University of BRISTOL

Computer Science Year 2

Algorithms & Data

Estimation, Regression, Classification

Prof Alin Achim

University of BRISTOL

Contact Details

Office:
Room 4.09
Merchant Venturers
Building

Tel:
+44 (0)117 456 1359

E-mail:
Alin.Achim@bristol.ac.uk

Course Delivery

- Lectures - please take notes
  - Lecture slides
  - Examples
- Lab
  - One hour session per week
- BlackBoard
- Github

General Outline

- Data: Estimation, Regression, Classification
  - Mathematical Preliminaries
    - Linear Algebra
    - Random variables and random processes
  - Estimation Theory
    - MVUE & CRLB
    - Maximum Likelihood Estimation (MLE)
    - Method of Moments
    - Least squares
  - Regression, Loss, and Curve Fitting
  - Classification
    - Bayes Classifiers

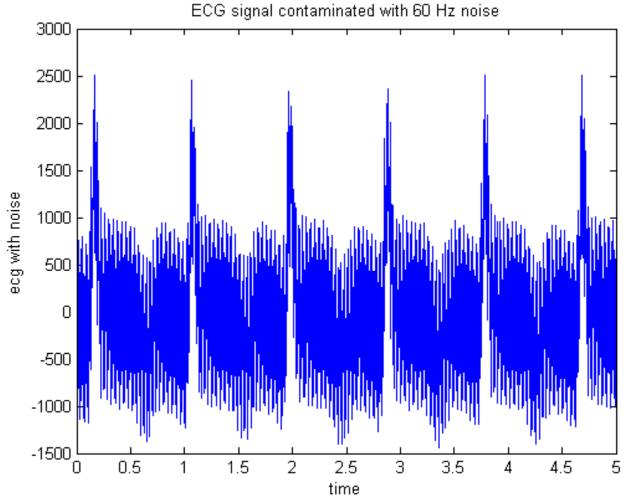
Definitions

- Signals and Signal Processing
- Data & Information
- Estimation Theory
- Pattern Recognition

- Statistical Signal (Data) Processing (SSP)
  - Digital → sampled, discrete time, quantized
  - Signal → waveform, sequence of measurements or observations
  - Processing → analyze, modify, filter, synthesize
- Examples
  - Sampled speech waveform
  - ECG
  - “pixelized” image
- SSP applications
  - Filtering (noise reduction)
  - Pattern recognition (speech, faces, fingerprints)
  - Compression

A major difficulty!!

- In many (perhaps most) SSP applications we don't have complete or perfect knowledge of the signals we wish to process. We are faced with many unknowns and uncertainties.



Challenges are measurement noise and intrinsic uncertainties in signal behaviour

A major difficulty!!

- Examples
- noisy measurements
- unknown signal parameters
- noisy system or environmental conditions
- natural variability in the signals encountered

- Questions:
  - How can we design data processing and analysis algorithms in the face of such uncertainty?
  - Can we model the uncertainty and incorporate this model into the design process?
- The answer: Statistical signal processing is concerned with the study of these questions.

Modelling uncertainties

- The most widely accepted and commonly used approach to modelling uncertainty is Probability Theory (although other alternatives exist such as *Fuzzy Logic*).
  - Probability Theory models uncertainty by specifying the chance of observing certain signals.
  - Alternatively, one can view probability as specifying the degree to which we believe a signal reflects the true state of nature.
- Examples of Probabilistic Models

  • errors in a measurement (due to an imprecise measuring device) modelled as realizations of a Gaussian random variable.

  • uncertainty in the phase of a sinusoidal signal modelled as a uniform random variable on $[0, 2\pi)$.

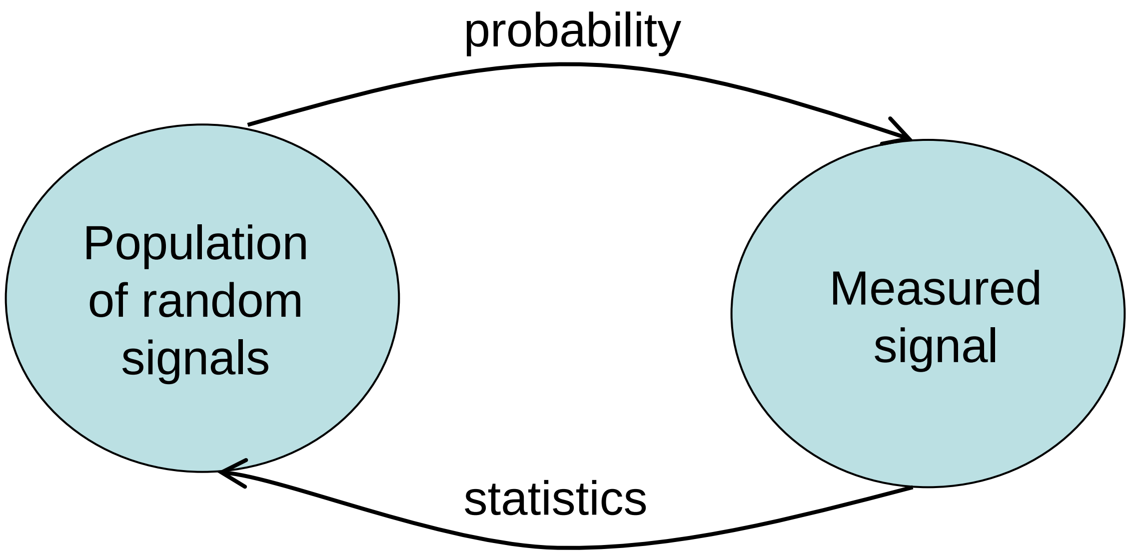
  • uncertainty in the number of photons striking a CCD per unit time modelled as a Poisson random variable.

Statistical Inference

- A statistic is a function of the observed data.
- Example: Suppose we observe N scalar values $x_1, x_2, \dots x_N$. The following are statistics:
  - Sample mean: $$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$
  - The data itself: $x_1, x_2, \dots x_N$
  - Order statistic: $\min\{x_1, x_2, \dots x_N\}$
- A statistic CANNOT depend on unknown parameters!!

Statistical Inference

- Probability is used to model uncertainty
- Statistics are used to draw conclusions about probability models



- **Probability models our uncertainty about signals we may observe.**
- **Statistics reason from the measured signal to the population of possible signals.**

Statistical Signal Processing (SSP)

- A Three-step approach:
  - Step 1 - Postulate a probability model (or models) that reasonably capture the uncertainties at hand
  - Step 2 - Collect data
  - Step 3 - Formulate statistics that allow us to interpret or understand our probability model(s)
- There are two major kinds of problems that are studied: detection and estimation. Most SSP problems fall under one of these two headings.

Detection Theory

- Given two (or more) probability models, which one best explains the signal?
- Examples:
  - Decode wireless comm signal into string of 0's and 1's
  - Pattern recognition
    - voice recognition
    - face recognition
    - handwritten character recognition
  - Anomaly detection
    - radar, sonar
    - irregular heartbeats in ECG signals
    - Extragalactic point sources (EPS) in CMB maps

Detection example

- Suppose we observe N tosses of an unfair coin. We would like to decide which side the coin favours, heads or tails.

o Step 1 - Assume each toss is a realization of a Bernoulli random variable:

$$\Pr[\text{Heads}] = p = 1 - \Pr[\text{Tails}]$$

The detection problem consists now in deciding p=1/4 vs p=3/4

o Step 2 – Collect data $x_1, x_2, \dots x_N$

$$X_i=1 \rightarrow \text{Heads}$$

$$X_i=0 \rightarrow \text{tails}$$

o Step 3 – Formulate a useful statistic $k = \sum_{n=1}^{N} x_n$

If k<N/2 guess p=1/4; if k>N/2 guess p=3/4

Estimation Theory

- If our probability model has free parameters, what are the best parameter settings to describe the signal we have observed?
- Examples
  - Noise reduction
  - Determine parameters of a sinusoid (phase, amplitude, frequency)
  - Adaptive filtering
    - track trajectories of space-craft
    - automatic control systems
    - channel equalization
  - Determine location of a submarine (sonar)
  - Seismology: estimate depth below ground of an oil deposit

Estimation example

- Suppose we take N measurements of a DC voltage A with a noisy voltmeter. We would like to estimate A.
- Step 1 - Assume a Gaussian noise model

$$x_N = A + w_N \text{ where } w_N \sim N(0,1)$$

- Step 2 – Collect data $x_1, x_2, \ldots x_N$
- Compute the sample mean

$$\hat{A} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

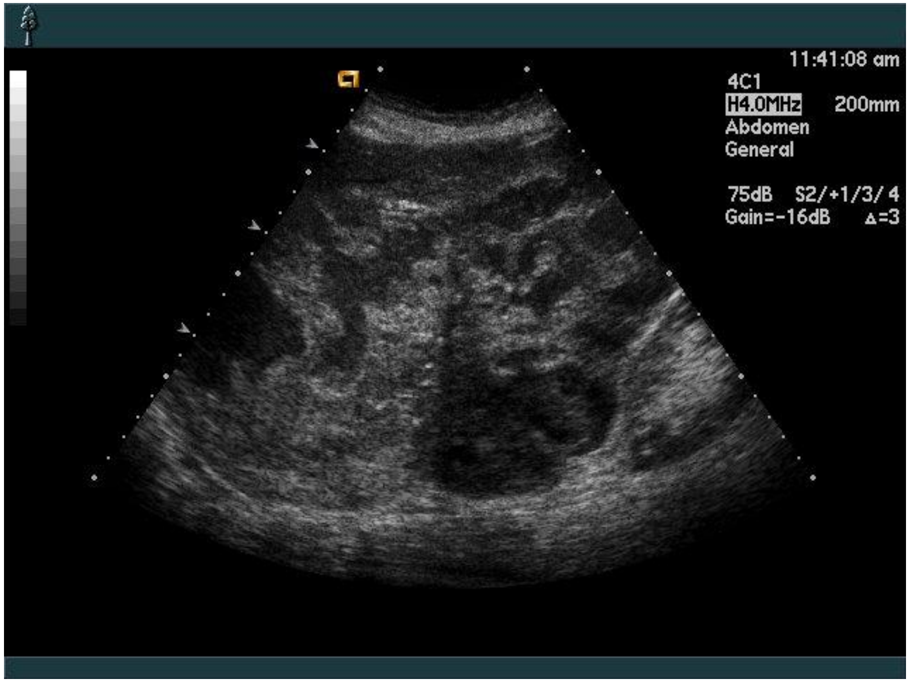
and use this as the estimate!

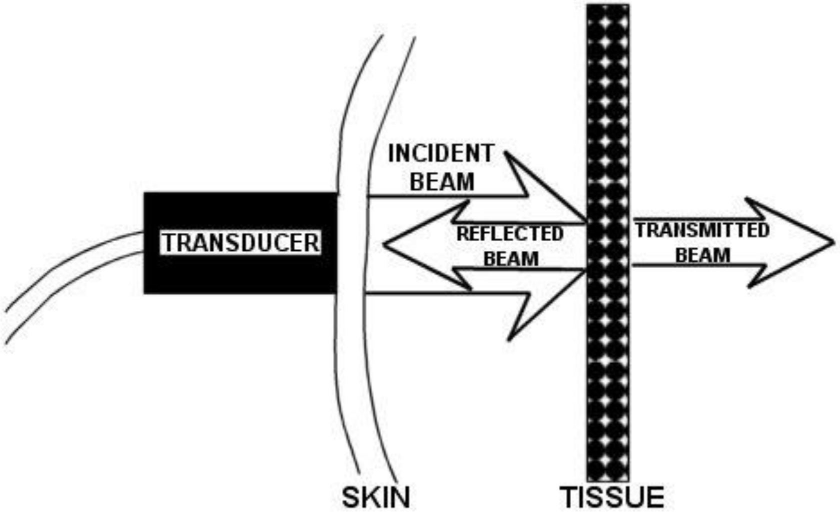
Parameter Estimation Applications

- Parameter estimation is an essential task in many signal and data processing applications
  - Range estimation in radar, sonar, navigation (robotics, aerospace vehicles guidance)
  - Direction of arrival estimation
  - Frequency domain and spectral analysis (e.g. medical images)
  - Estimation of model parameters in the design of different types of classifiers

Application to Ultrasound Image Despeckling

Medical ultrasonography produces a two-dimensional (2-D) signal. The ultrasound signal relies on the backscattering of acoustic waves.



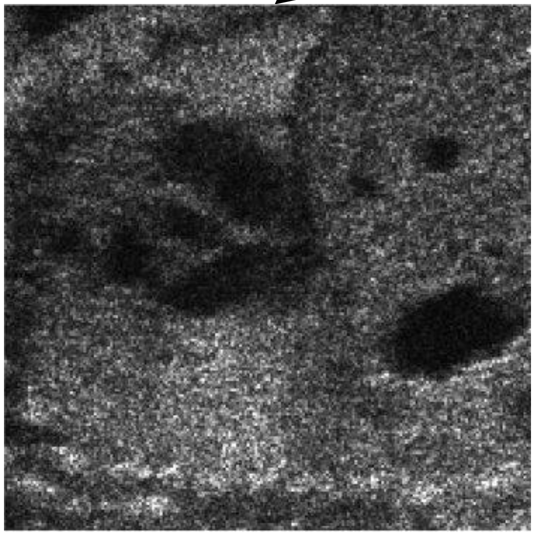


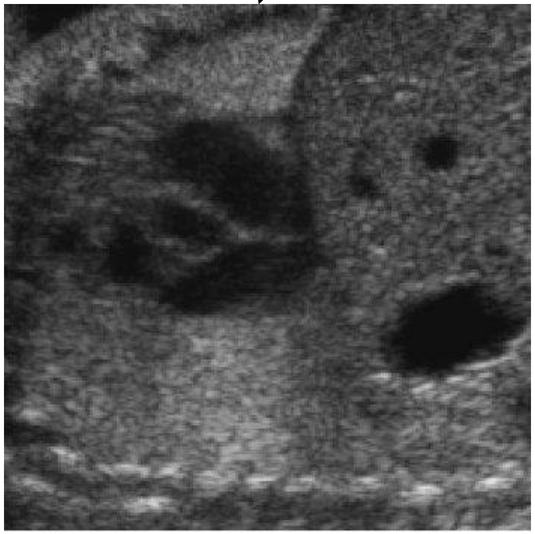
The principle of sonar (pulse-echo) is used to produce cross-sectional images of various organ and tissue interfaces in the body which are able to reflect high-frequency sound. The image above was obtained at a routine monitoring examination and represents a so-called 'B-mode' image.

University of BRISTOL

(Inverse) Problem

$$I(x,y) = S(x,y) \cdot \eta(x,y)$$





Speckle Noise (multiplicative): unit-mean, log-normal distributed.

Need to balance between speckle suppression and signal detail preservation!!!

The Symmetric Alpha-Stable (SαS) Model

SαS Characteristic Function:

$$\boxed{\varphi(\omega) = e^{-\gamma|\omega|^{\alpha}}}$$

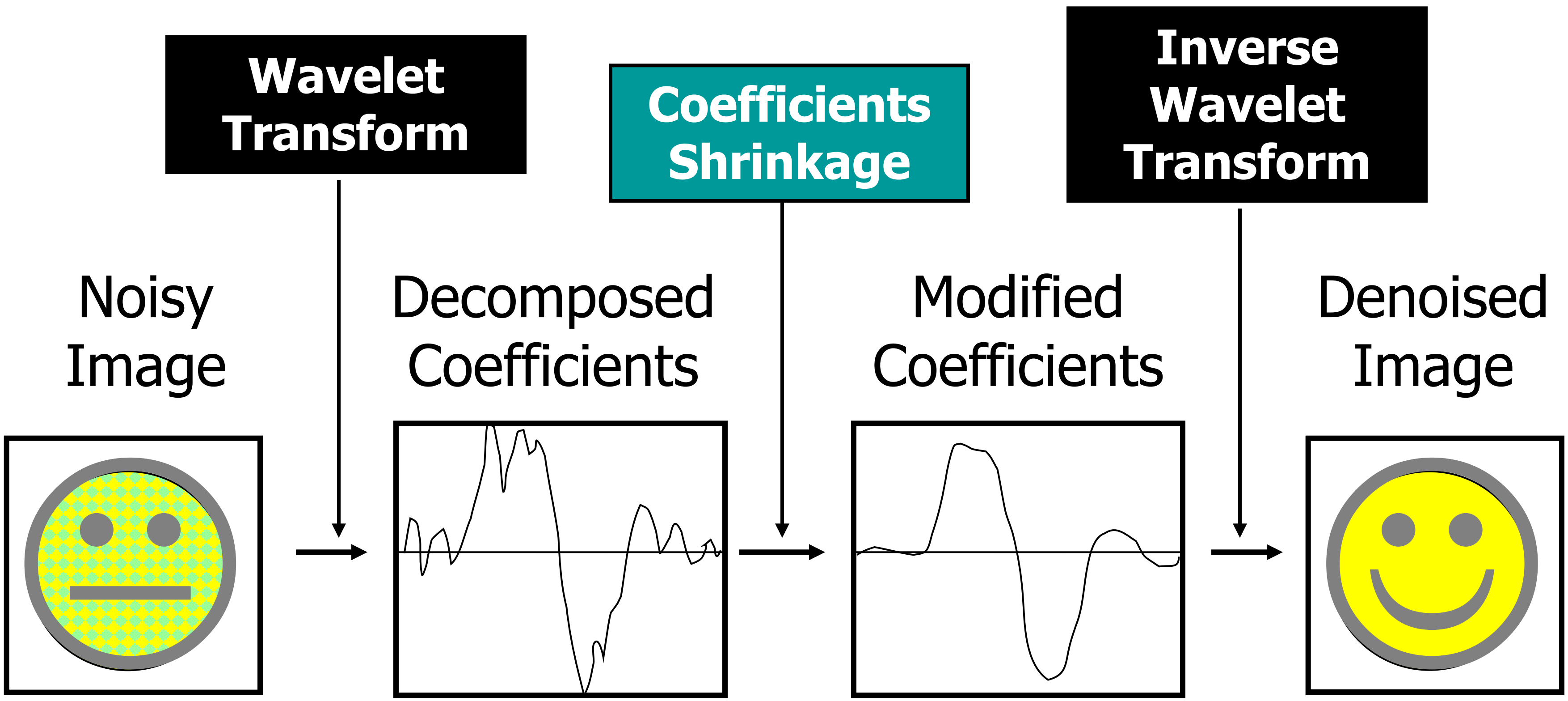
<u>α: characteristic exponent, 0<α<=2</u> *(determines thickness of the distribution tails, α=2: Gaussian, α=1: Cauchy)*

<u>γ: dispersion parameter</u>

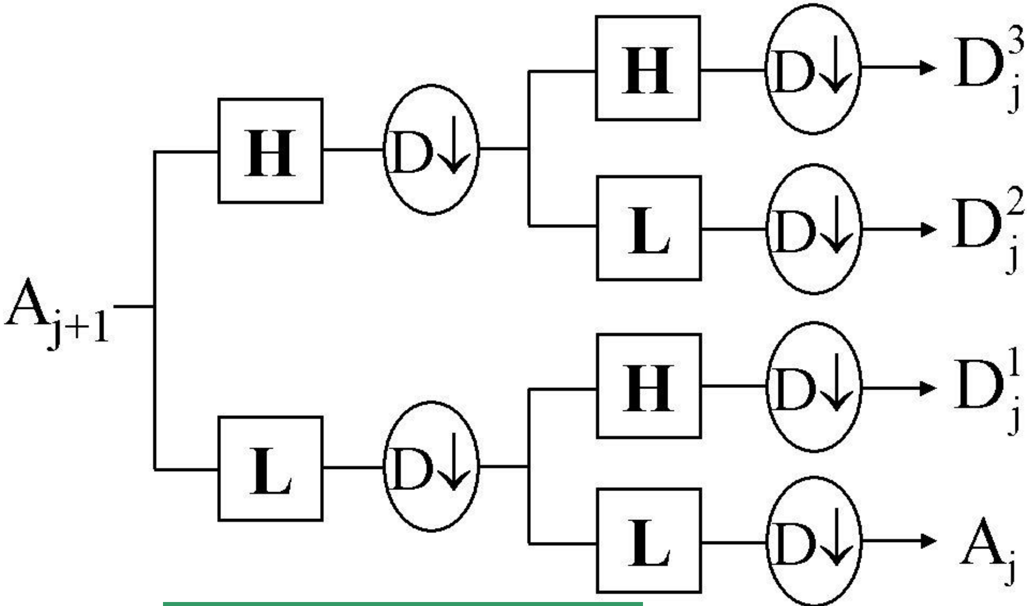
for Gaussian → variance = 2 x γ

for Cauchy → γ behaves like variance

Wavelets for Image Denoising

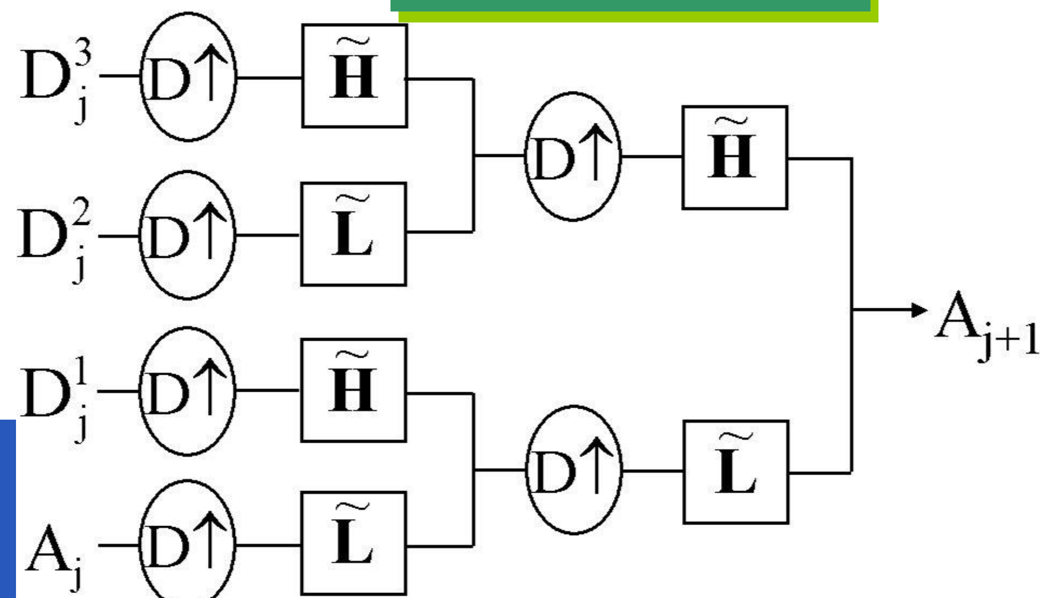


2-D Dyadic Wavelet Transform



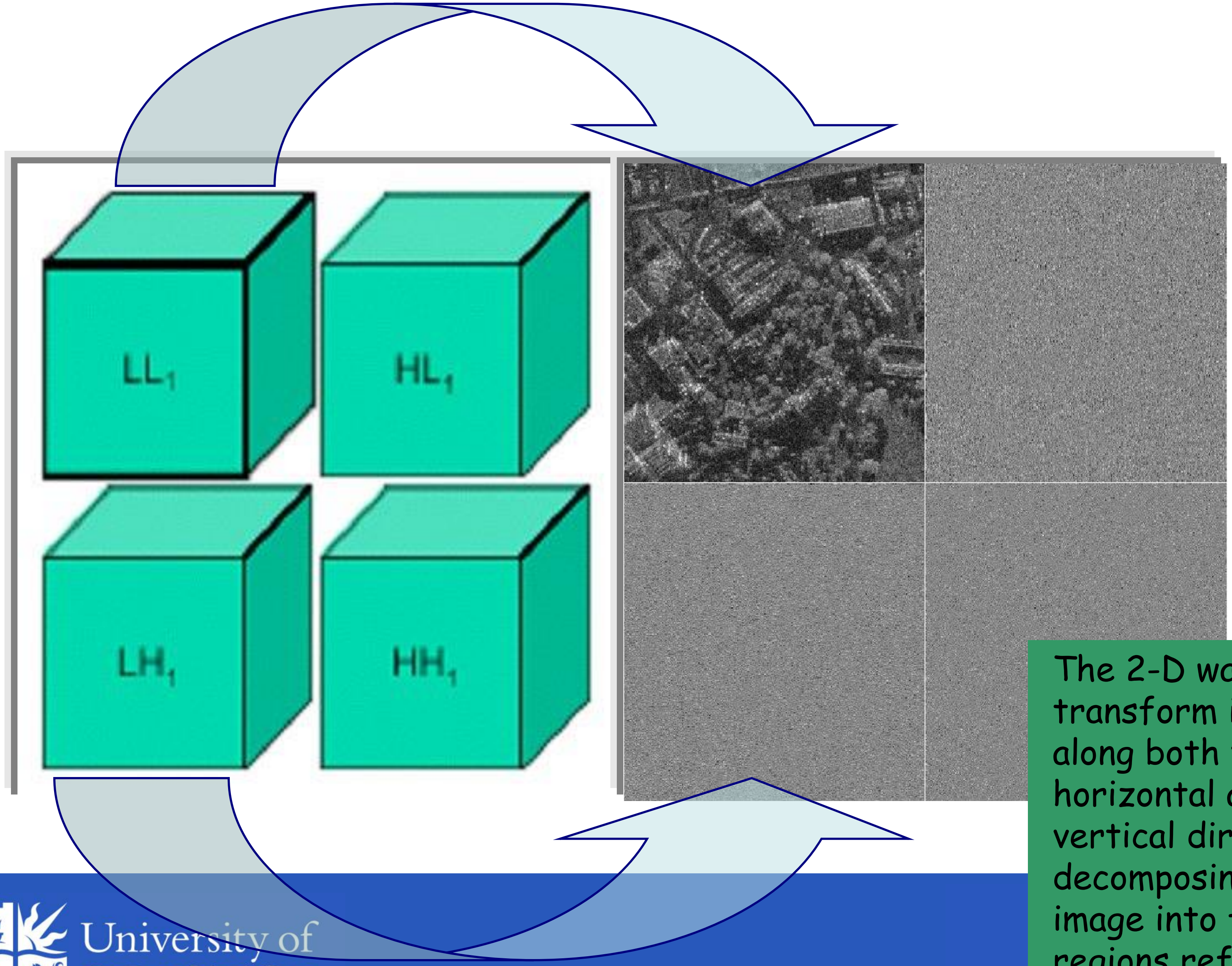
Decomposition of an image into an approximation and 3 detail subbands

Reconstruction of an image from its approximation and details



University of BRISTOL

Multiresolution decomposition – 1st level



The 2-D wavelet transform is applied along both the horizontal and vertical directions, decomposing the image into four regions referred as image subbands.

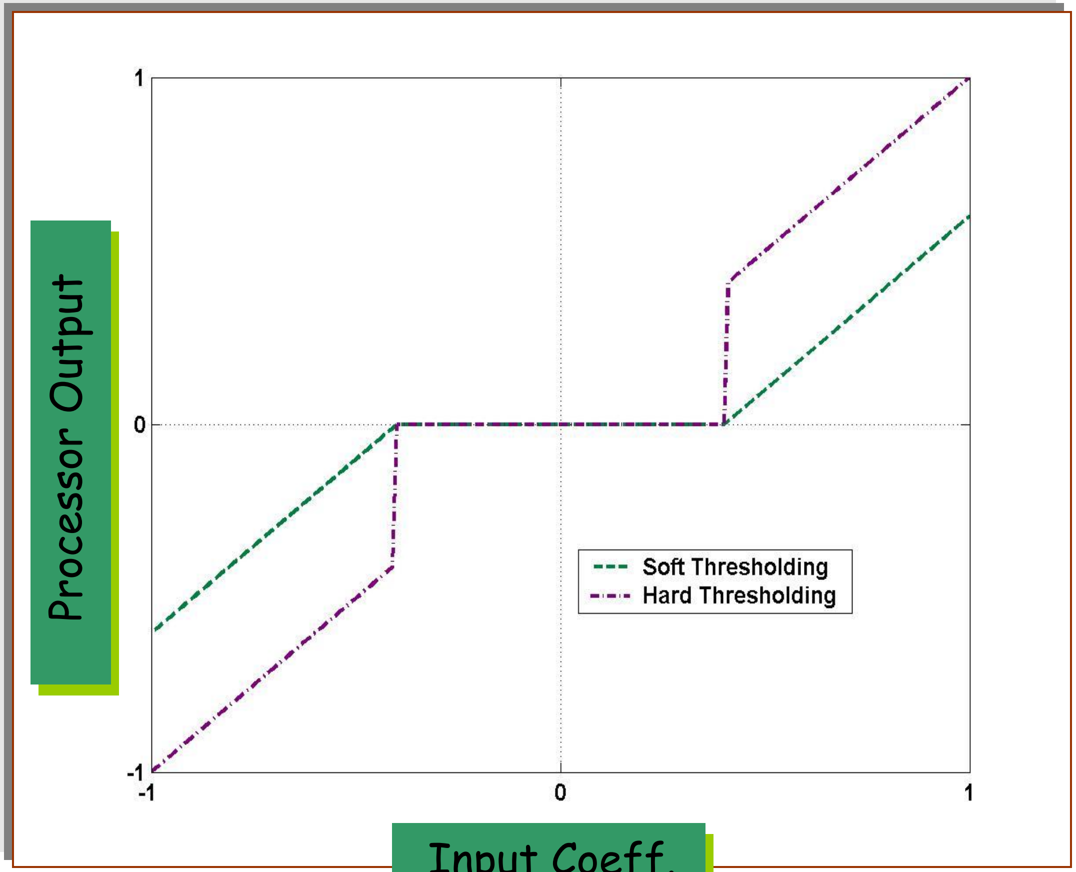
Wavelet Shrinkage Methods

- Soft Thresholding

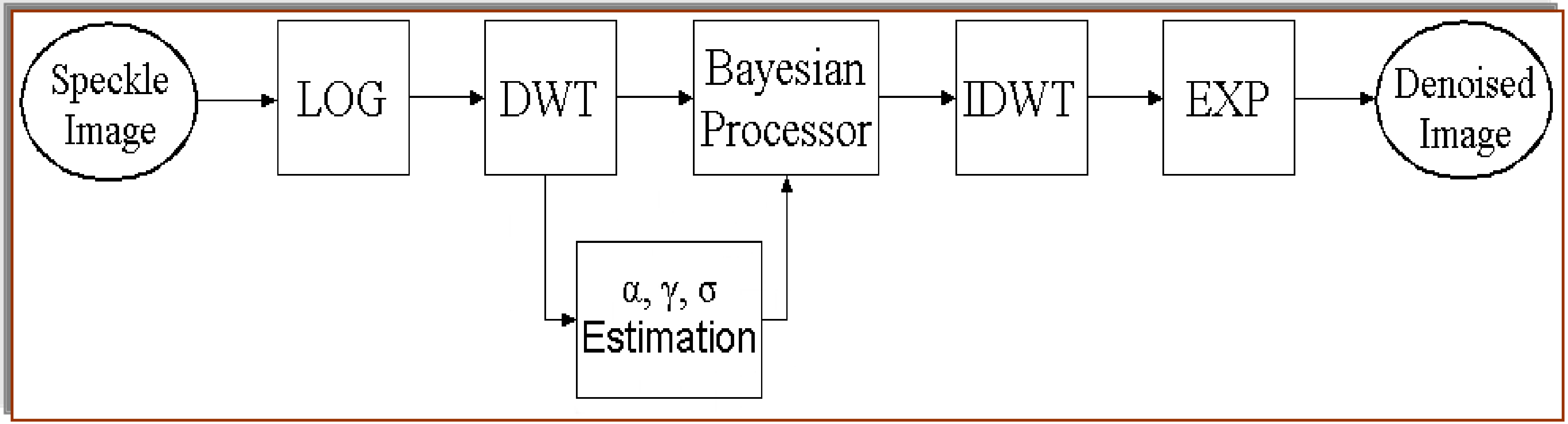
$$T_s^{soft}(s) = \begin{cases} \mathrm{sgn}(s)(|s| - t), & |s| > t \\ 0, & |s| \le t \end{cases}$$

- Hard Thresholding

$$T_s^{hard}(s) = \begin{cases} s, & |s| > t \\ 0, & |s| \le t \end{cases}$$



Alternative: Bayesian estimator of noise-free data



Estimator's fundamentals:

1. Wavelet transform the speckle ultrasound image.
2. SaS modeling of signal wavelet coefficients.
3. Bayesian processing of the wavelet coefficients.

The MAE Bayesian Estimator

- After applying the DWT: $d_{j,k}^{i} = s_{j,k}^{i} + \xi_{j,k}^{i}$
- The MAE estimator is the conditional median of s, given d, which coincides with the conditional mean (due to the symmetry of the distributions):

$$\hat{s}(d) = \int s \cdot P_{s|d}(s|d) \cdot ds = \frac{\int P_{\xi}(d-s)P(s)s \cdot ds}{\int P_{\xi}(d-s)P(s) \cdot ds}$$

- Noise Estimation: $\hat{\sigma} = \frac{1}{0.6745} MAD\left(\left\{d_{J,k},\ 0 \le k < 2^{J}\right\}\right)$

- Signal Parameter Estimation - by means of a LS fitting in the characteristic function domain:

$$\Phi(\omega) = \exp(-\gamma_s |\omega|^{\alpha_s}) \cdot \exp(-\frac{\sigma^2}{2}|\omega|^2)$$

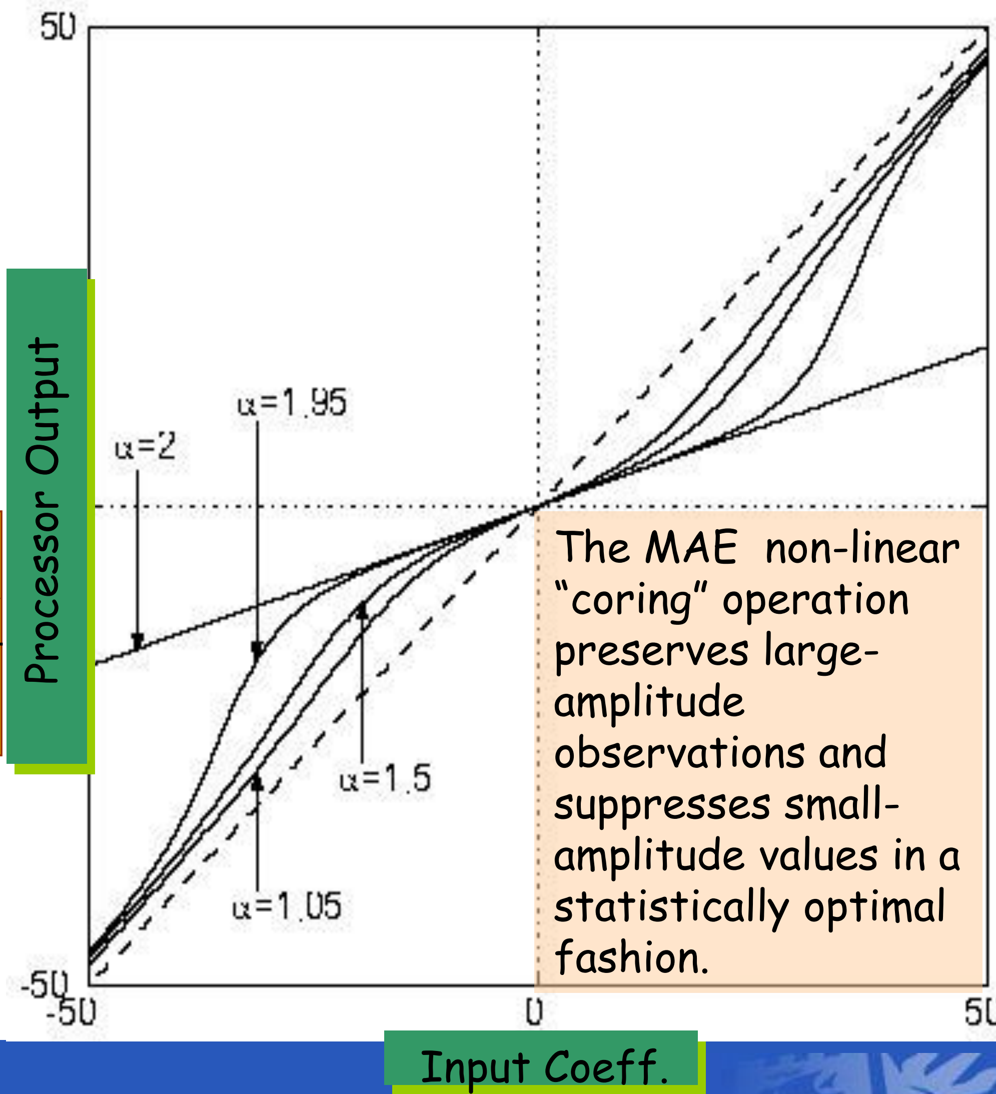
MAE Processor I/O Curves

- Wavelet domain measurements:

$$d_{j,k}^{i} = s_{j,k}^{i} + \xi_{j,k}^{i}$$

- Bayesian Estimation:

$$\hat{s}(d) = \int s \cdot P_{s|d}(s|d) \cdot ds = \frac{\int P_{\xi}(d-s)P(s)s \cdot ds}{\int P_{\xi}(d-s)P(s) \cdot ds}$$



The MAE non-linear "coring" operation preserves large-amplitude observations and suppresses small-amplitude values in a statistically optimal fashion.

The MAP Bayesian Estimator

• The MAP estimator is the Bayes risk estimator under an uniform cost function:

$$\widehat{s}\left(d\right) = \arg\max_{\widehat{s}} P_{s|d}(s|d) = \arg\max_{\widehat{s}} P_{d|s}(d|s)P(s) =$$

$$= \arg\max_{\widehat{s}} P_{\xi}(d-s)P_{s}(s) = \arg\max_{\widehat{s}} P_{\xi}(\xi)P_{s}(s)$$

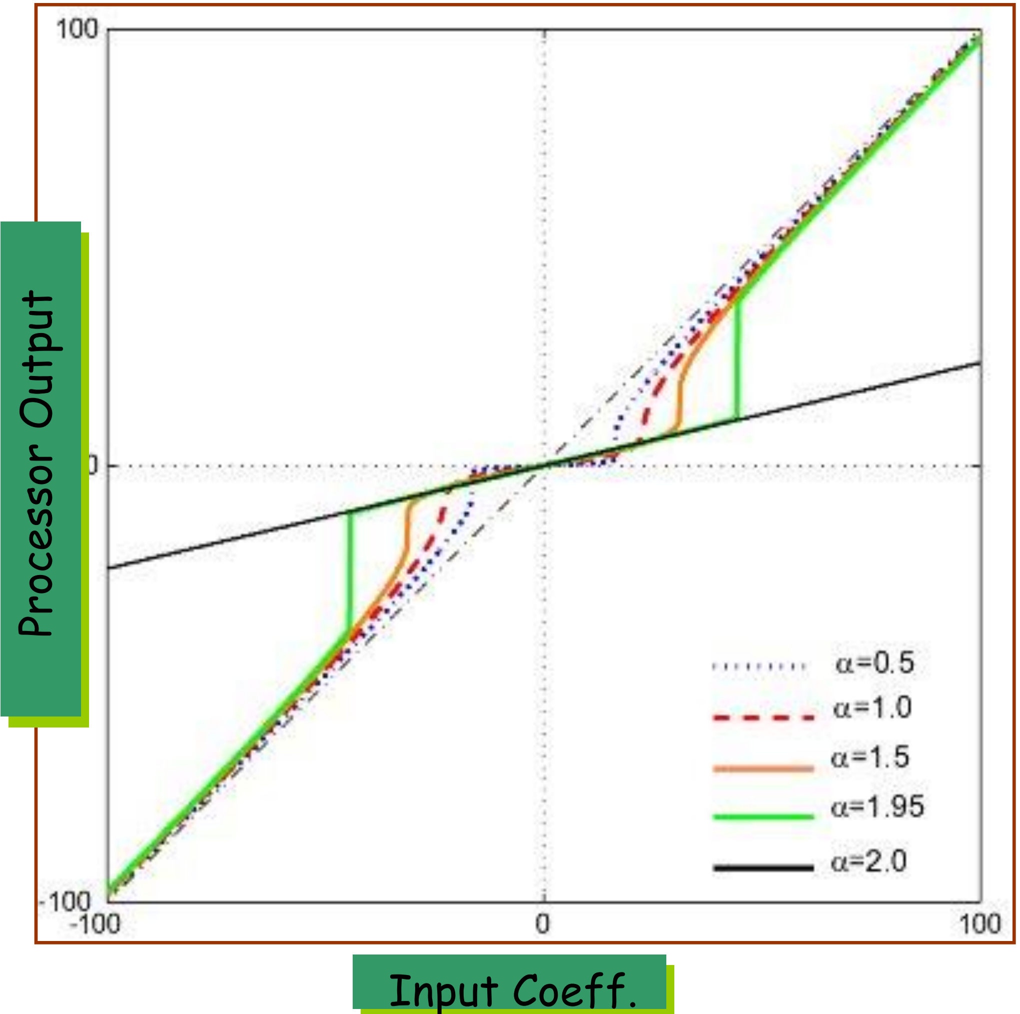
• Parameter estimation method: After estimating the level of noise σ we find the parameters α_s and γ_s by regressing

$$y = \log[-(\log|\Phi_d(\omega)|^2 + \sigma^2\omega^2)]$$

on $w = \log|\omega|$ in the model: $y_k = \mu + \alpha \cdot w_k + \varepsilon_k$

where: $\mu = \log(2\gamma)$, ε_k – error term, and $(\omega_\kappa, \kappa = 1,\ldots,K) \in \boldsymbol{R}$

MAP Processor I/O Curves

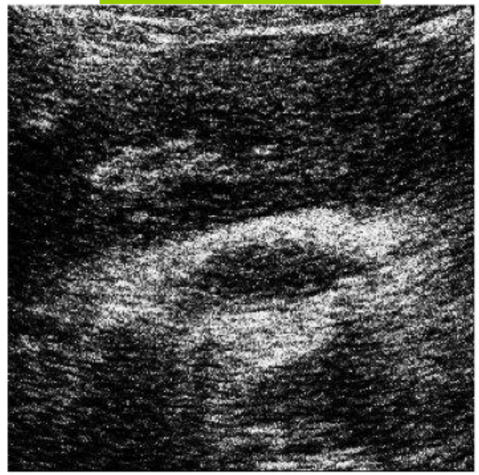


The plots illustrate the processor dependency on the parameter α of the signal *prior* PDF. For a given ratio γ/σ, the amount of shrinkage decreases as α decreases. The intuitive explanation for this behavior is that the smaller the value of α, the heavier the tails of the signal PDF and the greater the probability that the measured value is due to the signal.

Ultrasound Image Denoising Results

Original Image

Degraded Image



Soft Thresholding

Wiener

Bayesian