Wavelet Methods for Time Series Analysis

Part XI: Continuous Wavelet Transforms

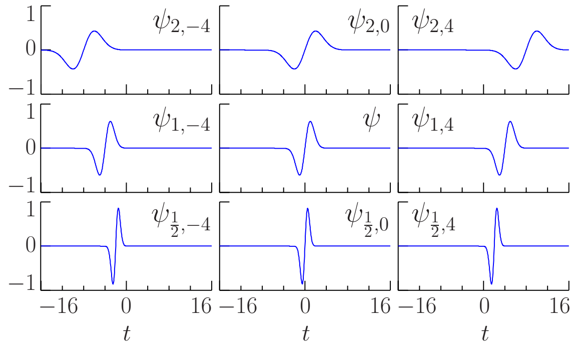
- will look at connection between discrete wavelet transforms (DWTs) and underlying continuous wavelet transforms (CWTs)
- will discuss about one form of analysis (wavelet transform modulus maxima) that benefits from use of CWT rather than DWT
- close with some brief comments about CWTs that do not easily fall into 'differences of weighted averages' framework

Definition of Continuous Wavelet Transform: I

- start with basic wavelet $\psi(\cdot)$ satisfying conditions

$$\int_{-\infty}^{\infty} \psi^2(u)\,du = 1 \text{ and } \int_{-\infty}^{\infty} \psi(u)\,du = 0$$

- use $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}}\psi\left(\frac{u-t}{\tau}\right)$ to stretch/shrink and relocate $\psi(\cdot)$



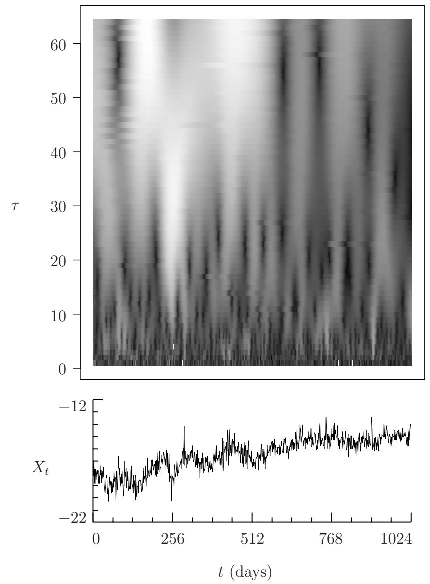
Definition of Continuous Wavelet Transform: II

- define CWT of 'signal' $x(\cdot)$ as

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

where $\tau > 0$ and $t \in \mathbb{R} \equiv (-\infty, \infty)$

Example: Mexican Hat CWT of Clock Data



Connection Between DWT and CWT

- motivated DWT by claiming it could be regarded as subsamples along slices of CWT at scales $\tau = \tau_j = 2^{j-1}$ (MODWT would be slices with nonsparse subsampling)
- Q: what precisely is the wavelet $\psi(\cdot)$ corresponding to, say, an LA(8) DWT?
- once $\psi(\cdot)$ is known, can approximate $W(\tau, t)$ for arbitrary τ using numerical integration

Scaling Functions: I

- start with a scaling function $\phi(\cdot)$ satisfying 3 conditions:
  1. integral of $\phi^2(\cdot)$ is unity: $\int_{-\infty}^{\infty} \phi^2(u)\,du = 1$
  2. integral of $\phi(\cdot)$ is some positive value: $\int_{-\infty}^{\infty} \phi(u)\,du > 0$
  3. orthogonality to its integer shifts:

$$\int_{-\infty}^{\infty} \phi(u)\phi(u-k)\,du = 0 \text{ when } k \text{ is a nonzero integer}$$

Scaling Functions: II

- denote stretched/shrunk and relocated $\phi(\cdot)$ by

$$\phi_{j,k}(u) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{u-k}{2^j}\right)$$

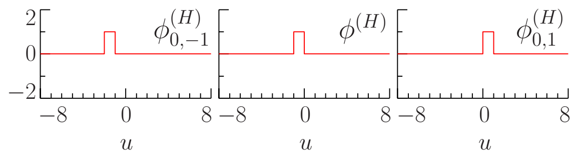
  – differs from $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}}\psi\left(\frac{u-t}{\tau}\right)$, but need only deal with
    - ∗ scales τ that are powers of 2 and
    - ∗ shifts t that are integers $k \in \mathbb{Z}$ (the set of all integers),

    so we have simplified notation (avoids subscript with a superscript)
- note that $\phi_{0,0}(u) = \phi(u)$
- note also that $\phi(\cdot)$ is associated with unit scale

Scaling Functions: III

- simple example is the Haar scaling function:

$$\phi^{(H)}(u) \equiv \begin{cases} 1, & -1 < u \leq 0; \\ 0, & \text{otherwise} \end{cases}$$



- note that $\phi^{(H)}(\cdot)$ satisfies conditions 1, 2 and 3

Approximation Spaces: I

- let V_0 be the space of functions with finite energy that can be expressed as linear combinations of $\{\phi_{0,k}(\cdot) : k \in \mathbb{Z}\}$
- V_0 is called an approximation space of scale unity
- let $x(\cdot)$ be a signal with finite energy:

$$\|x\|^2 \equiv \int_{-\infty}^{\infty} x^2(t)\, dt < \infty;$$

as shorthand for the above, we say that $x(\cdot)$ belongs to $L^2(\mathbb{R})$

Approximation Spaces: II

- can analyze $x(\cdot)$ using $\{\phi_{0,k}(\cdot)\}$:

$$v_{0,k} \equiv \langle x(\cdot), \phi_{0,k}(\cdot)\rangle \equiv \int_{-\infty}^{\infty} x(t)\phi_{0,k}(t)\, dt,$$

where $\{v_{0,k}\}$ are called the scaling coefficients for unit scale

- if $x(\cdot) \in V_0$, can synthesize $x(\cdot)$ from a linear combination of functions in $\{\phi_{0,k}(\cdot)\}$ weighted by $\{v_{0,k}\}$:

$$x(t) = \sum_{k=-\infty}^{\infty} v_{0,k}\phi_{0,k}(t) = \sum_{k=-\infty}^{\infty} v_{0,k}\phi(t-k)$$

- if $x(\cdot) \notin V_0$, right-hand side can be regarded as an approximation to $x(\cdot)$ (also called the 'projection' of $x(\cdot)$ onto V_0)

Approximation Spaces: III

- let V_j be the space of functions with finite energy that can be expressed as linear combinations of $\{\phi_{j,k}(\cdot) : k \in \mathbb{Z}\}$;
- V_j is called an approximation space of scale $\lambda_j = 2^j$

Approximation Spaces: IV

- can also analyze $x(\cdot)$ using $\{\phi_{j,k}(\cdot)\}$:

$$v_{j,k} \equiv \langle x(\cdot), \phi_{j,k}(\cdot)\rangle = \int_{-\infty}^{\infty} x(t)\phi_{j,k}(t)\, dt,$$

where $\{v_{j,k}\}$ are called the scaling coefficients for scale λ_j

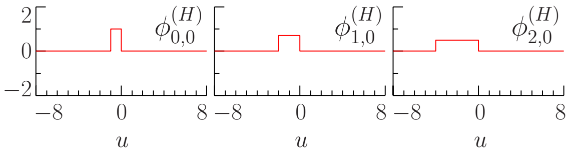
- if $x(\cdot) \in V_j$, can synthesize $x(\cdot)$ from $\{\phi_{j,k}(\cdot)\}$ and $\{v_{j,k}\}$:

$$x(t) = \sum_{k=-\infty}^{\infty} v_{j,k}\phi_{j,k}(t)$$

- if $x(\cdot) \notin V_j$, right-hand side is the approximation to $x(\cdot)$ obtained by projecting it onto V_j

Approximation Spaces: V

- if $j > 0$, $\phi_{j,k}(\cdot)$ is a stretched out version of $\phi_{0,k}(\cdot)$:



- hence

$$\sum_{k=-\infty}^{\infty} v_{j,k}\phi_{j,k}(t) \text{ is 'coarser' than } \sum_{k=-\infty}^{\infty} v_{0,k}\phi_{0,k}(t)$$

- can formalize this notion of coarser/finer approximations by stipulating relationships between the spaces V_j's

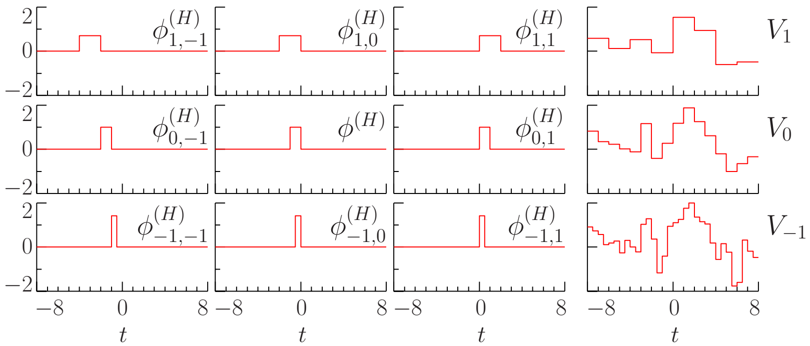
Multiresolution Analysis: I

- formal definition of multiresolution analysis (MRA): sequence of closed subspaces $V_j \subset L^2(\mathbb{R})$, $j \in \mathbb{Z}$, such that:

1. $\cdots \subset V_3 \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots$
   $\longleftarrow$ coarser finer $\longrightarrow$
2. $x(\cdot) \in V_0$ if and only if $x_{0,k}(\cdot) \in V_0$ also
3. $x(\cdot) \in V_0$ if and only if $x_{j,0}(\cdot) \in V_j$ also
4. following two technical conditions hold:
$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \text{ and } \bigcap_{j\in\mathbb{Z}} V_j = \{0\}$$
   (here '0' refers to function that is 0 for all t)
5. there exists a scaling function $\phi(\cdot) \in V_0$ such that $\{\phi_{0,k}(\cdot) : k \in \mathbb{Z}\}$ forms an orthonormal basis for V_0

Multiresolution Analysis: II

- let's illustrate these ideas using the Haar scaling function:



Approximation of Finite Energy Signals: I

- suppose $x(\cdot) \in V_0$; i.e., $x(\cdot)$ is signal of scale $\lambda_0 = 1$
- can write

$$x(t) = \sum_{k=-\infty}^{\infty} v_{0,k}\phi_{0,k}(t) \text{ with } v_{0,k} \equiv \langle x(\cdot), \phi_{0,k}(\cdot)\rangle$$

- can approximate $x(\cdot)$ by projecting it onto coarser $V_1 \subset V_0$
- this approximation is given by

$$s_1(t) \equiv \sum_{k=-\infty}^{\infty} \langle x(\cdot), \phi_{1,k}(\cdot)\rangle \phi_{1,k}(t) = \sum_{k=-\infty}^{\infty} v_{1,k}\phi_{1,k}(t)$$

(its DWT analog is the first level smooth $\mathcal{S}_1$)

Approximation of Finite Energy Signals: II

- an easy exercise shows that $\{v_{0,k}\}$ and $\{v_{1,k}\}$ are related by
$$v_{1,k} = \sum_{l=-\infty}^{\infty} g_l v_{0,2k-l},$$
where
$$g_l \equiv \int_{-\infty}^{\infty} \phi(u+l)\frac{\phi(\frac{u}{2})}{\sqrt{2}}\,du;$$
i.e., $g_l = \langle \phi_{0,-l}(\cdot), \phi_{1,0}(\cdot)\rangle$
- thus: filter $\{v_{0,k}\}$ and downsample to get $\{v_{1,k}\}$!!!

Approximation of Finite Energy Signals: III

- Haar scaling function example:
$$g_l = \int_{-\infty}^{\infty} \phi(u+l)\frac{\phi(\frac{u}{2})}{\sqrt{2}}\,du = \frac{1}{\sqrt{2}}\int_{-2}^{0} \phi(u+l)\,du$$
since $\phi(\frac{u}{2}) \neq 0$ for $-1 < u/2 \le 0$, i.e., $-2 < u \le 0$
- since $\phi(u+l) \neq 0$ for $-1 < u+l \le 0$, i.e., $-1-l < u \le -l$, obtain
$$g_l = \begin{cases} 1/\sqrt{2}, & l = 0, 1; \\ 0, & \text{otherwise}, \end{cases}$$
in agreement with the Haar scaling filter introduced earlier

Approximation of Finite Energy Signals: IV

- more generally, can approximate $x(\cdot)$ by projecting it onto V_j:
$$s_j(t) \equiv \sum_{k=-\infty}^{\infty} \langle x(\cdot), \phi_{j,k}(\cdot)\rangle \phi_{j,k}(t) = \sum_{k=-\infty}^{\infty} v_{j,k}\phi_{j,k}(t)$$
where
$$v_{j,k} = \sum_{l=-\infty}^{\infty} g_l v_{j-1,2k-l}$$
(the DWT analog of $s_j(\cdot)$ is the jth level smooth $\mathcal{S}_j$)
- compare the above with a very similar equation relating the DWT scaling coefficients for levels $j-1$ and j:
$$V_{j,t} = \sum_{l=0}^{L-1} g_l V_{j-1,2t+1-l \bmod N/2^{j-1}},$$
where we defined $V_{0,t}$ to be equal X_t

Approximation of Finite Energy Signals: V

- schemes quite similar if we equate $v_{0,k}$ with X_k
- note that
$$v_{0,k} = \int_{-\infty}^{\infty} x(t)\phi_{0,k}(t)\,dt,$$
for which we would need to know $x(t)$ for all t
- usually we just have $X_k = x(k)$, and we get the DWT by
  - replacing projections $v_{0,k}$ by the samples X_k
  - making a periodic extension
- other initialization schemes possible (e.g., estimate $v_{0,k}$ using samples X_k with quadrature formulae)

Relating Scaling Functions to Filters: I

- since $V_0 \subset V_{-1}$ and since $\phi(\cdot) \in V_0$, it follows that $\phi(\cdot) \in V_{-1}$, so we can write

$$\phi(t) = \sum_{l=-\infty}^{\infty} \langle \phi(\cdot), \phi_{-1,l}(\cdot) \rangle \phi_{-1,l}(t) = \sum_{l=-\infty}^{\infty} g_l \phi_{-1,-l}(t)$$

- yields a '2 scale' relationship

$$\phi(t) = \sqrt{2} \sum_{l=-\infty}^{\infty} g_l \phi(2t + l)$$

- let $\Phi(\cdot)$ be the Fourier transform of $\phi(\cdot)$:

$$\Phi(f) \equiv \int_{-\infty}^{\infty} \phi(t) e^{-i2\pi f t}\, dt$$

Relating Scaling Functions to Filters: II

- can use the 2 scale relationship to solve for $\Phi(\cdot)$ (and hence $\phi(\cdot)$) in terms of the discrete Fourier transform $G(\cdot)$ of $\{g_l\}$:

$$\Phi(-f) = \prod_{m=1}^{\infty} \frac{G(\frac{f}{2^m})}{\sqrt{2}};$$

i.e., knowing $\{g_l\}$ is equivalent to knowing the scaling function

- taking the FT of the above and recalling that the DFT of the jth level equivalent scaling filter $\{g_{j,l}\}$ takes the form

$$\prod_{l=0}^{j-1} G(2^l f)$$

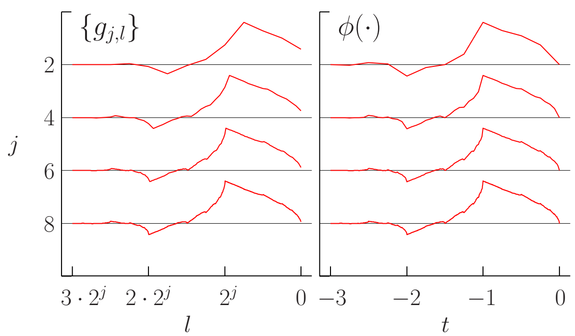
leads us to the connection

$$\phi(-\tfrac{l}{2^j}) \approx 2^{j/2} g_{j,l},$$

with the approximation getting better and better as j increases

Relating Scaling Functions to Filters: III

- specialize to Daubechies scaling filters $\{g_l\}$
  - $\{g_l\}$ has finite width L; i.e., $g_l = 0$ for $l < 0$ or $l \geq L$
  - implies $\phi(t) = 0$ for $t \notin (-L + 1, 0]$
- comparison of $\{g_{j,l}\}$ and $\phi(\frac{l}{2^j})$ for D(4)



  - note: can actually compute $\phi(\frac{l}{2^j})$ *exactly*!

Wavelet Functions and Detail Spaces: I

- let V_j, $j \in \mathbb{Z}$, be an MRA (in particular, $V_0 \subset V_{-1}$)
- let $W_0 \subset V_{-1}$ be the orthogonal complement of V_0 in V_{-1}; i.e., if $\varphi(\cdot) \in W_0$, then $\varphi(\cdot) \in V_{-1}$ but

$$\langle \varphi(\cdot), x(\cdot) \rangle = 0$$

for any $x(\cdot) \in V_0$

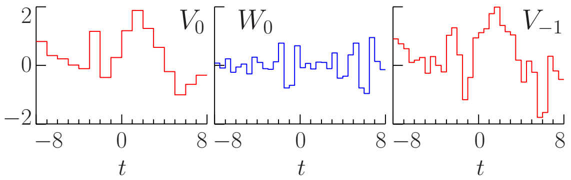
- by definition can write

$$V_{-1} = V_0 \oplus W_0;$$

i.e., V_{-1} is direct sum of V_0 and W_0 (meaning that any element in V_{-1} is the sum of 2 orthogonal functions, one from V_0, and the other from W_0)

Wavelet Functions and Detail Spaces: II

- examples of functions in Haar-based V_0, W_0 and V_{-1}



- in general $V_j \subset V_{j-1}$ and $W_j \subset V_{j-1}$ is the orthogonal complement in V_{j-1} of V_j:

$$V_j = V_{j+1} \oplus W_{j+1}$$

- W_j is called the detail space for scale $\tau_j = 2^{j-1}$, (the functions in W_j are analogous to the jth level DWT details $\mathcal{D}_j$)

Wavelet Functions and Detail Spaces: III

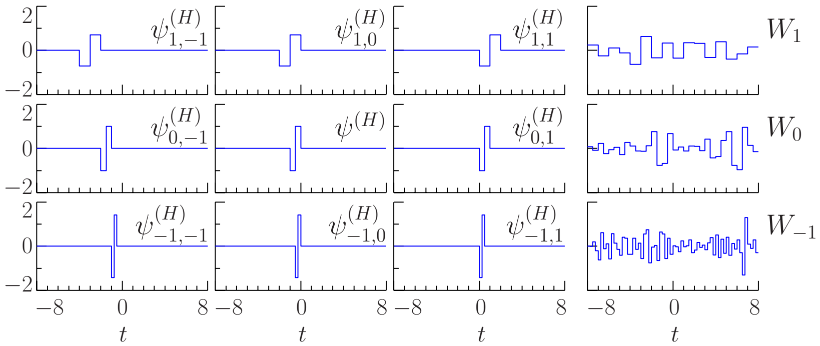
- define wavelet function

$$\psi(t) \equiv \sqrt{2} \sum_{l=-\infty}^{\infty} h_l \phi(2t + l), \text{ where } h_l \equiv (-1)^l g_{1-l-L}$$

- can argue that $\{\psi_{0,m}(\cdot) : m \in \mathbb{Z}\}$ forms orthonormal basis for detail space W_0, and, in general, $\{\psi_{j,m}(\cdot)\}$ forms orthonormal basis for W_j

Wavelet Functions and Detail Spaces: III

- examples of basis functions and of functions in W_1, W_0 and W_{-1} for Haar case



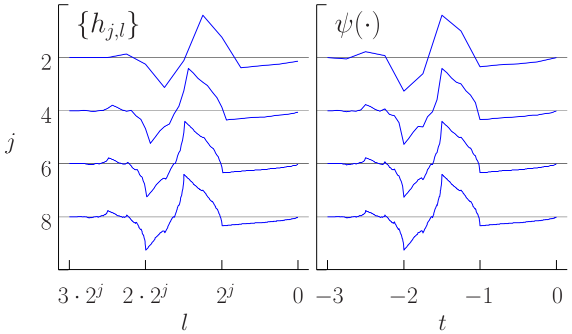
Wavelet Functions and Detail Spaces: IV

- projections of $x(\cdot)$ onto space W_j give us the difference between the successive approximations $s_{j-1}(\cdot)$ and $s_j(\cdot)$
- construction emphasizes that wavelets are the connection between adjacent approximations in an MRA
- consistent with DWT where $\mathcal{D}_j = \mathcal{S}_{j-1} - \mathcal{S}_j$ follows from

$$\mathbf{X} = \sum_{k=1}^{j} \mathcal{D}_k + \mathcal{S}_j \text{ and } \mathbf{X} = \sum_{k=1}^{j-1} \mathcal{D}_k + \mathcal{S}_{j-1}$$

Relating Wavelet Functions to Filters

- can argue that the wavelet function $\psi(\cdot)$ is related to the jth level equivalent wavelet filter $\{h_{j,l}\}$ via $\psi(-\frac{l}{2^j}) \approx 2^{j/2} h_{j,l}$, with the approximation improving as j increases
- comparison of $\{h_{j,l}\}$ and $\psi(\frac{l}{2^j})$ for D(4)



Evaluating CWT at an Arbitrary Scale τ: I

- could in principle compute a Daubechies CWT at any desired scale τ (i.e., not just $\tau_j = 2^j$)
- to do so, start by computing the jth level equivalent filter $\{h_{j,l}\}$ for some large j
- use the relationship

$$\psi(\tfrac{l}{2^j}) \approx 2^{j/2} h_{j,-l}$$

and an interpolation scheme to approximate $\psi(\cdot)$ over the interval $(-L+1, 0]$ (it is zero outside of this interval)

Evaluating CWT at an Arbitrary Scale τ: II

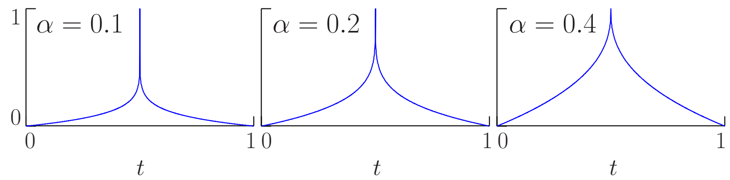
- given $x(\cdot)$, use the approximation to $\psi(\cdot)$ to evaluate

$$W(\tau, t) = \int_{-\infty}^{\infty} x(u)\psi_{\tau,t}(u)\,du = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u)\psi\left(\frac{u-t}{\tau}\right) du$$
$$= \sqrt{\tau} \int_{-L+1}^{0} x(\tau u + t)\psi(u)\,du$$

using a numerical integration scheme
- note that Haar wavelet is easy to adjust to an arbitrary scale since it can be expressed simply in closed form

Analysis of Singularites Using the CWT: I

- have argued that the DWT is usually an adequate summary of the information in the CWT, but there are some analysis techniques for which use of the full CWT is helpful
- consider a signal with a cusp at $t = t_0$: $x(t) = a + b|t - t_0|^{\alpha}$,



- degree of 'cuspiness' might characterize, e.g., ECG data

Analysis of Singularites Using the CWT: II

- since $\psi(\cdot)$ must integrate to zero, we have (letting $v = (u - t_0)/\tau$)

$$W(\tau, t_0) = \frac{b}{\sqrt{\tau}} \int_{-\infty}^{\infty} |u - t_0|^{\alpha} \psi\left(\frac{u - t_0}{\tau}\right) du$$
$$= \frac{b}{\sqrt{\tau}} \int_{-\infty}^{\infty} |v\tau|^{\alpha} \psi(v)\, dv = C\tau^{\alpha - \frac{1}{2}},$$

  where $C \neq 0$ for a suitably chosen wavelet (e.g., Mexican hat wavelet or Daubechies wavelet associated with $L \geq 4$)
- plot of $\log |W(\tau, t_0)|$ versus $\log \tau$ should be linear with a slope of $\alpha - \frac{1}{2}$, i.e., related to degree of cusp

Analysis of Singularites Using the CWT: III

- for fixed τ, $|W(\tau, t)|$ has its maximum value at $t = t_0$, so can identify location and determine nature of cusp by searching for the wavelet transform modulus maxima (WTMM)
- in applications, presence of noise and other components can perturb maxima away from t_0, so need to track through CWT

A Wavelet Tour of Signal Processing
Stéphane Mallat, Academic Press 1999 (2nd edition)

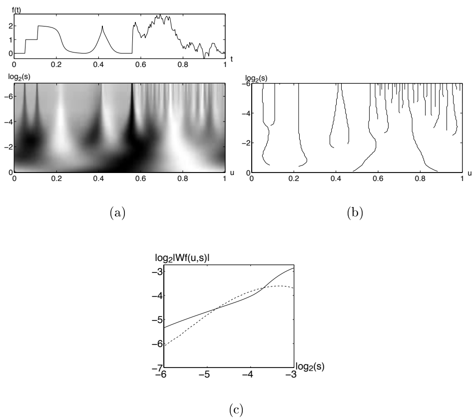


Figure 6.5: (a): Wavelet transform $Wf(u,s)$. The horizontal and vertical axes give respectively u and $\log_2 s$. (b): Modulus maxima of $Wf(u,s)$. (c): The full line gives the decay of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along the maxima line that converges to the abscissa $t = 0.05$. The dashed line gives $\log_2 |Wf(u,s)|$ along the left maxima line that converges to $t = 0.42$.

Morlet Wavelet: I

- in one of the first article on wavelets Goupillaud, Grossmann and Morlet (1984) introduced

$$\psi(u) = Ce^{-i\omega_0 u}\left(e^{-u^2/2} - \sqrt{2}e^{-\omega_0^2/4}e^{-u^2}\right),$$

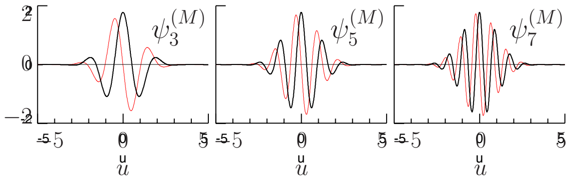
  where C and ω_0 are positive constants
- $\psi(\cdot)$ is complex-valued, but, like real-valued wavelets, $\psi(\cdot)$ integrates to zero and, if C is chosen properly, satisfies a unit energy condition (now taken to be $\int |\psi(u)|^2\, du = 1$)
- as ω_0 increases, the second term in the parentheses comes negligible, yielding an approximation known as the Morlet wavelet:

$$\psi(u) \approx \psi_{\omega_0}^{(M)}(u) \equiv \pi^{-1/4} e^{-i\omega_0 u} e^{-u^2/2}$$

  (strictly speaking, $\psi_{\omega_0}^{(M)}(\cdot)$ is *not* a wavelet because it doesn't integrate to zero exactly, but it does so approximationly)

Morlet Wavelet: II

- here are three Morlet wavelets (real/imaginary components given by thick/thin curves)



- does not yield a CWT that is easily interpreted as changes in weighted averages, but rather is a localized Fourier analysis (with ω_0 controlling the number of local cycles)