Lecture 7

Basis functions & decomposition
The generic basis model

We consider a model for the observations in terms of a function of a set of basis functions.

In the majority of methods, a **generic linear model** is tacitly assumed.

\[ x[t] = \sum_i w_i \phi_i(X) \]
What common models fall into this category?

Almost all!

- AR processes, including the Kalman process
- PCA, ICA, NMF
- Fourier basis, Wavelets
- Hilbert-Huang (EMD)
- ....
Fixed basis models

These include Fourier & Wavelet approaches and make the tacit assumption that the function can be modelled using a combination of basis functions that are **universal approximators**

This just means that, for finite bounds on the approximation error, a finite mixture of basis functions meets the requirements

Fourier and wavelet models, for example, form a **dictionary of basis functions**

Many, but not all, such models impose **orthogonality in the basis** to ensure better conditioning

For example, the **Fourier basis** takes $\phi_n = \exp(i n \omega t)$ this means that the coefficients $w$ in $x[t] = \sum_i w_i \phi_i(X)$ are **complex**
The time-frequency plane

We are familiar with the Fourier basis, either seen in complex notation, or seen using a sine, cosine basis set.

Although these change in frequency the support of the basis, how 'wide' it is in the time-domain, is always the same.

It's worth at this point considering what this imposes on the way we decompose information in time and frequency.
The time-frequency plane

The time-frequency **Heisenberg** \[\Delta t \Delta \omega = k\]

The **short-time Fourier transform** works on fixed \(\Delta t\)

Better time resolution  

better frequency resolution
We can divide the t-f as we want

What if we propose $\Delta \omega$ as a function of $\omega$?

This is one of the reasons for the creation of the wavelet transform and multiresolution analysis.

We get good time resolution for high-frequency events and good frequency resolution for low-frequency events, the combination best suited for many real signals.
Wavelets

This means that the width of the wavelet in the time domain is related to the frequency it represents.

This also requires wavelets to be local – as opposed to global, such as the sines and cosines of the Fourier basis.
Wavelets

We can start by considering a simple construction – using the notion of a scale space.

The WT is defined by convolution with the wavelet function – in the same way that the FT is defined by convolving with harmonic functions.

\[ \mathcal{W}(x[t]; s) = x[t] \ast \psi_s(x) \]

The canonical building block in the WT is the mother wavelet – which is scaled to produce the wavelet basis at scale \( s \)

\[ \psi_s(x) = \frac{1}{s} \psi(x/s) \]

One constructor for the mother wavelet is as the derivative of a smoothing function:

\[ \psi(x) = \frac{dG(x)}{dx} \]
For example

We can take $G(x)$ to be the Gaussian

We *could* change the scale $s$ in any consistent way we want – but **full coverage of the time-frequency plane** is easiest achieved using

$$s_n = 2s_{n-1} \text{ hence } s_n = 2^n s_0$$
DWT decomposition
Wavelets – finishing comments

Wavelets form an **optimal basis** (under least squares) for **sparse reconstruction**

There exists a vast array of mother wavelet functions – differing in their properties, i.e. differentiability, compact support etc

They are **widely used**, forming e.g. the core of JPEG2000 and mp4 encoding
Creating a bespoke basis

• So far we have considered only fixed basis functions

• A variety of approaches, in effect, create a bespoke basis given the observed data

• This can be very effective – we now explore the simplest of these approaches, by considering the decomposition of the embedding matrix
The embedding matrix or lag matrix consists of copies of the timeseries lagged by differing amounts.

Consider a digital signal \( x[n] \) for index \( n = N : M \). Construct an embedding matrix from this sequence from a set of \( p \) lagged versions of \( x \).

\[
X = \begin{pmatrix}
  x[N - 1] & x[N] & \ldots & x[M - 1] \\
  x[N - 2] & x[N - 1] & \ldots & x[M - 2] \\
  x[N - 3] & x[N - 2] & \ldots & x[M - 3] \\
  \vdots & \vdots & \ddots & \vdots \\
  x[N - p] & x[N - p + 1] & \ldots & x[M - p]
\end{pmatrix}
\]

The methods will now differ according to the way in which we decompose this matrix \( X \) into \( X = AB \)
Principle Components

We first look at the PCA (or SVD) basis, which is just a generalised eigen decomposition

\[ X = U S V^T \]

where \( U \) is an orthogonal set of projections of \( X \) onto the vectors contained in (the orthonormal matrix) \( V \). The matrix \( S \) is diagonal and contains the singular values. You do not need worry about this method - although it is worth noting that the singular values are the positive square roots of the eigenvalues of the matrix \( XX^T \). The main point is that, given \( X \) the matrices \( U, S, V \) are uniquely determined.
Data-dependent FIR filter

So what use is all this? Taking the above equation we can re-write it such that

\[ U = WX \]

Let’s consider just the first component of \( U \), a set of samples \( u_1[n] \) say.

\[ u_1[n] = \sum_{k=1}^{p} w_1[k] x[n-k] \]

which is just a general moving average FIR filter. This means that each column of \( U \) is a different FIR filter, whose coefficients are determined from the observed data, rather than being ‘designed’.
Example
What about the 'filters'? 

Filter 'coefficients' $w$ and corresponding power spectra (linear scale)
Now what about ICA?

Note: the SVD basis produces a basis similar to the FT, and ICA to that of Wavelets
Other decompositions are available

We could consider doing $\mathbf{U} = \mathbf{A}\mathbf{B}$ using any linear decomposition

- Non-negative factorisation
- Zero-phase component analysis
- Minor component analysis
- etc....

Though the 'meaningfulness' of the basis may not be obvious and they are not widely used

Next, though we do look at a data-driven basis related to the Hilbert \textbf{Transform} of the timeseries
Empirical Mode Decomposition

The Hilbert-Huang transform
The Hilbert transform

Time series \( X(t) \)

\( P \) Cauchy principal value

Complex Hilbert representation \( Z(t) \)

\[
Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} \, dt'
\]

\[
Z(t) = X(t) + iY(t) = a(t)e^{i\theta(t)}
\]

\[
a(t) = \left[ X^2(t) + Y^2(t) \right]^{1/2}, \quad \theta(t) = \arctan \left( \frac{Y(t)}{X(t)} \right)
\]

\[
\omega = \frac{d\theta(t)}{dt}
\]

Forms the core of the **analytic signal transform**
Single harmonic system

\[ x = \sin([1:2000]' \times 2\pi \times 0.05); \]
Chirp

\[ x = \sin([1:2000]'*2*\pi.*(0.01+0.04*(1:2000)/2000)) \];

All works well – we get good representations for the **amplitude and frequency**
This has caused problems. The Hilbert / AST has made the tacit assumption that there is a single process here – so what we obtain is wrong.
EMD: The sifting process

Fit a smooth function to the **extrema** – both sides

At each iteration, we remove the 'mean' which is

\[ m = \frac{(a - b)}{2} \]
We iterate until the relative change in $m$ is below some threshold. This then provides an **intrinsic mode function**
Final decomposition: intrinsic mode functions
Hilbert of IMF

Freq = 0.05

Freq = 0.03
Histograms of the frequencies from Hilbert of the IMFs
Non sine data
We recover the mixture of sine wave and triangle (kind of)…
Sunspot data
FX mid-price
Open issues

- EMD is non-Bayesian
- Number of components
  - Stopping criteria in sifting
  - ‘quick hacks’ work ok, but...
- Fitting process?
  - Splines impose a model
- Non-causal at present
  - Some literature on ‘pseudo-causal’ models
Interesting avenues

• Go Bayesian!
• Links with other decomposition processes
• Applications to multi-d data [e.g. tracks]
• Links with wavelets & W-V [STFT] already discussed, but need tying up [Huang et al. 1998]
• Applications – nice property regarding the likelihood of zero crossings:
  • ideal for financial modelling [bond yield curves, predictive dislocation points etc]
  • Physical system signals
Core reference

The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis

By Norden E. Huang\(^1\), Zheng Shen\(^2\), Steven R. Long\(^3\), Manli C. Wu\(^4\), Hsing H. Shih\(^5\), Quanan Zheng\(^6\), Nai-Chyuan Yen\(^7\), Chi Chao Tung\(^8\) and Henry H. Liu\(^9\)

Order-statistic filters

Consider a set of digital samples from $x[n-p] : x[n]$. Order the set from smallest to largest, and let this sample set be $z[1] : z[p]$. Order statistic filters then perform linear filtering on this ordered set, rather than the original.

For example, the **median filter** considers a linear filtering of $z$ such that all coefficients are 0, save for the middle element which is 1.