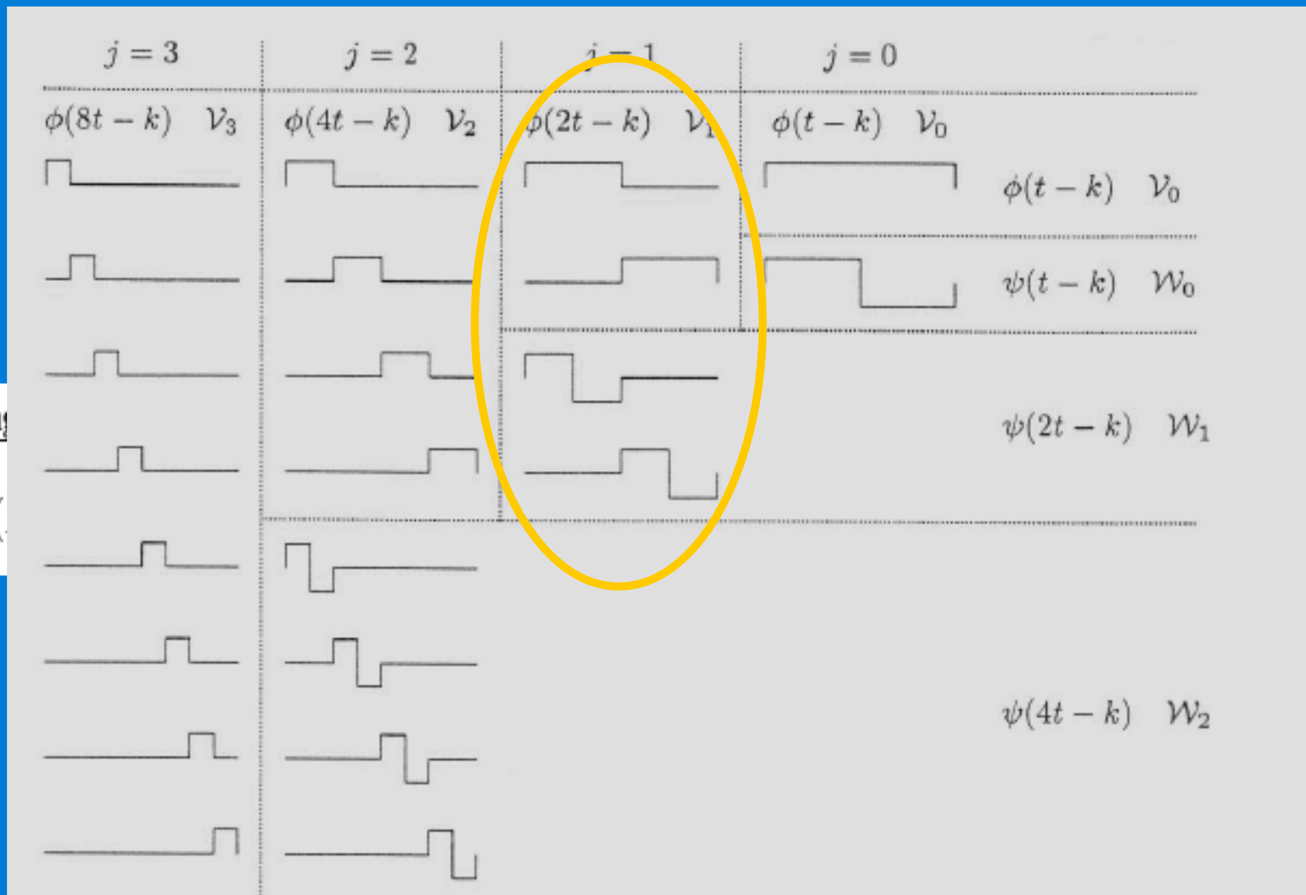
f(x)= [9 7 3 5]

V_2



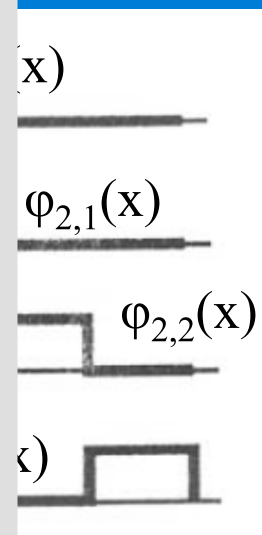
Example (cont’d)

f(x)=



usin

$f(x) = c_0^2\phi_0^2($



Example (cont'd)

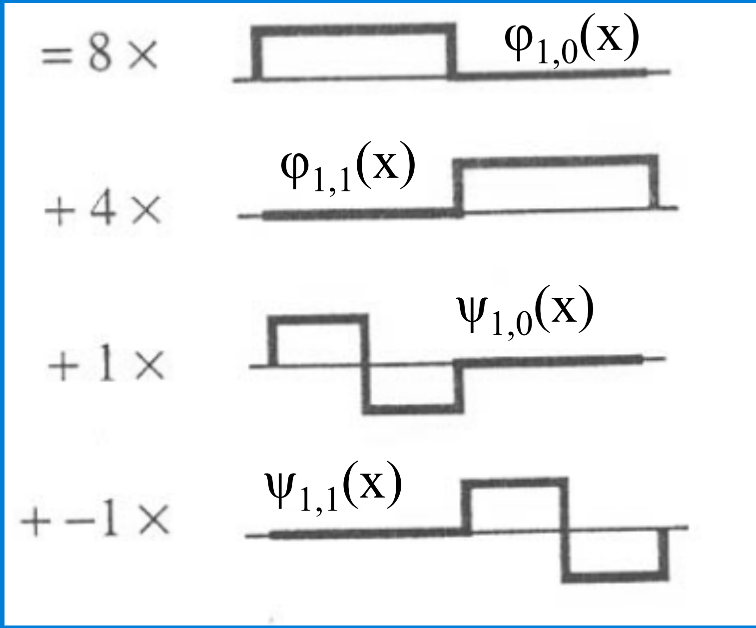
using the basis functions in V_1 and W_1

$$V_2 = V_1 + W_1$$

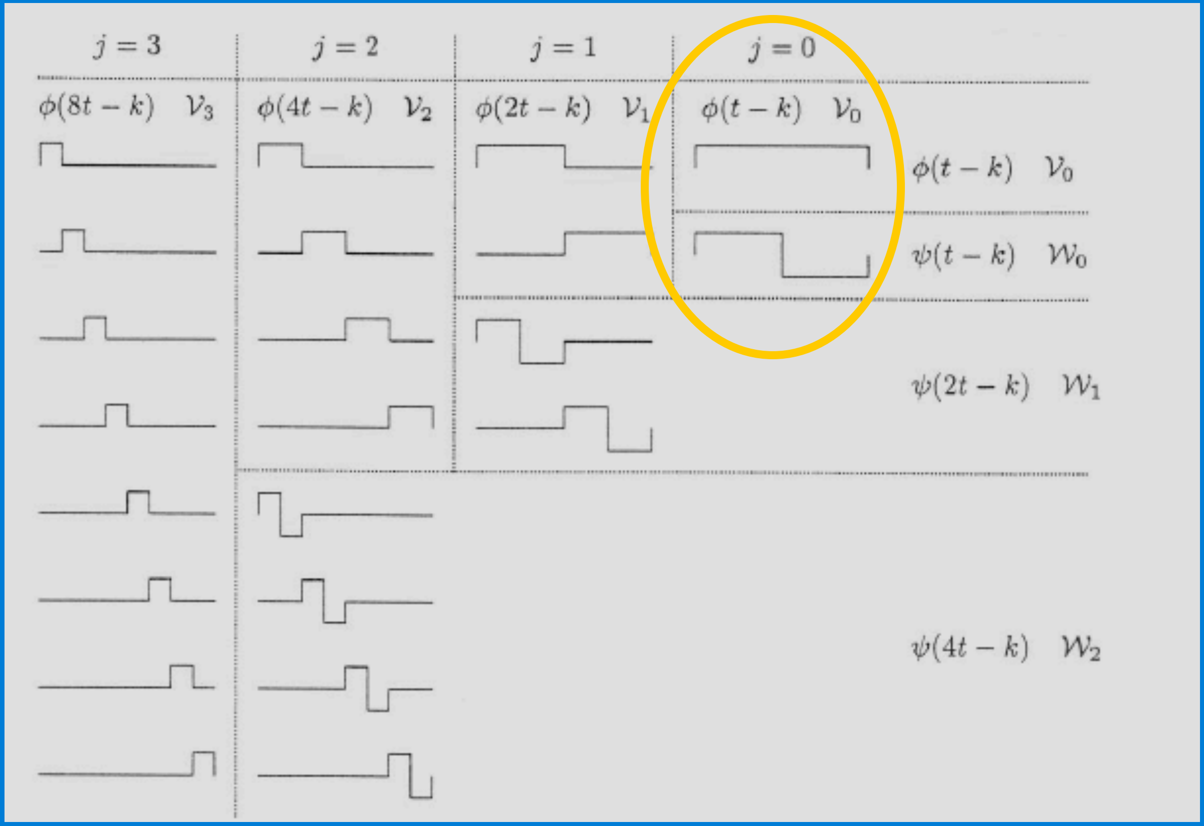
$$f(x) = c_0^1 \phi_0^1(x) + c_1^1 \phi_1^1(x) + d_0^1 \psi_0^1(x) + d_1^1 \psi_1^1(x)$$

Resolution	*Averages*	*Detail Coefficients*
4	[9 7 3 5]	[]
2	[8 4]	[1 − 1]
4	[6]	[2]

(divide by 2 for normalization)



Example (cont'd)



Example (cont'd)

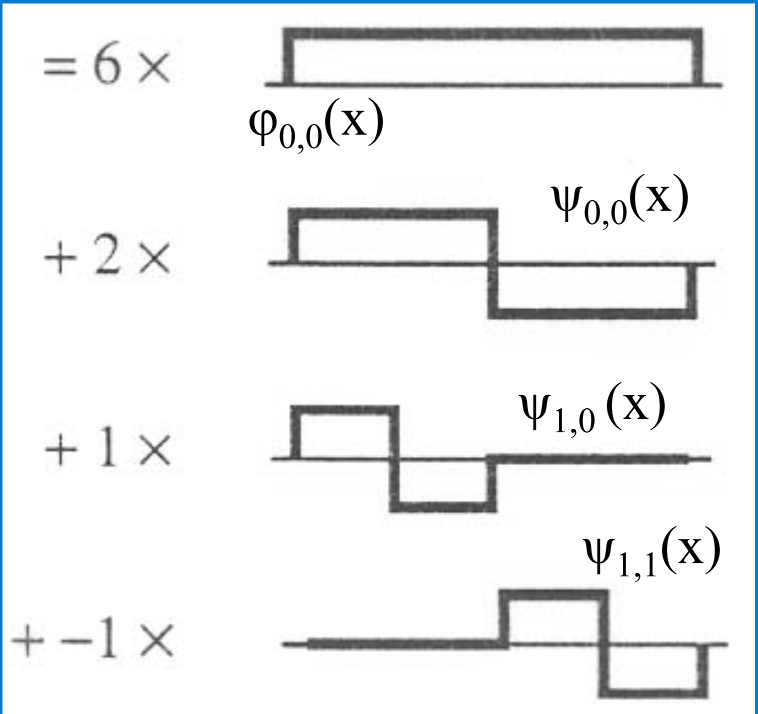
using the basis functions in V_0, W_0 and W_1

$$V_2=V_1+W_1=V_0+W_0+W_1$$

$$f(x) = c_0^0\phi_0^0(x) + d_0^0\psi_0^0(x) + d_0^1\psi_0^1(x) + d_1^1\psi_1^1(x)$$

Resolution	*Averages*	*Detail Coefficients*
4	[9 7 3 5]	[]
2	[8 4]	[1 −1]
4	[6]	[2]

(divide by 2 for normalization)

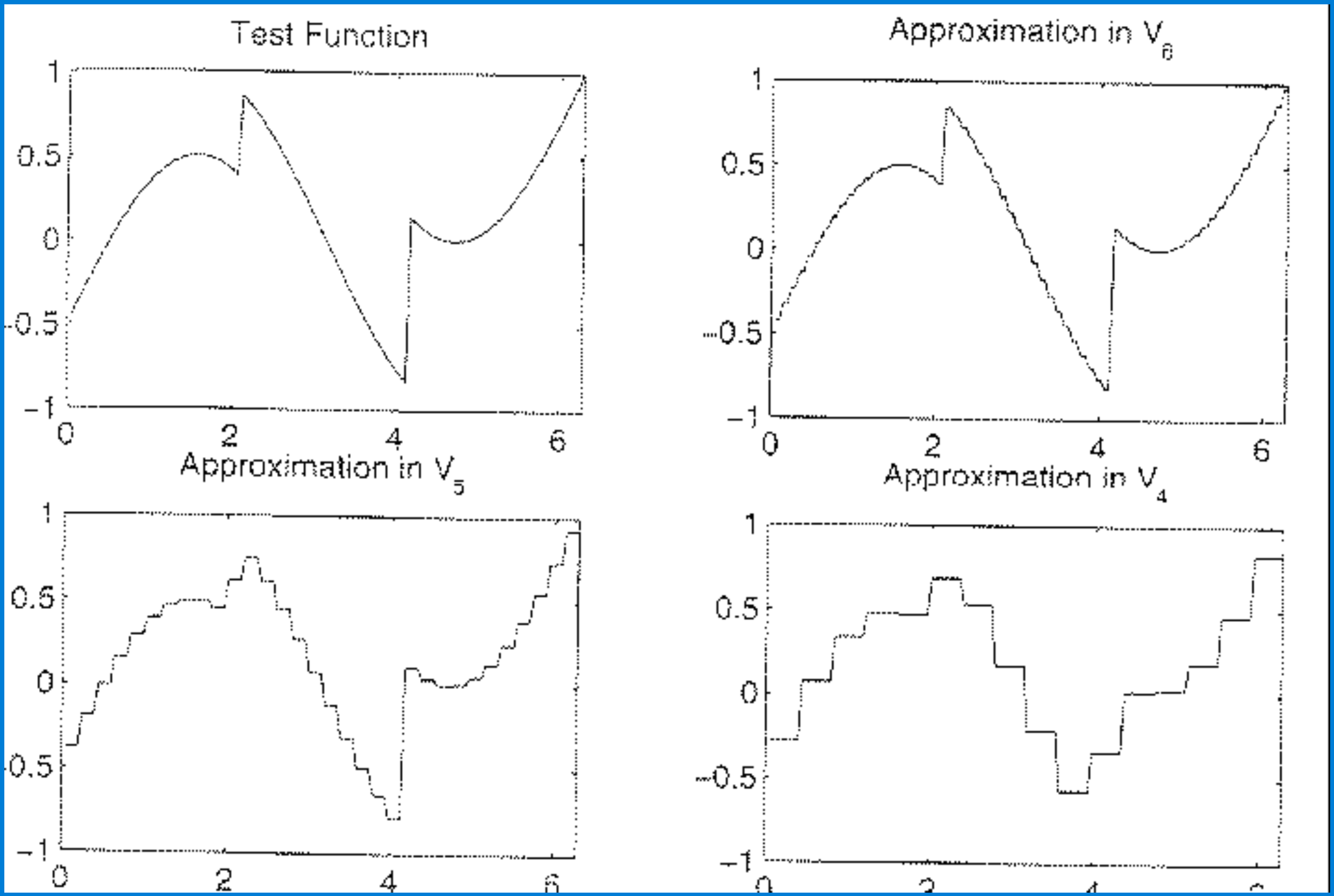


$$f(t) = \sum_k c_k \varphi(t-k) + \sum_k \sum_j d_{jk} \psi(2^j t - k)$$

scaling function

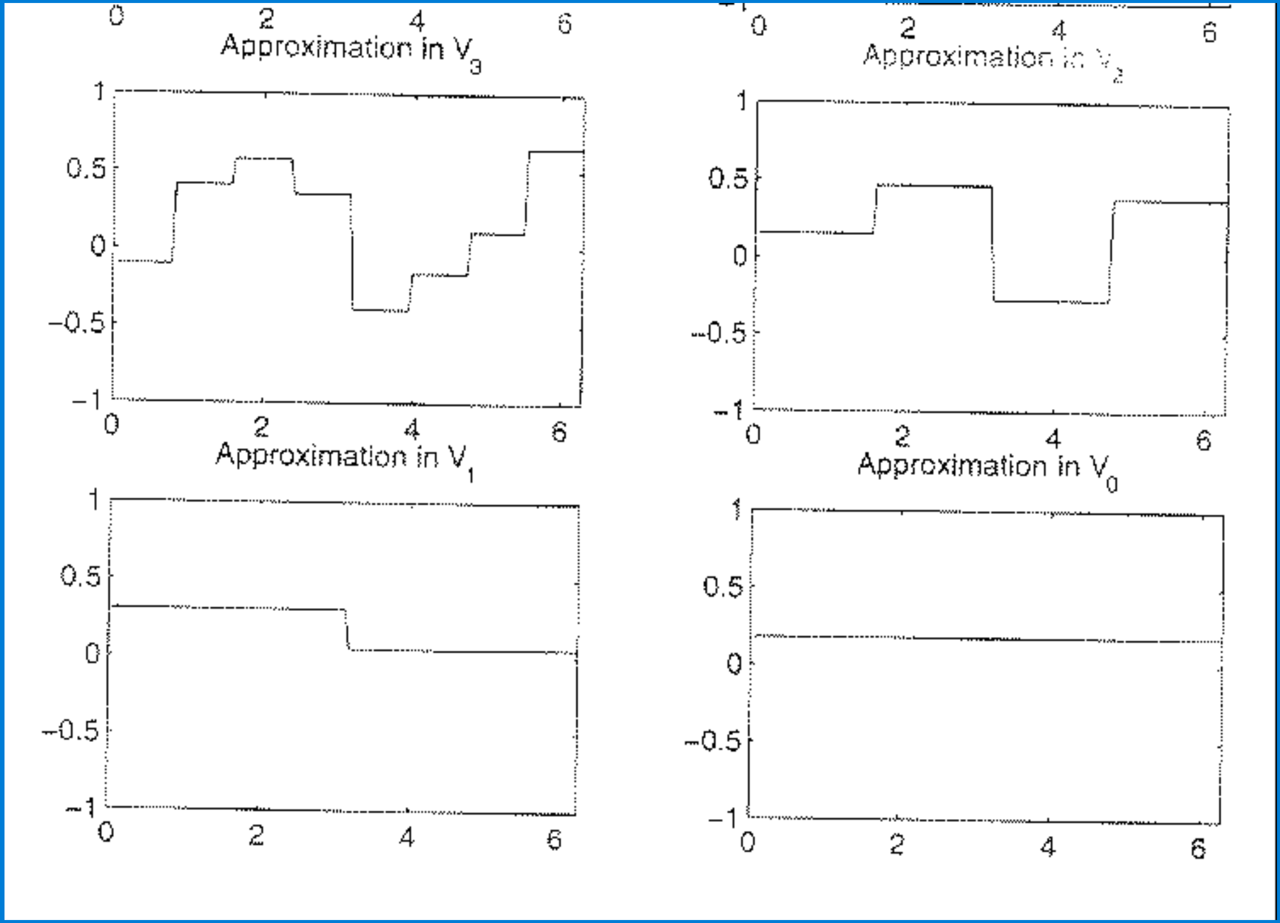
wavelet function

Example



Test Function
Approximation in V_6
Approximation in V_5
Approximation in V_4

Example (cont'd)



Summary

- Structure extraction
  - If the coefficient $d_{j,k}$ is large, then this means that there is some oscillatory variation in f(x).
- Localization in time
- Efficiency
  - Execution times compared with FT.
  - Good recover of discountinuities and corners.
  - A few amount of terms are needed to approximate.

References

- Percibal D. B., Walden A. T., Wavelet Methods for Time Series Analysis
- Ivezic Z., Connolly J., VanderPlas J., Gray A., Statistics, Data Mining, and Maching Learning in Astronomy
- Härdle W., Kerkyacharian G., Picard D., Tsybakov A., Wavelets, Approximation, and Statistical Applications
- Nason, G.P., Wavelet Methods in Statistics with R

- Thanks for your attention…