Multitaper Spectral Estimation

- tapering useful for $S(\cdot)$ with large dynamic range . . .
  . . . but increases variance of $\hat{S}^{(lw)}(f)$ by $C_h > 1$
- alternatives to $\hat{S}^{(lw)}(\cdot)$ (i.e., smoothing $\hat{S}^{(d)}(\cdot)$):
  - prewhitening (cf. Chapter 9)
  - WOSA
  - multitapering (Thomson, 1982)
- why multitapering?
  - works automatically on high dynamic range sdf’s
  - natural definition of resolution
  - can tradeoff bias/variance/resolution easily
  - for some processes, can argue
    * superior to prewhitening (Thomson, 1990a)
    * superior to WOSA (Bronez, 1992)
  - produces ‘$\hat{S}^{(d)}(\cdot)$’ with $2 < \nu \leq 2K$ dof
    * $K =$ number of tapers (2 to 10 typically)
    * Figure 257: shrinks 95% c.i.’s considerably
  - can get internal estimate of variance
  - can handle mixed spectra (i.e., line components)
  - extends naturally to irregularly sampled processes
Outline

• present basic ideas behind multitapering
• discuss two family of multitapers
• argue multitapering recovers ‘lost information’
• consider multitapering of white noise
• relate multitapering to quadratic estimators
• conclude with example (ocean wave data)
Basics of Multitapering: I

- average of $K$ direct spectral estimators:
  \[
  \hat{S}^{(mt)}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f)
  \]
  is basic multitaper estimator, where
  \[
  \hat{S}_k^{(mt)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_{t,k}X_te^{-i2\pi ft\Delta t} \right|^2
  \]
  is called $k$th eigenspectrum; uses $k$th taper $\{h_{t,k}\}$ (note resemblance to WOSA estimator)

- each taper normalized such that $\sum_t h_{t,k}^2 = 1$

- spectral window for $k$th eigenspectrum:
  \[
  \mathcal{H}_k(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_{t,k}e^{-i2\pi ft\Delta t} \right|^2
  \]

- each eigenspectrum is example of $\hat{S}^{(d)}(\cdot)$, so
  \[
  E\{\hat{S}_k^{(mt)}(f)\} = \int_{-f(N)}^{f(N)} \mathcal{H}_k(f - f')S(f') \, df
  \]
  thereby have
  \[
  E\{\hat{S}^{(mt)}(f)\} = \int_{-f(N)}^{f(N)} \mathcal{H}(f - f')S(f') \, df', \quad \mathcal{H}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \mathcal{H}_k(f)
  \]

- leakage for $\hat{S}^{(mt)}(\cdot)$ ok if $\mathcal{H}_k(\cdot)$’s have small sidelobes
Basics of Multitapering: II

• assume \( S(\cdot) \) locally constant about \( f \)

• for \( j \neq k \), can argue \( \text{cov} \{ \hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f) \} \approx 0 \)
  if tapers are orthogonal; i.e.,
  \[
  \sum_{t=1}^{N} h_{t,j} h_{t,k} = 0
  \]

• \( \text{var} \{ \hat{S}_k^{(mt)}(f) \} \approx S^2(f) \implies \text{var} \{ \hat{S}^{(mt)}(f) \} \approx S^2(f)/K \)

• two sets of orthonormal tapers in common use
  – dpss tapers (Thomson, 1982)
  – sine tapers (Riedel and Sidorenko, 1995)
DPSS Multitapers: I

• tapers minimize spectral window sidelobes

• for fixed resolution bandwidth $2W$, measure sidelobes via

$$\beta_k^2(W) \equiv \frac{\int_{-W}^{W} \mathcal{H}_k(f) \, df}{\int_{-f(N)}^{f(N)} \mathcal{H}_k(f) \, df}$$

• for given $W$ and $N$
  - $\{h_{t,0}\}$ maximizes $\beta_0^2(W)$
  - $\{h_{t,k}\}$ maximizes $\beta_k^2(W)$ amongst sequences orthogonal to $\{h_{t,0}\}, \ldots, \{h_{t,k-1}\}$

• computation of $\{h_{t,k}\}$’s requires solution of

$$Ah_{k} = \beta_k^2(W)h_{k} \quad \text{with} \quad h_{k}^T = [h_{1,k}, \cdots, h_{N,k}];$$

$$A_{t',t} = \frac{\sin(2\pi W(t' - t))}{\pi(t' - t)}$$

is $(t',t)$th element of $N \times N$ matrix $A$

  - Section 8.1: inverse iteration (stable, but slow)
  - Section 8.2: numerical integration (Thomson, 1982)
  - Section 8.3: tridiagonal formulation (fast)

• note: simple approx. to $\{h_{t,0}\}$ in Equation (386)
DPSS Multitapers: II

- number of $\{h_{t,k}\}$'s with good leakage protection is $2NW \Delta t - 1$ or less

- strategy & considerations for picking $K$
  - set resolution bandwidth $2W$
    * $\Delta f \equiv 1/N \Delta t =$ spacing between $f_k$'s (i.e., Fourier frequencies)
    * usually set $W = J \Delta f \iff NW = J/\Delta t$
      for $J = 2, 3, 4, \ldots$
  - set $K < 2NW \Delta t = 2J$, noting that
    * leakage gets worse as $K$ increases
    * variance decreases as $K$ increases
  - increasing $W$ implies
    * resolution decreases
    * more leakage free tapers (i.e., can increase $K$)
  - Figures 336–41 use $NW = 4$ (i.e, $2NW = 8$)
    * maximum $K$ should be is 7
    * eighth taper $\{h_{t,7}\}$ also depicted
Sine Multitapers: I

- tapers minimize ‘spectral window bias’
  
  - recall notion of smoothing window bias:
    \[
    b_W \approx \frac{S''(f)}{2} \int_{-f(N)}^{f(N)} \phi^2 W_m(\phi) d\phi = \frac{S''(f)}{24} \beta_W^2
    \]
    (used to derive Papoulis lag window)
  
  - same approach yields spectral window bias:
    \[
    b_{\mathcal{H}_k} \approx \frac{S''(f)}{2} \int_{-f(N)}^{f(N)} \phi^2 \mathcal{H}_k(\phi) d\phi \equiv \frac{S''(f)}{24} \beta_{\mathcal{H}_k}^2
    \]

- for given \(N\)
  
  - \(\{h_{t,0}\}\) minimizes \(\beta_{\mathcal{H}_0}^2\)
  
  - \(\{h_{t,k}\}\) minimizes \(\beta_{\mathcal{H}_k}^2\) amongst sequences
    orthogonal to \(\{h_{t,0}\}, \ldots, \{h_{t,k-1}\}\)

- note: Riedel & Sidorenko (1995) actually used continuous parameter processes

- can approximate solutions well using
  \[
  h_{t,k} = \left\{ \frac{2}{N + 1} \right\}^{1/2} \sin \left\{ \frac{(k + 1)\pi t}{N + 1} \right\},
  \]
  which is very easy to compute!
Sine Multitapers: II

- all \( \{ h_{t,k} \} \)'s offer moderate leakage protection
- strategy & considerations for picking \( K \)
  - resolution bandwidth = \( (K + 1)/(N + 1) \) (i.e., increases with \( K \))
  - leakage relatively unchanged as \( K \) increases
  - can trade off resolution & variance:
    * decrease resolution by increasing \( K \)
    * decrease variance by increasing \( K \)
- sine tapers vs. dpss tapers
  - 1 parameter (\( K \)) vs. 2 parameters (\( 2W & K \))
  - moderate leakage protection vs. adjustable leakage protection
  - juggler resolution/variance vs.
    juggler leakage/resolution/variance
  - simple expression vs. need software to compute
- Figures 336s–41s show first 8 sine tapers
Recovery of ‘Lost Information’

- dpss tapers are solutions to $A h_k = \beta_k^2(W) h_k$
- $N$ orthonormal solutions $h_0, \ldots, h_{N-1}$
- can order via eigenvalues (concentration measure):
  $$1 > \beta_0^2(W) > \beta_1^2(W) > \cdots > \beta_{N-1}^2(W) > 0$$
- only first $K < 2NW \Delta t$ have $\beta_k^2(W) \approx 1$
- form $V = [h_0, h_1, \ldots, h_{N-1}]$
- $V^T V = I$ restates orthonormality:
  $$\sum_{t=1}^N h_{t,j} h_{t,k} = \begin{cases} 1, & j = k; \\ 0, & j \neq k. \end{cases}$$
- since $V^T = V^{-1}$, also have $VV^T = I$, yielding
  $$\sum_{k=0}^{N-1} h_{t,k} h_{t',k} = \begin{cases} 1, & t = t'; \\ 0, & t \neq t'. \end{cases}$$
- thus have (because $(*)$ is unity)
  $$\sum_{k=0}^{N-1} \sum_{t=1}^N (h_{t,k} X_t)^2 = \sum_{t=1}^N X_t^2 \underbrace{\sum_{k=0}^{N-1} h_{t,k}^2}_{(*)} \sum_{t=1}^N X_t^2$$
- Figure 345 plots $\sum_{k=0}^{K-1} h_{t,k}^2$ versus $t$
  - $K = 1, \ldots, 8$ for $NW \Delta t = 4$
  - shows relative influence of $X_t$’s (cf. $(*)$)
Multitapering of White Noise: I

- assume $X_1, \ldots, X_N$ is Gaussian white noise, mean zero, unknown variance $s_0$, sdf $S(f) = s_0$
  (setting $\Delta t = 1$ here for convenience)
- best estimate of $s_0$ is $\frac{\sum_{t=1}^{N} X_t^2}{N} = \hat{s}_0^{(p)}$
- implies best estimate of $S(f)$ is $\hat{s}_0^{(p)}$
- can obtain best estimator from $\hat{S}^{(p)}(\cdot)$ via
  \[ \int_{-1/2}^{1/2} \hat{S}^{(p)}(f) \, df = \hat{s}_0^{(p)}; \]
  i.e., ‘smoothing’ with $W_m(f) = 1$ (cf. Exercise [6.6a])
- Equation (278b) says var $\{\hat{s}_0^{(p)}\} = 2s_0^2/N$
Multitapering of White Noise: II

- let $\hat{S}^{(d)}(\cdot)$ be direct spectral estimator using $\{\tilde{h}_{t,0}\}$
- smoothing $\hat{S}^{(d)}(\cdot)$ with $W_m(f) = 1$ yields
  \[
  \int_{-1/2}^{1/2} \hat{S}^{(d)}(f) \, df = \sum_{t=1}^{N} \tilde{h}_{t,0}^2 X_t^2 = s_0^{(d)}.
  \]
- since $\text{var} \{X_t^2\} = 2s_0^2$ (Equation (40)), have
  \[
  \text{var} \{s_0^{(d)}\} = \sum_{t=1}^{N} \text{var} \{\tilde{h}_{t,0}^2 X_t^2\} = 2s_0^2 \sum_{t=1}^{N} \tilde{h}_{t,0}^4 = 2s_0^2 C_h / N
  \]
- use Cauchy inequality
  \[
  \left| \sum_{t=1}^{N} a_t b_t \right|^2 \leq \sum_{t=1}^{N} |a_t|^2 \sum_{t=1}^{N} |b_t|^2,
  \]
  with $a_t = \tilde{h}_{t,0}^2$ and $b_t = 1$ to argue $\sum_{t=1}^{N} \tilde{h}_{t,0}^4 \geq 1/N$; i.e., $C_h \geq 1$ (equality if and only if $\tilde{h}_{t,0} = 1/\sqrt{N}$)
- can conclude $\text{var} \{s_0^{(d)}\} > 2s_0^2 / N = \text{var} \{s_0^{(p)}\}$ for any nonrectangular taper
Multitapering of White Noise: III

- claim: multitapering reclaims best estimator
- let \( \{\tilde{h}_{t,0}\}, \{\tilde{h}_{t,1}\}, \ldots, \{\tilde{h}_{t,N-1}\} \) be orthonormal
- let \( \tilde{V} \) be the \( N \times N \) matrix given by

\[
\tilde{V} \equiv \begin{bmatrix}
\tilde{h}_{1,0} & \tilde{h}_{1,1} & \ldots & \tilde{h}_{1,N-1} \\
\tilde{h}_{2,0} & \tilde{h}_{2,1} & \ldots & \tilde{h}_{2,N-1} \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{h}_{N,0} & \tilde{h}_{N,1} & \ldots & \tilde{h}_{N,N-1}
\end{bmatrix}
\]

- orthonormality says \( \tilde{V}^{T} \tilde{V} = I \) & hence \( \tilde{V} \tilde{V}^{T} = I \)
- \( k \)th eigenspectrum: \( \tilde{S}_{k}^{(mt)}(f) \equiv |\sum_{t=1}^{N} \tilde{h}_{t,k}X_{t}e^{-i2\pi ft}|^{2} \)
- form \( \tilde{S}^{(mt)}(\cdot) \) by averaging all \( \tilde{S}_{k}^{(mt)}(\cdot) \)'s:

\[
\tilde{S}^{(mt)}(f) \equiv \frac{1}{N} \sum_{k=0}^{N-1} \tilde{S}_{k}^{(mt)}(f)
\]

\[
= \frac{1}{N} \sum_{k=0}^{N-1} \left( \sum_{t=1}^{N} \tilde{h}_{t,k}X_{t}e^{-i2\pi ft} \right) \left( \sum_{u=1}^{N} \tilde{h}_{u,k}X_{u}e^{i2\pi fu} \right)
\]

\[
= \frac{1}{N} \sum_{t=1}^{N} \sum_{u=1}^{N} X_{t}X_{u} \left( \sum_{k=0}^{N-1} \tilde{h}_{t,k}\tilde{h}_{u,k} \right) e^{-i2\pi f(t-u)}
\]

\[
1 \text{ if } t = u; \ 0 \text{ if } t \neq u
\]

\[
= \frac{1}{N} \sum_{t=1}^{N} X_{t}^{2} = \hat{s}_{0}^{(p)}
\]

- note: holds for any set of orthonormal tapers!
Multitapering of White Noise: IV

- as $K$ increases, can study rate of decay

$$\text{var} \{ \hat{S}^{(mt)}(f) \} = \text{var} \left\{ \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f) \right\}$$

$$= \frac{1}{K^2} \sum_{j=0}^{K-1} \sum_{k=0}^{K-1} \text{cov} \{ \hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f) \}$$

- Exer. [7.1b] gives how to compute for white noise

- Figure 350: example for $f = 1/4$ using dpss tapers
  - $N = 64; NW = 4; s_0 = 1; S(f) = 1$
  - thick curve: var $\{ \hat{S}^{(mt)}(1/4) \}$ vs. $K$
    * $K = 1$: var $\{ \hat{S}^{(mt)}(1/4) \} = S^2(f) = 1$
    * $K = N$: var $\{ \hat{S}^{(mt)}(1/4) \} = 2/N \approx 0.03$
    * curve agrees with these values
  - thin curve: computed assuming
    $$\text{cov} \{ \hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f) \} = 0 \text{ when } j \neq k$$

- thin vertical line marks Shannon number $2NW = 8$

- two curves agree closely for $K \leq 2NW$

- variance decreases slowly for $K > 2NW$
  (bias then can be bad for nonwhite processes)
Quadratic Spectral Estimators: I

• provides important motivation for multitapering

• let $X_1, \ldots, X_N$ be portion of real-valued stationary process with mean 0; sdf $S(\cdot)$; acvs $\{s_\tau\}$

• for fixed $f$, define $Z_t \equiv X_t e^{-i2\pi ft} \Delta t$

• Exer. [5.7a]: $\{Z_t\}$ stationary with $S_Z(f') = S(f+f')$ and $s_{\tau,Z} = s_\tau e^{-i2\pi f \tau} \Delta t$

• note: $S_Z(0) = S(f)$, so can estimate $S(f)$ by estimating $S_Z(\cdot)$ at $f = 0$

• let $Z$ be vector with $t$th element $Z_t$

• let $Z^H$ be its Hermitian transpose:

$$Z^H \equiv \begin{bmatrix} Z_1^*, \ldots, Z_N^* \end{bmatrix}$$

note: if $A$ real-valued matrix, then $A^H = A^T$

• since $X_t X_{t'} \Delta t$ has same units as $S(f)$, consider

$$\hat{S}^{(q)}(f) \equiv \hat{S}_Z^{(q)}(0) \equiv \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} Z_s^* Q_{s,t} Z_t = \Delta t Z^H Q Z;$$

$Q_{s,t}$ is $(s, t)$th element of weight matrix $Q$

• $\hat{S}^{(q)}(f)$ called quadratic spectral estimator
Quadratic Spectral Estimators: II

• assumptions about $N \times N$ matrix $Q$:
  
  - $Q_{s,t}$ is real-valued
  - $Q$ is symmetric; i.e., $Q_{s,t} = Q_{t,s}$
  - $Q_{s,t}$ does not depend on $\{Z_t\}$

• if $Q$ positive semidefinite (psd), then $\hat{S}^{(q)}(f) \geq 0$

• three examples of quadratic estimators
  
  - lag window estimator (need not be psd):
    
    $$
    \hat{S}^{(lw)}(f) \equiv \Delta t \sum_{\tau=-(N-1)}^{N-1} w_{\tau, m} \hat{S}_{\tau}^{(d)} e^{-i2\pi f \tau \Delta t}
    $$
    
    $$
    = \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} h_s X_s h_t X_t w_{t-s, m} e^{-i2\pi f (t-s) \Delta t}
    $$
    
    $$
    = \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} Z_s^* h_s w_{t-s, m} h_t Z_t = Q_{s,t}
    $$

  - direct spectral estimator (always psd):
    
    $$
    \hat{S}^{(d)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \Delta t} \right|^2
    $$
    
    $$
    = \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} Z_s^* h_s h_t Z_t = Q_{s,t}
    $$

  - WOSA (always psd)
**Quadratic Spectral Estimators: III**

- **goal:** set $Q$ so $\hat{S}^{(q)}(\cdot)$ unbiased & has small variance

- to get $\hat{S}^{(q)}(f) \geq 0$, assume $Q$ is psd:
  let $K = \text{rank of } Q$ & assume $1 \leq K \leq N$

- Exer. 7.2: can write $Q = AA^T$, where
  - $A$ is $N \times K$ real-valued matrix
  - $A^T A$ is $K \times K$ diagonal matrix

- $a_k = k$th column of $A$; $a_{t,k} = t$th element of $a_k$; then
  $$\hat{S}^{(q)}(f) = \Delta t Z^H AA^T Z$$

  $$= \Delta t Z^H \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_K^T \end{bmatrix} Z$$

  $$= \Delta t \begin{bmatrix} Z^H a_1 \\ Z^H a_2 \\ \vdots \\ Z^H a_K \end{bmatrix} \begin{bmatrix} a_1^T Z \\ a_2^T Z \\ \vdots \\ a_K^T Z \end{bmatrix}$$

  $$= \Delta t \sum_{k=1}^K Z^H a_k a_k^T Z = \Delta t \sum_{k=1}^K (a_k^T Z)^* a_k^T Z = \Delta t \sum_{k=1}^K |a_k^T Z|^2$$

  $$= \Delta t \sum_{k=1}^K \left| \sum_{t=1}^N a_{t,k} Z_t \right|^2 = \frac{\Delta t}{K} \sum_{k=0}^{K-1} \sum_{t=1}^N \tilde{h}_{t,k} X_t e^{-i2\pi ft \Delta t}$$

where $\tilde{h}_{t,k} \equiv a_{t,k+1} \sqrt{K}$
Quadratic Spectral Estimators: IV

- conclusion: can write all psd quadratic estimators as

\[ \hat{S}^{(q)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(q)}(f) \]

\[ \hat{S}_k^{(q)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} \tilde{h}_{t,k} X_t e^{-i2\pi ft \Delta t} \right|^2 \]

- \{\tilde{h}_{t,k}\} pairwise orthogonal because \(A^T A\) diagonal

- Q: what conditions on \(Q\) ensure \(\hat{S}^{(q)}(f)\) has good bias & variance properties?

- will study line of thought leading to dpss tapers (Bronez, 1985)
First Moment of $\hat{S}^{(q)}(\cdot)$: I

- because each $\hat{S}^{(q)}_k(f)$ is a direct estimator, have

$$E\{\hat{S}^{(q)}(f)\} = \frac{1}{K} \sum_{k=0}^{K-1} E\{\hat{S}^{(q)}_k(f)\}$$

$$= \frac{1}{K} \sum_{k=0}^{K-1} \int_{f(N)}^{f(N)} \mathcal{H}_k(f - f') S(f') df'$$

$$= \int_{-f(N)}^{f(N)} \mathcal{H}(f - f') S(f') df'$$

where

$$\mathcal{H}_k(f) \equiv \Delta t \sum_{t=1}^{N} \hat{h}_{t,k} e^{-i2\pi ft \Delta t}^2, \quad \mathcal{H}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \mathcal{H}_k(f)$$

- Exer. [7.3] gives equivalent ‘time domain’ expression:

$$E\{\hat{S}^{(q)}(f)\} = \Delta t \text{ tr } \{Q \Sigma_Z\} = \Delta t \text{ tr } \{A^T \Sigma_Z A\},$$

where tr = trace & $\Sigma_Z =$ covariance matrix for $Z_t$’s
First Moment of $\hat{S}^{(q)}(\cdot)$: II

- require $\hat{S}^{(q)}(\cdot)$ be unbiased for white noise:

$$\int_{-f(N)}^{f(N)} \tilde{H}(f') df' = 1 \iff \text{tr} \{A^T A\} = 1;$$

since $\Sigma_Z = s_0 I$, trace result follows from

$$E\{\hat{S}^{(q)}(f)\} = s_0 \Delta t = \Delta t \text{ tr} \{A^T \Sigma_Z A\} = \Delta t \text{ tr} \{A^T [s_0 I] A\} = s_0 \Delta t \text{ tr} \{A^T A\}$$

- using $a_{t,k+1} = \tilde{h}_{t,k}/\sqrt{K}$ & orthogonality, have

$$\text{tr} \{A^T A\} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{t=1}^{N-1} \tilde{h}^2_{t,k}, \text{ so unbiased if } \sum_{k=0}^{K-1} \sum_{t=1}^{N-1} \tilde{h}^2_{t,k} = K,$$

which holds under usual normalization $\Sigma \tilde{h}^2_{t,k} = 1$

- requirement provides normalization for tapers
First Moment of $\hat{S}^{(q)}(\cdot)$: III

- for general $\{X_t\}$, can get handle on first moment by incorporating notion of resolution (key idea!)
- given resolution bandwidth $2W > 0$, seek $Q$’s so

$$E\{\hat{S}^{(q)}(f)\} \approx \frac{1}{2W} \int_{f-W}^{f+W} S(f') \, df' \equiv \overline{S}(f)$$

(i.e., no longer seek $E\{\hat{S}^{(q)}(f)\} \approx S(f)$)

- rationale
  - ‘regularizes’ sdf estimation problem:
    $\overline{S}(\cdot)$ smooth to some degree; $S(\cdot)$ need not be
  - incorporates filter bandwidth in
    filtering interpretation of $S(\cdot)$ (Section 5.6)

- strategy
  - set resolution bandwidth $2W$ appropriately
  - optimize bias/variance within limitations imposed by choice of $2W$

- basically giving up finest possible resolution of $1/N \Delta t$
  to get handle on bias/variance
Broad-Band & Local Bias: I

• with estimation problem redefined, bias is

\[
 b\{\hat{S}^{(q)}(f)\} \equiv E\{\hat{S}^{(q)}(f)\} - \mathcal{S}(f)
\]

\[
 = \int_{f(N)}^{f(-N)} \overline{H}(f - f')S(f') df' - \frac{1}{2W} \int_{f-W}^{f+W} S(f') df'
\]

\[
 = \int_{f-W}^{f+W} \left[ \overline{H}(f - f') - \frac{1}{2W} \right] S(f') df'
\]

\[
 + \int_{f' \notin [f-W, f+W]} \overline{H}(f - f')S(f') df'
\]

\[
 \equiv b^{(l)}\{\hat{S}^{(q)}(f)\} + b^{(b)}\{\hat{S}^{(q)}(f)\}
\]

local bias broad-band bias

• to bound bias terms, assume \( S(\cdot) \) bounded by \( S_{\text{max}} \); i.e., \( S(f) \leq S_{\text{max}} < \infty \) for all \( f \)

• bound on magnitude of local bias:

\[
|b^{(l)}\{\hat{S}^{(q)}(f)\}| \leq \int_{f-W}^{f+W} \left| \overline{H}(f - f') - \frac{1}{2W} \right| S(f') df'
\]

\[
 \leq S_{\text{max}} \int_{-W}^{W} \left| \overline{H}(f'') - \frac{1}{2W} \right| df'';
\]

integral gives useful measure of local bias

• local bias small if \( \overline{H}(f) \approx 1/2W \) over \([-W, W]\)
Broad-Band & Local Bias: II

- bound on broad-band bias (must be positive!):
\[
b^{(b)} \{\hat{S}^{(q)}(f)\} = \int_{f' \notin [f-W,f+W]} \hat{H}(f - f') S(f') \, df'
\[
\leq S_{\text{max}} \int_{f' \notin [f-W,f+W]} \hat{H}(f - f') \, df'
\[
= S_{\text{max}} \int_{f' \notin [-W,W]} \hat{H}(f) \, df''
\[
= S_{\text{max}} \left( \int_{f(\langle N \rangle)}^{f(\langle N \rangle)} \hat{H}(f') \, df'' - \int_{-W}^{W} \hat{H}(f') \, df'' \right)
\[
= S_{\text{max}} \Delta t \left( \text{tr} \{ A^T \Sigma^{(bl)} A \} - \text{tr} \{ A^T \Sigma^{(bl)} A \} \right),
\]

where \( \Sigma^{(bl)} \) arises from the following argument:

- suppose \( \{X_t\} \) is band-limited white noise; i.e.,
  \( S^{(bl)}(f) \equiv \begin{cases} 1, & |f| \leq W; \\ 0, & W < |f| \leq f(\langle N \rangle), \end{cases} \)

and acvs
\[
s^{(bl)}_\tau \equiv \begin{cases} 2W, & \tau = 0; \\ \sin \left( 2\pi W \tau \Delta t \right) / \left( \pi \tau \Delta t \right), & \tau \neq 0. \end{cases}
\]

- for this sdf (and letting \( f = 0 \) so \( Z_t = X_t \)), have
\[
E\{ \hat{S}^{(q)}(0) \} = \int_{f(\langle N \rangle)}^{f(\langle N \rangle)} \hat{H}(0 - f') S^{(bl)}(f') \, df'
\[
= \int_{-W}^{W} \hat{H}(f'') \, df'' = \Delta t \text{ tr} \{ A^T \Sigma^{(bl)} A \},
\]

where \( (j, k) \)th element of \( \Sigma^{(bl)} \) is \( s^{(bl)}_{j-k} \).
Minimizing Broad-Band Bias Measure

- measure of broad-band bias (leakage) is thus
  \[ \text{tr} \{ A^T A \} - \text{tr} \{ A^T \Sigma^{(bl)} A \} \]

- setting \( \text{tr} \{ A^T A \} = 1 \) ensures unbiasedness for white noise

- to minimize broad-band bias under this restriction,
  maximize \( \text{tr} \{ A^T \Sigma^{(bl)} A \} \) subject to \( \text{tr} \{ A^T A \} = 1 \)

- Exer. [7.4] gives solution:
  - set \( K = 1 \)
  - \( A = a_1 \) is normalized eigenvector associated with largest eigenvalue \( \lambda_0(N, W) \) of \( \Sigma^{(bl)} \)
  - eigenvector is dpss of 0th order
    (technically: finite subsequence of dpss)
  - broad-band bias measure \( = 1 - \lambda_0(N, W) \)
    \( (\lambda_0(N, W) = \text{concentration ratio}) \)

- solution conflicts with variance in white noise case:
  as \( K \) increases, variance decreases
Managing Bias & Variance

• reasonable balance: use $K$ orthonormal dpss tapers
  – broad-band bias: measure given by Exer. [7.5]:
    \[ 1 - \frac{1}{K} \sum_{k=0}^{K-1} \lambda_k(N, W); \]
    \[ \lambda_k(N, W) \] close to unity as long as $K < 2NW \Delta t$
  – variance: Section 7.4 argues that approximately
    \[ \hat{S}^{(mt)}(f) \] is approximately
    \[ \frac{S(f)}{2K} \chi^2_{2K} \]
    if $S(\cdot)$ not rapidly varying over $[f - W, f + W]$;
    thus have \( \text{var} \{ \hat{S}^{(mt)}(f) \} \approx S^2(f) / K \)
  – local bias: as $K$ increases, local bias decreases
    (cf. Figures 340–1: $NW = 4$ with $N = 1024$
    \( \implies 1/2W = N/8 = 128 \div 21 \text{ dB} \))
Adaptive Multitaper Estimation: I

- Section 7.4 gives refinement to basic multitapering (developed for dpss tapers)
- idea: weight eigenspectra adaptively according to need for leakage suppression at each $f$
  - if $S(f)$ relatively large, leakage not a concern
    $\Rightarrow$ can make $K$ large
  - if $S(f)$ relatively small, leakage is a concern
    $\Rightarrow$ should make $K$ small
- adaptive multitaper estimator given by
  $$\hat{S}^{(amt)}(f) \equiv \frac{\sum_{k=0}^{K-1} b_k^2(f)\lambda_k \hat{S}_k^{(mt)}(f)}{\sum_{k=0}^{K-1} b_k^2(f)\lambda_k}$$
  where $\lambda_k \approx 1 - 1/10^j$ (with $j \downarrow$ as $k \uparrow$) &
  $$b_k(f) = \frac{1}{\lambda_k + (1 - \lambda_k)\frac{s_0 \Delta t}{S(f)}} \approx \frac{1}{1 + \frac{s_0 \Delta t}{10^j S(f)}}$$
  - $\lambda_k$’s downweight higher eigenspectra (slightly)
  - $s_0 \Delta t = \text{average value of } S(\cdot)$
  - $b_k(f)$ small if $10^j S(f) \ll s_0 \Delta t$
  - $b_k(f)$ large if $10^j S(f) \gg s_0 \Delta t$
- determine $b_k(f)$ using preliminary estimate of $S(\cdot)$; can iterate to refine $b_k(f)$’s if desired
Adaptive Multitaper Estimation: II

• assume
  \[ \hat{S}_k^{(mt)}(f) \overset{d}{=} S(f) \chi^2_{2}/2 \] for each eigenspectrum
  \[ \hat{S}_k^{(mt)}(f) \]’s are pairwise uncorrelated

• as before, assume \( \hat{S}^{(amt)}(f) \overset{d}{=} a\chi^2 \)

• edof argument similar to \( \hat{S}^{(lw)}(\cdot) \) & \( \hat{S}^{(WOSA)}(\cdot) \) yields

\[
\nu = \frac{2 \left( E\{ \hat{S}^{(amt)}(f) \} \right)^2}{\text{var} \{ \hat{S}^{(amt)}(f) \}} \approx \frac{2 \left( \sum_{k=0}^{K-1} b^2_k(f) \lambda_k \right)^2}{\sum_{k=0}^{K-1} b^4_k(f) \lambda_k^2}
\]
Example: Ocean Wave Data

- \( N = 1024; \Delta t = 1/4 \) second
- Figure 373a: basic multitaper estimate \( \hat{S}^{(mt)}(\cdot) \)
  - set \( NW = 4 \) (resolution not main concern)
  - maximum of 7 possible reasonable tapers, but \( \hat{S}^{(mt)}_6(\cdot) \) poor at high frequencies
  - set \( K = 6 \), yielding \( \nu = 12 \) edof
  - width of crisscross = 2\( W \)
- Figure 373b: 2nd \( \hat{S}^{(mt)}(\cdot) \) (thick curve)
  - set \( NW = 6; K = 10 \) so \( \nu = 20 \)
  - thin curve: \( m = 150 \) Parzen estimate (Fig. 301a)
    (bandwidth \( \approx 0.049 \) Hz \( \approx 2W \approx 0.047 \) Hz)
  - good agreement between \( \hat{S}^{(lw)}(\cdot) \) and \( \hat{S}^{(mt)}(\cdot) \)
- Figure 373c: adaptive estimate (thick curve)
  - \( NW = 4 \) with \( K = 7 \)
  - agrees well with 373a between 0 \& 1 Hz
  - more structure for \( f > 1 \) Hz due to \( \nu \downarrow \)
    (cf. Figure 373d, which plots \( \nu \) vs. \( f \))
  - thin curves: 95\% confidence intervals