Introduction to Wavelets: Overview

• wavelets are analysis tools for time series and images
• as a subject, wavelets are
  – relatively new (1983 to present)
  – a synthesis of old/new ideas
  – keyword in 50,000+ articles and books since 1989
    (an inundation of material!!!)
• broadly speaking, there have been two waves of wavelets
  – continuous wavelet transform (1983 and on)
  – discrete wavelet transform (1988 and on)
• will introduce subject via CWT & then concentrate on DWT
What is a Wavelet?

- sines & cosines are ‘big waves’

- wavelets are ‘small waves’ (left-hand is Haar wavelet $\psi^{(\text{H})}(\cdot)$)
Technical Definition of a Wavelet: 1

- real-valued function $\psi(\cdot)$ defined over real axis is a wavelet if
  1. integral of $\psi^2(\cdot)$ is unity: $\int_{-\infty}^{\infty} \psi^2(u) \, du = 1$
     (called ‘unit energy’ property, with apologies to physicists)
  2. integral of $\psi(\cdot)$ is zero: $\int_{-\infty}^{\infty} \psi(u) \, du = 0$
     (technically, need an ‘admissibility condition,’ but this is almost equivalent to integration to zero)
Technical Definition of a Wavelet: II

- \( \int_{-\infty}^{\infty} \psi^2(u) \, du = 1 \) & \( \int_{-\infty}^{\infty} \psi(u) \, du = 0 \) give a wavelet because:
  - by property 1, for every small \( \epsilon > 0 \), have
    \[
    \int_{-T}^{-\infty} \psi^2(u) \, du + \int_{T}^{\infty} \psi^2(u) \, du < \epsilon
    \]
    for some finite \( T \)
  - ‘business’ part of \( \psi(\cdot) \) is over interval \([−T, T]\)
  - width \( 2T \) of \([−T, T]\) might be huge, but will be insignificant compared to \((−\infty, \infty)\)
  - by property 2, \( \psi(\cdot) \) is balanced above/below horizontal axis

- matches intuitive notion of a ‘small’ wave
Two Non-Wavelets and Three Wavelets

- Two failures: $f(u) = \cos(u)$ & same limited to $[-3\pi/2, 3\pi/2]$:

- Haar wavelet $\psi^{(H)}(\cdot)$ and two of its friends:
What is Wavelet Analysis?

• wavelets tell us about variations in local averages
• to quantify this description, let $x(\cdot)$ be a ‘signal’
  — real-valued function of $t$ defined over real axis
  — will refer to $t$ as time (but it need not be such)
• consider ‘average value’ of $x(\cdot)$ over $[a, b]$:

$$\frac{1}{b - a} \int_{a}^{b} x(t) \, dt$$
Approaching Average Value of a Signal

- can approximate integral using Riemann sum
  - break \([a, b]\) into \(N\) subintervals of equal width \((b - a)/N\)
  - sample \(x(\cdot)\) at midpoint of each subinterval:
    \[x_j = x \left( a + \left[ j + \frac{1}{2} \right] \frac{b-a}{N} \right), \quad j = 0, 1, \ldots, N - 1\]
  - Riemann sum = sum of \(x_j\)'s \(\times\) width \((b - a)/N\)
  - yields approximation to average value of \(x(\cdot)\) over \([a, b]\):
    \[
    \frac{1}{b-a} \int_a^b x(t) \, dt \approx \frac{1}{b-a} \left( \frac{b-a}{N} \sum_{j=0}^{N-1} x_j \right) = \frac{1}{N} \sum_{j=0}^{N-1} x_j
    \]
- average value of \(x(\cdot)\) \(\approx\) sample mean of sampled values
Example of Average Value of a Signal

- let $x(\cdot)$ be step function taking on values $x_0, x_1, \ldots, x_{15}$ over 16 equal subintervals of $[a, b]$:

- here we have

$$\frac{1}{b-a} \int_a^b x(t) \, dt = \frac{1}{16} \sum_{j=0}^{15} x_j = \text{height of dashed line}$$
Average Values at Different Scales and Times

• define the following function of $\lambda$ and $t$

$$A(\lambda, t) \equiv \frac{1}{\lambda} \int_{t - \frac{\lambda}{2}}^{t + \frac{\lambda}{2}} x(u) \, du$$

  – $\lambda$ is width of interval – referred to as ‘scale’
  – $t$ is midpoint of interval

• $A(\lambda, t)$ is average value of $x(\cdot)$ over scale $\lambda$ centered at $t$

• average values of signals have wide-spread interest
  – one second average temperatures over forest
  – ten minute rainfall rate during severe storm
  – yearly average temperatures over central England
Defining a Wavelet Coefficient $W$

- multiply Haar wavelet & time series $x(\cdot)$ together:

$$\int_{-\infty}^{\infty} \psi^{(H)}(t)x(t) \, dt = W(1, 0)$$

- integrate resulting function to get ‘wavelet coefficient’ $W(1, 0)$:

$$W(1, 0) \propto \frac{1}{1} \int_{0}^{1} x(t) \, dt - \frac{1}{1} \int_{-1}^{0} x(t) \, dt = A(1, \frac{1}{2}) - A(1, -\frac{1}{2})$$
Defining Wavelet Coefficients for Other Scales

- $W(1, 0)$ proportional to difference between averages of $x(\cdot)$ over $[-1, 0]$ & $[0, 1]$, i.e., two unit scale averages before/after $t = 0$
  - ‘1’ in $W(1, 0)$ denotes scale 1 (width of each interval)
  - ‘0’ in $W(1, 0)$ denotes time 0 (center of combined intervals)

- stretch or shrink wavelet to define $W(\tau, 0)$ for other scales $\tau$:

\[
\begin{align*}
WMTSA: 9–10 & \quad I–11
\end{align*}
\]
Defining Wavelet Coefficients for Other Locations

• relocate to define $W(\tau, t)$ for other times $t$:

$\text{yields } W(1, 1)$

$\text{yields } W(2, -\frac{1}{2})$
Haar Continuous Wavelet Transform (CWT)

• for all $\tau > 0$ and all $-\infty < t < \infty$, can write

$$W(\tau, t) = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u)\psi^{(H)}\left(\frac{u-t}{\tau}\right) du$$

  - $\frac{u-t}{\tau}$ does the stretching/shrinking and relocating
  - $\frac{1}{\sqrt{\tau}}$ needed so $\psi^{(H)}_{\tau,t}(u) \equiv \frac{1}{\sqrt{\tau}}\psi^{(H)}\left(\frac{u-t}{\tau}\right)$ has unit energy
  - since it also integrates to zero, $\psi^{(H)}_{\tau,t}(\cdot)$ is a wavelet

• $W(\tau, t)$ over all $\tau > 0$ and all $t$ is Haar CWT for $x(\cdot)$

• analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of averages
Other Continuous Wavelet Transforms: I

- can do the same for wavelets other than the Haar
- start with basic wavelet $\psi(\cdot)$
- use $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}} \psi \left( \frac{u-t}{\tau} \right)$ to stretch/shrink & relocate
- define CWT via
  $$W(\tau, t) = \int_{-\infty}^{\infty} x(u) \psi_{\tau,t}(u) \, du = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u) \psi \left( \frac{u-t}{\tau} \right) \, du$$
- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of weighted averages
Other Continuous Wavelet Transforms: II

- consider two friends of Haar wavelet

\[
\psi^{(H)}(u) \quad \psi^{(fdG)}(u) \quad \psi^{(Mh)}(u)
\]

- \(\psi^{(fdG)}(\cdot)\) proportional to 1st derivative of Gaussian PDF
- ‘Mexican hat’ wavelet \(\psi^{(Mh)}(\cdot)\) proportional to 2nd derivative
- \(\psi^{(fdG)}(\cdot)\) looks at difference of adjacent weighted averages
- \(\psi^{(Mh)}(\cdot)\) looks at difference between weighted average and sum of weighted averages occurring before & after
First Scary-Looking Equation

- CWT equivalent to \( x(\cdot) \) because we can write

\[
x(t) = \int_{0}^{\infty} \left[ \frac{1}{C\tau^2} \int_{-\infty}^{\infty} W(\tau, u) \frac{1}{\sqrt{\tau}} \psi \left( \frac{t - u}{\tau} \right) \, du \right] \, d\tau,
\]

where \( C \) is a constant depending on specific wavelet \( \psi(\cdot) \)

- can synthesize (put back together) \( x(\cdot) \) given its CWT; i.e., nothing is lost in reexpressing signal \( x(\cdot) \) via its CWT

- regard stuff in brackets as defining ‘scale \( \tau \)’ signal at time \( t \)

- says we can reexpress \( x(\cdot) \) as integral (sum) of new signals, each associated with a particular scale

- similar additive decompositions will be one central theme
energy in $x(\cdot)$ is reexpressed in CWT because

$$\text{energy} = \int_{-\infty}^{\infty} x^2(t) \, dt = \int_{0}^{\infty} \left[ \frac{1}{C\tau^2} \int_{-\infty}^{\infty} W^2(\tau, t) \, dt \right] \, d\tau$$

• can regard $x^2(t)$ versus $t$ as breaking up the energy across time (i.e., an ‘energy density’ function)

• regard stuff in brackets as breaking up the energy across scales

• says we can reexpress energy as integral (sum) of components, each associated with a particular scale

• function defined by $W^2(\tau, t)/C\tau^2$ is an energy density across both time and scale

• similar energy decompositions will be a second central theme
Example: Atomic Clock Data

- example: average daily frequency variations in clock 571

\[
X_t
\]

- \( t \) is measured in days (one measurement per day)
- plot shows \( X_t \) versus integer \( t \)
- \( X_t = 0 \) would mean that clock 571 could keep time perfectly
- \( X_t < 0 \) implies that clock is losing time systematically
- can easily adjust clock if \( X_t \) were constant
- inherent quality of clock related to changes in averages of \( X_t \)
Mexican Hat CWT of Clock Data: I
Mexican Hat CWT of Clock Data: II

\[
\begin{align*}
\tau & = \text{Parameter of interest} \\
X_t & = \text{Time series data}
\end{align*}
\]
Mexican Hat CWT of Clock Data: III

\[
\begin{align*}
\tau & \quad \text{(scale)} \\
X_t & \quad \text{(time)}
\end{align*}
\]
Mexican Hat CWT of Clock Data: IV
Beyond the CWT: the DWT

- can often get by with subsamples of $W(\tau, t)$
- leads to notion of discrete wavelet transform (DWT) (can regard as discretized ‘slices’ through CWT)
Rationale for the DWT

• DWT has appeal in its own right
  – most time series are sampled as discrete values (can be tricky to implement CWT)
  – can formulate as orthonormal transform (makes meaningful statistical analysis possible)
  – tends to decorrelate certain time series
  – standardization to dyadic scales often adequate
  – generalizes to notion of wavelet packets
  – can be faster than the fast Fourier transform
• will concentrate primarily on DWT for remainder of course
Addendum on First Scary-Looking Equation: I

• can synthesize signal $x(\cdot)$ from its CWT $W(\cdot, \cdot)$:

$$x(t) = \int_0^\infty \left[ \frac{1}{C \tau^2} \int_{-\infty}^{\infty} W(\tau, u) \frac{1}{\sqrt{\tau}} \psi \left( \frac{t-u}{\tau} \right) \, du \right] \, d\tau,$$  (*)

where $C$ is a constant depending on specific wavelet $\psi(\cdot)$

• Q: what is the constant $C$ all about?

• as mentioned on overhead I–3, for a function $\psi(\cdot)$ to be a wavelet, it must satisfy a so-called ‘admissibility condition’

• to state admissibility condition, let $\Psi(\cdot)$ denote Fourier transform of $\psi(\cdot)$ (assumed to be a square-integrable function):

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(u) e^{-i2\pi fu} \, du$$
Addendum on First Scary-Looking Equation: II

• admissibility condition says that
  \[ C \equiv \int_0^\infty \frac{|\Psi(f)|^2}{f} df \text{ must be such that } 0 < C < \infty \]
  (note: above implies that \( \psi(\cdot) \) must integrate to zero)

• \( C \) above is same \( C \) appearing in (*)

• as to why \( C \) appears, need to work through proof of (*), which
  is not trivial
  
  – see Mallat, 1998, §4.3 for a clear proof
  – proof in the wavelet literature due to Grossman and Morlet, 1984, who discuss why admissibility condition is needed
  – Grossman and Morlet’s result actually appeared earlier in 1964 paper by Calderón