Lecture 25: Wavelets for signal processing

Reading: ESL 5.9

STATS 202: Data mining and analysis

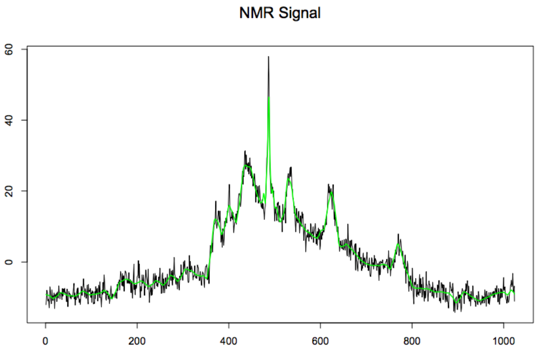
Sergio Bacallado
November 20, 2013

Friday's lecture

Lester Mackey:

Predicting ALS Disease Progression with Bayesian Additive Regression Trees.

Wavelet smoothing



Goal: Find a smoother approximation of this signal, which has a *sparse* representation.

Why would we want to smooth?

This is often useful in spatial data, time series, and image analysis.

1. We want to analyze patterns in a financial time series. Which trends are real and which are noise?
2. A seismic study produces a map of the density of the underground for oil exploration. If the density measurement is noisy, what is the best estimate we can make?
3. An MRI produces a 3D image of water density in the brain. What is the best estimate of the actual density?
4. Image and video compression.

Basis functions

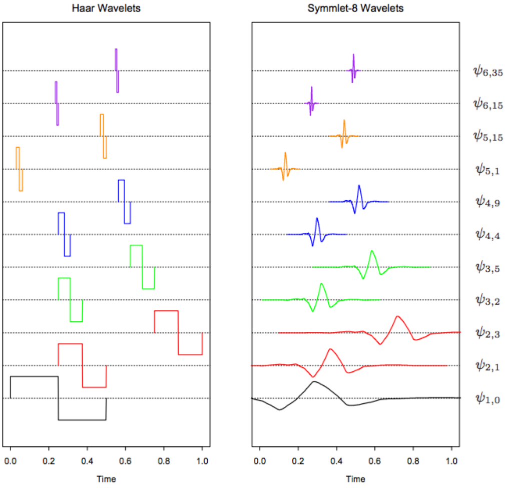
Cubic splines are a way to smooth one-dimensional data:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \xi_1)_+^3 + \cdots + \beta_{K+3} (x - \xi_K)_+^3.$$

If we choose a good set of knots, this is a *sparse* representation.
We are compressing n samples into $K + 4$ parameters.

Wavelet smoothing simply uses a different set of basis functions.

Wavelet basis

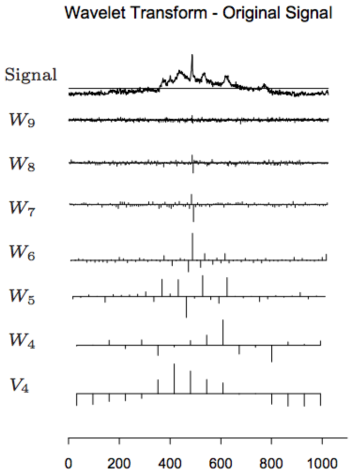


- Each basis function $\psi_{i,j}$ is indexed by a scale i and a location j.
- The functions $\psi_{i,j}$ are orthonormal:

$$\int \psi_{i,j}(x)\psi_{i',j'}(x)dx$$

$$= \begin{cases} 1 & \text{if } i = i', j = j' \\ 0 & \text{otherwise} \end{cases}$$

NMR example

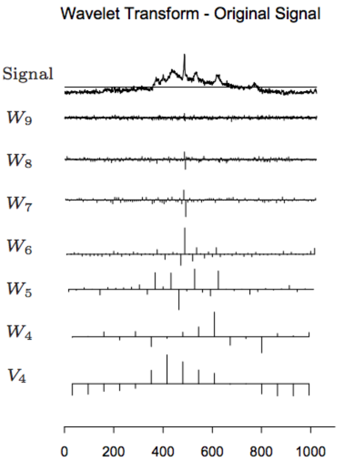


Wavelet Transform - Original Signal
Signal
W_9
W_8
W_7
W_6
W_5
W_4
V_4
0
200
400
600
800
1000

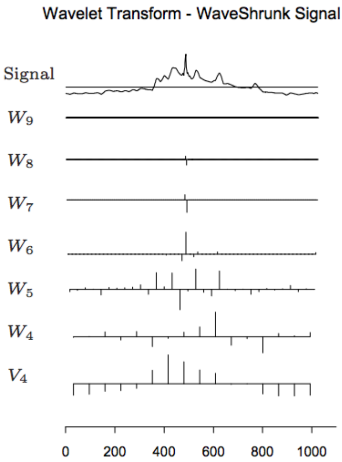
Advantages of wavelets

1. They use a complete orthogonal basis, so:
   - We can represent almost any function.
   - The coefficients are unique.
2. Unlike the Fourier Transform, each wavelet represents an effect localized in time and frequency. This leads to sparser representations for many signals.
3. They can be generalized to 2 or 3 dimensions.
4. They have vanishing moments, which means that polynomials of order d can be represented exactly with wavelets of limited resolution.
5. By shrinking the coefficients, we can denoise the signal and obtain a sparse representation!

NMR example



Wavelet Transform - Original Signal
Signal
W_9
W_8
W_7
W_6
W_5
W_4
V_4
0
200
400
600
800
1000



Wavelet Transform - WaveShrunk Signal
Signal
W_9
W_8
W_7
W_6
W_5
W_4
V_4
0
200
400
600
800
1000

SURE Shrinkage

If we minimize the residual sum of squares:

$$\min_{\beta} \|y - \mathbf{W}\beta\|_2^2,$$

where the columns of $\mathbf{W}$ are wavelets evaluated at the n samples $x_1, \ldots, x_n$, we obtain the wavelet transform:

$$\hat{\beta} = \mathbf{W}^T y.$$

The Stein Unbiased Risk Estimate (SURE) solves the penalized problem:

$$\min_{\beta} \|y - \mathbf{W}\beta\|_2^2 + 2\lambda\|\beta\|_1.$$

SURE Shrinkage

The Stein Unbiased Risk Estimate (SURE) solves the problem:

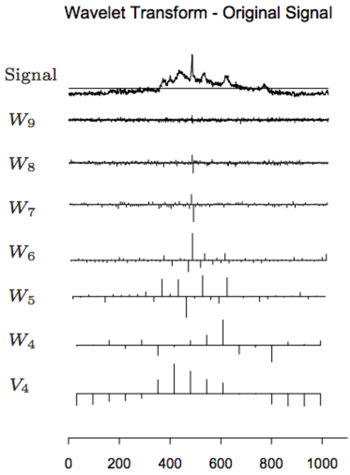
$$\min_{\beta} \|y - \mathbf{W}\beta\|_2^2 + 2\lambda\|\beta\|_1.$$

This is just a Lasso problem. Since the columns of $\mathbf{W}$ are orthogonal, the solution is easy to derive:

$$\hat{\beta}_j^{(\lambda)} = \mathsf{sign}(\hat{\beta}_j)(|\hat{\beta}_j| - \lambda)_+$$

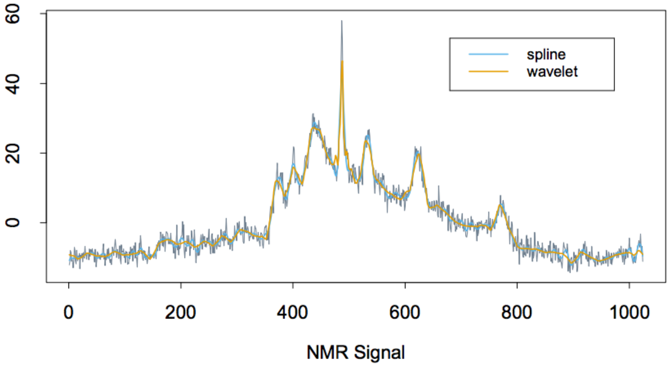
This is a soft thresholding of the wavelet transform $\hat{\beta}$.

NMR example

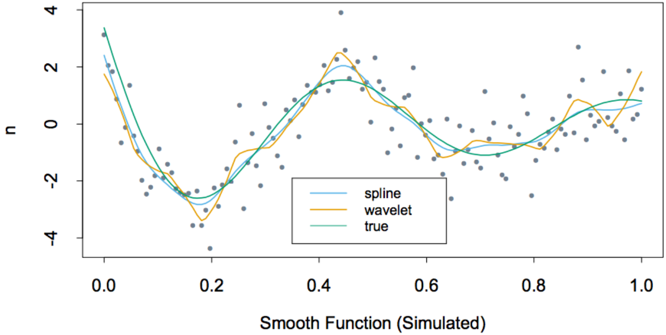


Wavelet Transform - Original Signal
Signal
W_9
W_8
W_7
W_6
W_5
W_4
V_4
0
200
400
600
800
1000

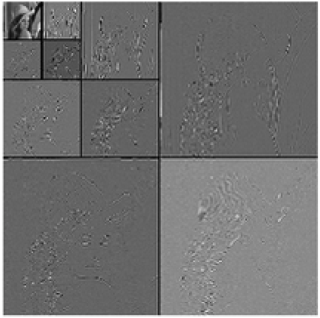
Wavelets vs. splines



Wavelets vs. splines



Two dimensional wavelets



Other useful bases

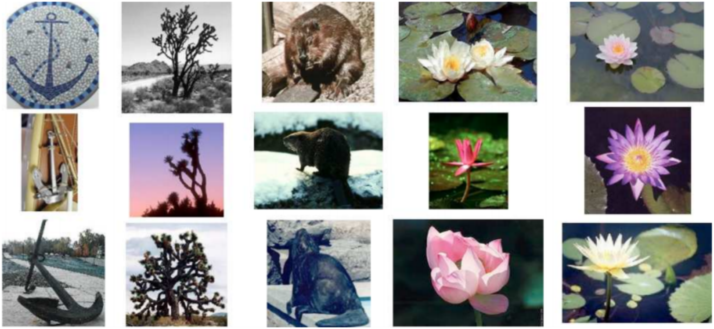
- Curvelets
- Bandelets
- Chirplets
- Contourlets
- ...

Prediction using complex signals

Digit recognition: Classify the images into digits.

Prediction using complex signals

Caltech dataset: Classify the images into categories.



Prediction using complex signals

Challenge: Summarize the information in each signal into a set of relevant features (featurization).

Ideas based on wavelets:

- Use the spectrum of the signal, i.e. a histogram of the frequencies present in the signal.
- The **scattering transform** is a recent method that can produce a representation of the data which is:
  - Invariant to translations.
  - Robust to deformations.