EE269
Signal Processing for Machine Learning
Wavelets Part II

Instructor: Mert Pilanci
Stanford University

October 5, 2021
What makes a good wavelet

Application specific
- Compact time support vs frequency support
- Smoothness
- Orthogonality
Other Wavelets

- Meyer
- Morlet
- Mexican hat
Other Wavelets

- In MATLAB

\[ [c,1] = \text{wavedec}(x,n,wname) \] returns the wavelet decomposition of the signal \( x \) at level \( n \) using the wavelet \( wname \)
Other Discrete Wavelets

```matlab
[c,1] = wavedec(sumsin,3,'db2');
apprx = appcoef(c,1,'db2');
[cd1,cd2,cd3] = detcoef(c,1,[1 2 3]);

Plot the coefficients.
```

```matlab
subplot(4,1,1)
plot(apprx)
title('Approximation Coefficients')
subplot(4,1,2)
plot(cd3)
title('Level 3 Detail Coefficients')
subplot(4,1,3)
plot(cd2)
title('Level 2 Detail Coefficients')
subplot(4,1,4)
plot(cd1)
title('Level 1 Detail Coefficients')
```
Fourier vs Wavelet Transforms

- Fourier Transform has convolution theorem and mathematical relationships
- No closed form relations exist for wavelet transforms
- Fourier transform has uniform spectral resolution
- Wavelet transform has adaptive resolution
- 100 Hz resolution at 400 Hz and at 4000 Hz are not the same
Short-time Fourier Transform

- window signal

  e.g. \( w[m] = \begin{cases} 
  0 & m < 0, m \geq L \\
  1 & 0 \leq m \leq L - 1 
\end{cases} \)

- Short Time Fourier Transform (STFT)

\[
X[n, k] = \sum_{m=0}^{L-1} x[n + m]w[m]e^{-j(2\pi/N)km}, \quad 0 \leq k \leq N - 1.
\]
Short-time Fourier Transform

- window signal
  
  \[ w[m] = \begin{cases} 
  0 & m < 0, m \geq L \\
  1 & 0 \leq m \leq L - 1 
  \end{cases} \]

- Short Time Fourier Transform (STFT)
  
  \[
  X[n, k] = \sum_{m=0}^{L-1} x[n + m]w[m]e^{-j(2\pi/N)km}, \quad 0 \leq k \leq N - 1.
  \]

- Continuous Frequency STFT
  
  \[
  X[n, \lambda] = \sum_{m=0}^{L-1} x[n + m]w[m]e^{-j\lambda m},
  \]
Short-time Fourier Transform

\[
X[n, \lambda] = \sum_{m=0}^{L-1} x[n + m]w[m]e^{-j\lambda m},
\]

- time-windowed signal
Short-time Fourier Transform vs Wavelet Transform

- windowed signal = windowed complex exponential basis
- STFT has uniform time and frequency resolution
- In contrast, wavelets have adaptive windows:
  - short windows for higher frequencies (small scale)
  - long windows for lower frequencies (large scale)
Wavelet Transform vs STFT

Wavelet transform analyzes a signal at different frequencies with different resolutions:

- good time resolution and relatively poor frequency resolution at high frequencies
- good frequency resolution and relatively poor time resolution at low frequencies

Wavelet transform is better for signals with non-periodic and fast transient features (i.e., high frequency content for short duration)
Wavelet Transform vs STFT

Wavelet Transform

Short Time Fourier Transform

Signal

STFT

CWT
Wavelet Transform vs STFT: Locality

**TIME PLOT OF SIGNAL WITH SMALL DISCONTINUITY**

- **a. Large sinewave with small glitch**

**CLOSEUP OF THE DISCONTINUITY**

- **b. Closeup of glitch**
Wavelet Transform vs STFT: Locality

Stretched “low frequency” wavelet compares better to long sinusoidal (wave) signal. It “finds” peaks and valleys.

Short “high frequency” wavelet compares well to discontinuity. It “finds” its location at 180.
Human Activity Recognition (HAR) Dataset
Application

- Human Activity Recognition Using Smartphones Data Set (Reyes-Ortiz et al, 2012)
- Compute DFT of the training signals $X_1[k], X_2[k], \ldots X_m[k]$
- DFT Magnitude $|X_1[k]|, |X_2[k]|, \ldots |X_m[k]|$
Results: training set: 7724 signals, test set: 2575 signals

3-Nearest Neighbors, $\ell_2$-norm distance on $x[n]$. **Accuracy** : 0.77

3-Nearest Neighbors, $\ell_2$-norm distance on $|X[k]|$. **Accuracy** : 0.85
Continuous Wavelet Transform of HAR signals

- changing the number of scales
Continuous Wavelet Transform of HAR signals
Wavelet Transform Features

- mean, median
- variance
- zero crossing rate, mean crossing rate
- entropy

slide credit: A. Taspinar
Human Activity Recognition dataset

- 3-Nearest Neighbors, $\ell_2$-norm distance on $x[n]$.
  \textbf{accuracy} : 0.77%
- 3-Nearest Neighbors, $\ell_2$-norm distance on $|X[k]|$.
  \textbf{accuracy} : 0.85%
- 1D Convolutional Net
  \textbf{accuracy} : 91%
- Wavelet Transform Features (entropy, zero crossing, simple statistics) + linear classifier
  \textbf{accuracy} : 95%
Application: Arrhythmia Detection

Normal heart rhythm

Irregular heart rhythm
Application: Arrhythmia Detection

Figure 3. The decomposition process of the 8-level WMRA.

Figure 4. Low-dimensional feature vector generated by PCA using wavelet coefficients.
# Application: Arrhythmia Detection

<table>
<thead>
<tr>
<th>Literatures and feature extraction methods</th>
<th>Feature selection (dimension)</th>
<th>Beat types</th>
<th>Training/test beats</th>
<th>Classifiers</th>
<th>Independent training/test data</th>
<th>k-fold cross validation</th>
<th>SEN (%)</th>
<th>SPE (%)</th>
<th>ACC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral correlation¹</td>
<td>Yes (88)</td>
<td>5</td>
<td>Totally 6259</td>
<td>SVM</td>
<td>Unknown</td>
<td>10-fold</td>
<td>99.20</td>
<td>99.70</td>
<td>98.60</td>
</tr>
<tr>
<td>Wavelet transform, morphological features³</td>
<td>No (28)</td>
<td>5</td>
<td>10675/93894</td>
<td>Artificial neural network</td>
<td>No</td>
<td>88.60</td>
<td>96.18</td>
<td>97.86</td>
<td></td>
</tr>
<tr>
<td>Morphological features⁵</td>
<td>Yes (6)</td>
<td>6</td>
<td>35848/35848</td>
<td>Linear discriminant analysis</td>
<td>No</td>
<td>91.19</td>
<td>98.65</td>
<td>94.03</td>
<td></td>
</tr>
<tr>
<td>Morphological features⁶</td>
<td>No (13)</td>
<td>3</td>
<td>600/30273</td>
<td>SVM, neural network</td>
<td>No</td>
<td>98.52</td>
<td>99.19</td>
<td>97.14</td>
<td></td>
</tr>
<tr>
<td>Time domain features⁷</td>
<td>No (9)</td>
<td>6</td>
<td>42427/14142</td>
<td>Decision tree</td>
<td>No</td>
<td>97.50</td>
<td>99.80</td>
<td>99.51</td>
<td></td>
</tr>
<tr>
<td>Morphological features⁸</td>
<td>No (16)</td>
<td>3</td>
<td>15509/8081</td>
<td>SVM, neural network</td>
<td>Yes</td>
<td>92.82</td>
<td>93.74</td>
<td>92.85</td>
<td></td>
</tr>
<tr>
<td>Morphological features¹¹</td>
<td>No (8)</td>
<td>5</td>
<td>12570/12570</td>
<td>Regression neural network</td>
<td>No</td>
<td>85.50</td>
<td>99.40</td>
<td>94.40</td>
<td></td>
</tr>
<tr>
<td>Fourier transform, wavelet package¹²</td>
<td>Yes (70)</td>
<td>16</td>
<td>3345/2542</td>
<td>k-NN</td>
<td>No</td>
<td>85.59</td>
<td>99.56</td>
<td>93.59</td>
<td></td>
</tr>
<tr>
<td>Wavelet transform, cosine transform¹³</td>
<td>Yes (18)</td>
<td>4</td>
<td>720/360</td>
<td>SVM</td>
<td>Unknown</td>
<td>96.50</td>
<td>98.60</td>
<td>95.50</td>
<td>96.50</td>
</tr>
<tr>
<td>Wavelet transform¹⁴</td>
<td>Yes (24)</td>
<td>5</td>
<td>900/900</td>
<td>SVM, genetic algorithm</td>
<td>No</td>
<td>98.50</td>
<td>99.69</td>
<td>98.80</td>
<td></td>
</tr>
<tr>
<td>Higher order spectral¹⁷</td>
<td>No (7)</td>
<td>5</td>
<td>330/500</td>
<td>SVM</td>
<td>Unknown</td>
<td>85.79</td>
<td>90.00</td>
<td>87.93</td>
<td>85.79</td>
</tr>
<tr>
<td>Wavelet transform¹⁹</td>
<td>Yes (20)</td>
<td>4</td>
<td>360/360</td>
<td>SVM</td>
<td>Unknown</td>
<td>98.62</td>
<td>98.62</td>
<td>98.61</td>
<td>98.61</td>
</tr>
<tr>
<td>Temporal and spectral features³⁰</td>
<td>Yes (15)</td>
<td>6</td>
<td>1440/720</td>
<td>SVM</td>
<td>No</td>
<td>97.60</td>
<td>93.80</td>
<td>95.20</td>
<td></td>
</tr>
<tr>
<td>Temporal and spectral features³¹</td>
<td>Yes (13)</td>
<td>8</td>
<td>Totally 17857</td>
<td>SVM</td>
<td>No</td>
<td>5-fold</td>
<td>95.00</td>
<td>99.00</td>
<td>98.60</td>
</tr>
<tr>
<td>Higher order statistics, wavelet packet³²</td>
<td>Yes (28)</td>
<td>5</td>
<td>3345/2542</td>
<td>k-NN</td>
<td>Yes</td>
<td>89.80</td>
<td>97.80</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Hilbert-Huang transform¹²</td>
<td>Yes (18)</td>
<td>6</td>
<td>10700/10700</td>
<td>SVM</td>
<td>No</td>
<td>98.64</td>
<td>99.77</td>
<td>99.51</td>
<td></td>
</tr>
<tr>
<td>Wavelet transform⁴⁰</td>
<td>Yes (18)</td>
<td>5</td>
<td>Totally 101352</td>
<td>SVM</td>
<td>Yes</td>
<td>44-fold</td>
<td>—</td>
<td>—</td>
<td>86.40</td>
</tr>
<tr>
<td>Approximate entropy, wavelet packet⁴⁷</td>
<td>Yes (9)</td>
<td>5</td>
<td>145/145</td>
<td>SVM, PNN</td>
<td>Unknown</td>
<td>98.70</td>
<td>99.70</td>
<td>98.60</td>
<td></td>
</tr>
<tr>
<td>Non-linear and center-clipping transform⁴⁸</td>
<td>No (5)</td>
<td>5</td>
<td>13640/13640</td>
<td>Wavelet neural network</td>
<td>No</td>
<td>98.78</td>
<td>99.70</td>
<td>98.78</td>
<td></td>
</tr>
<tr>
<td>Eigenvector method¹⁹</td>
<td>Yes (12)</td>
<td>4</td>
<td>360/360</td>
<td>Recurrent neural network</td>
<td>Unknown</td>
<td>98.89</td>
<td>99.25</td>
<td>98.06</td>
<td></td>
</tr>
<tr>
<td>Higher order statistics³⁰</td>
<td>No (24)</td>
<td>5</td>
<td>4000/14299</td>
<td>RBF neural network</td>
<td>No</td>
<td>92.93</td>
<td>98.52</td>
<td>95.18</td>
<td></td>
</tr>
<tr>
<td>Geometrical features³¹</td>
<td>No (18)</td>
<td>7</td>
<td>4035/3150</td>
<td>SVM, k-NN, BPNN</td>
<td>No</td>
<td>97.52</td>
<td>99.65</td>
<td>98.06</td>
<td></td>
</tr>
<tr>
<td>Wavelet transform, morphological features⁴⁰</td>
<td>Yes (8)</td>
<td>3</td>
<td>50928/49636</td>
<td>Linear discriminant analysis</td>
<td>Yes</td>
<td>80.00</td>
<td>—</td>
<td>94.00</td>
<td></td>
</tr>
<tr>
<td>Wavelet transform, linear prediction model³⁵</td>
<td>No (12)</td>
<td>3</td>
<td>50554/49273</td>
<td>Linear discriminant analysis</td>
<td>Unknown</td>
<td>86.50</td>
<td>—</td>
<td>86.50</td>
<td></td>
</tr>
<tr>
<td>Cross correlation³⁴</td>
<td>No (30)</td>
<td>3</td>
<td>41961/51285</td>
<td>Artificial neural network</td>
<td>Unknown</td>
<td>97.49</td>
<td>—</td>
<td>95.24</td>
<td></td>
</tr>
<tr>
<td>WMRA [This work]</td>
<td>Yes (12)</td>
<td>6</td>
<td>Totally 107049</td>
<td>SVM</td>
<td>Yes</td>
<td>10-fold</td>
<td>99.09</td>
<td>99.82</td>
<td>99.70</td>
</tr>
</tbody>
</table>
Limitations of the Fourier Transform

- **Time Domain (Stationary Process)**
  - Amplitude vs. Time

- **Time Domain (Non-Stationary Process)**
  - Amplitude vs. Time with a burst indicated

- **Frequency Domain (Stationary Process)**
  - Amplitude vs. Frequency

- **Frequency Domain (Non-Stationary Process)**
  - Amplitude vs. Frequency with burst frequencies highlighted
Non-stationary signals
Chirp Signals

- linear chirp $x(t) = \sin(2\pi (ct^2 + f_0 t))$
Chirp Signals

- quadratic chirp \( x(t) = \sin(2\pi(ct^3 + dt^2 + f_0t)) \)
Resolving Signal Components

- Sum of two hyperbolic chirps: one with instantaneous frequency \( \frac{7.5}{(0.80-t)^2} \) and one with \( \frac{2.5}{(0.80-t)^2} \) sampled at 2048Hz
Discrete Fourier Transform
Short-Time Fourier Transform - long window
Short-Time Fourier Transform - short window
Continuous Wavelet Transform

Magnitude Scalogram
Human Activity Recognition (HAR) Dataset
Continuous Wavelet Transform of HAR signals

- changing the number of scales
Continuous Wavelet Transform of HAR signals
Deep Convolutional Neural Network

Confusion Matrix

Walking  Walking upstairs  Walking downstairs  Sitting  Standing  Laying
Walking  472          0             0           0        0       0
Walking upstairs  16         466           4           1        1       0
Walking downstairs  8          1           416         0        0       0
Sitting    0           1            0           423      48      0
Standing    0           0            0           67       483     0
Laying     0           3            0           0        0       537

94.91% test accuracy
1. input signal $x = [x[1], \ldots, x[N]]$
2. compute discrete wavelet transform $y = Wx$
3. perform thresholding in the wavelet domain (shrink coefficients by hard/soft thresholding)
4. reconstruct the signal from thresholded discrete wavelet coefficients $\hat{x} = W^{-1}S(y)$
Hard/Soft Thresholding
Hard/Soft Thresholding

- hard thresholding
  \[ H_\lambda(y) = y 1[|y| > \lambda] \]

- soft thresholding
  \[ S_\lambda(y) = \text{sign}(y)(|y| - \lambda)_+ \]
Sparse Signal Recovery and Optimization

- noisy observation model

\[ y = y^* + \sigma w \]

\[ w \sim N(0, 1) \text{ independent identically distributed Gaussian noise} \]

- \( y^* \) is a sparse clean signal

- recover \( y^* \) from \( y \) via

  - hard thresholding:
    \[ \hat{y} = H_\lambda(y) = \arg\min_z \|y - z\|_2^2 + \lambda \|z\|_0 \]

  - soft thresholding:
    \[ \hat{y} = S_\lambda(y) = \arg\min_z \|y - z\|_2^2 + \lambda \|z\|_1 \]
Wavelet Denoising
Wavelet Denoising
Wavelet Denoising
Hard vs Soft Thresholding

The original signal

The signal with noise

Hard threshold function

Soft threshold function
Image Denoising by Wavelet Thresholding

- 2D Biorthogonal mother wavelet