(34) FOURIER ANALYSIS & APPLICATIONS - Noise & Variability IV

Nonstationary Processes.

Stationarity allowed us two things:

① We could obtain new samples from Ω simply by waiting a long time & sampling again

② We could simplify the characterization of moments, since $\langle x(t)\,x(t+\tau)\rangle$ depended only on τ.

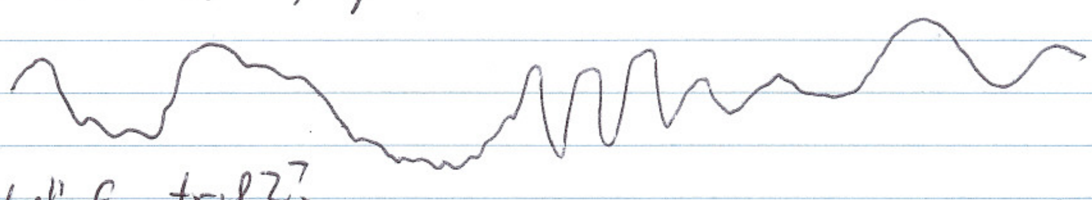
Now we relax both of these assumptions.

Still need a way to obtain multiple samples of Ω e.g., "trials", each trial has a start time [or, parallel preps!]

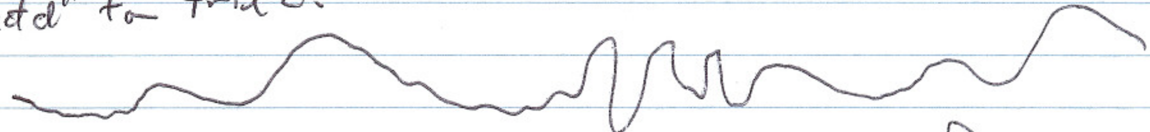
Still need to deal with the inter-dependence of direct time-domain estimates

$$\left.\begin{array}{l}\langle x(t)\,x(t+\tau)\rangle \\ \langle x(t)\,x(t+\tau')\rangle \\ \langle x(t+\tau)\,x(t+\tau')\rangle\end{array}\right] \text{ are mutually dependent.}$$

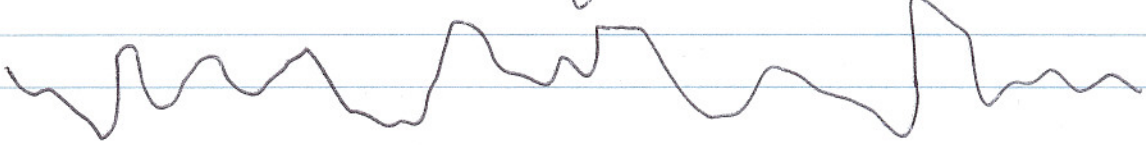
Still want to try to use intuitions about general properties of biologic systems. E.g., say you observe

trial 1 $x(t) =$ 

Which is more "expected" for trial 2?



vs

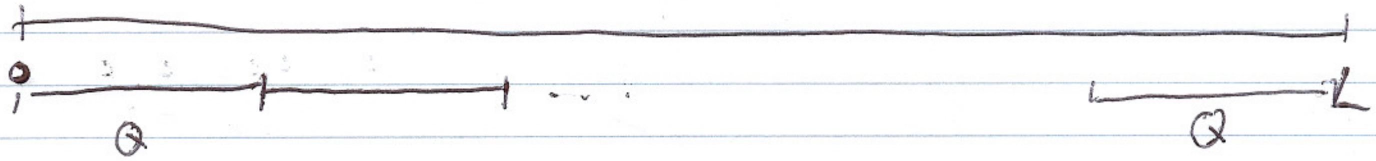


The intuition is that there are gradual changes in system properties over time, & interactions tend to be local.

This has a natural embodiment in terms of spectral estimation:

Get the best local spectral estimate, & see how it changes over time.

Full "trial"



Q = "quasi-stationary interval"

So everything we said before about getting the best estimate in length L still applies, but now over a period of time Q:

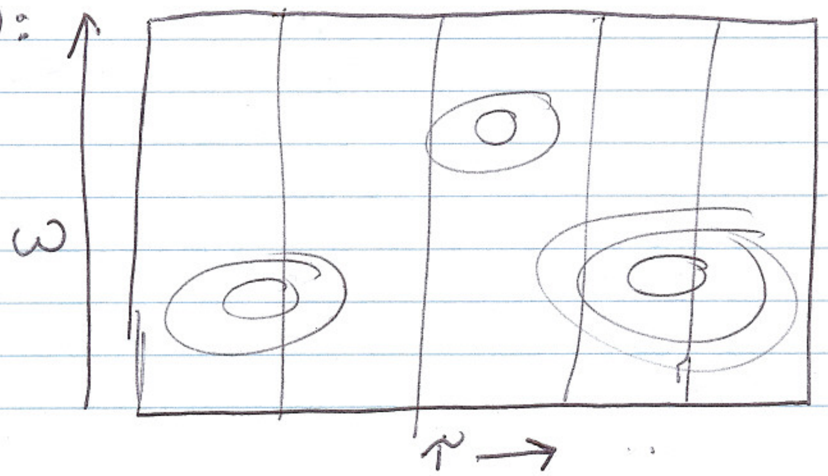
Multitaper estimate beginning at time τ, for signal from $t=\tau$ to $\tau+Q$:

$$P_X(\omega,\tau) = \frac{1}{K}\sum_{j=1}^{K} \left| \int_0^Q x(t+\tau)\, W_j(t,Q)\, e^{-i\omega t}\, dt \right|^2$$

where the $W_j(t,Q)$ are multitaper functions on $[0, Q]$

spectrogram

$P_X(\omega,\tau)$:

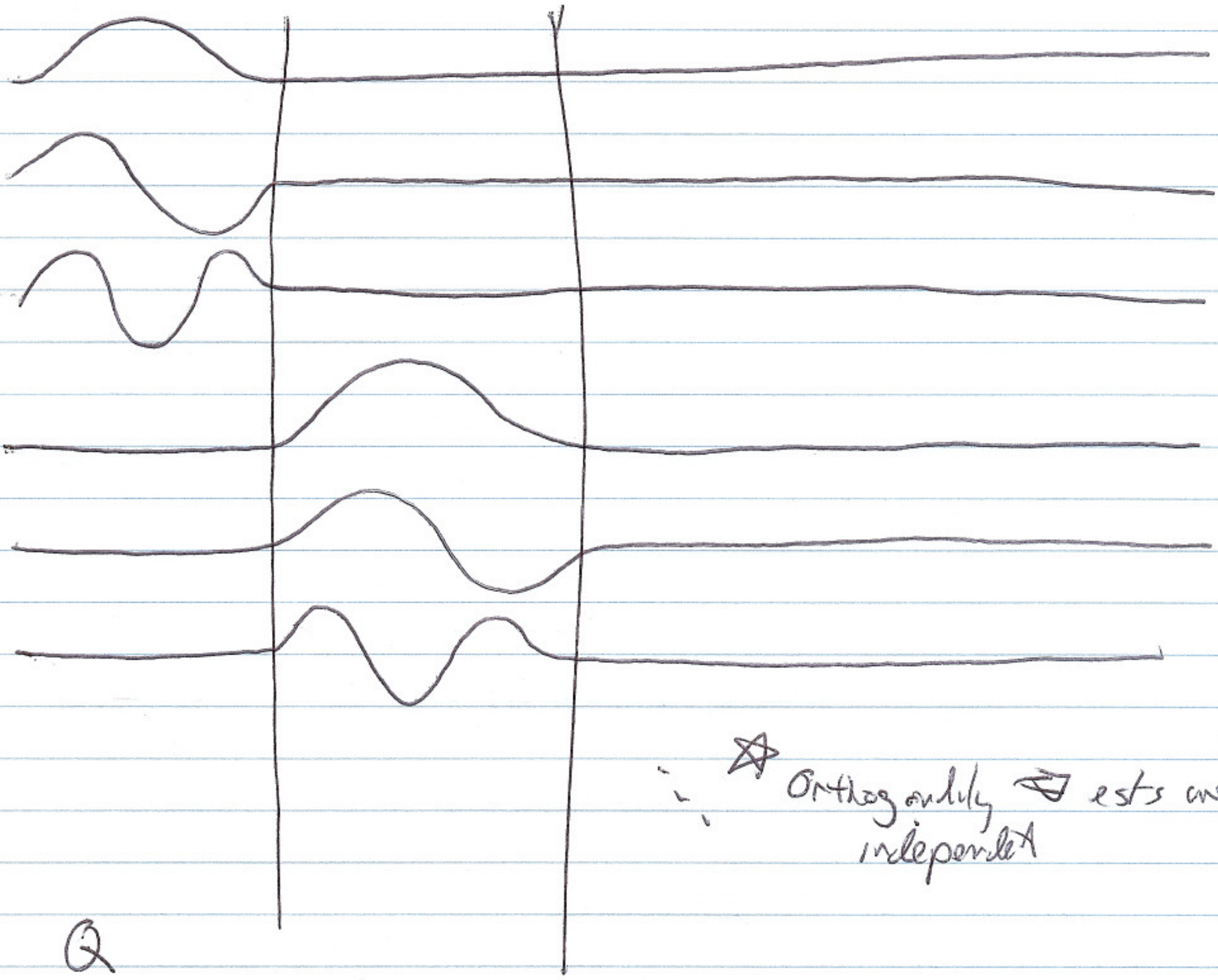


Frequency resolution is $\frac{2\pi K}{Q}$.

Time resolution is Q.

Larger $K \rightarrow$ better ests but worse resolution

Essentially we've made multiple local estimates of the power spectrum by multiplying $X(t)$ by tapers + their shifts:

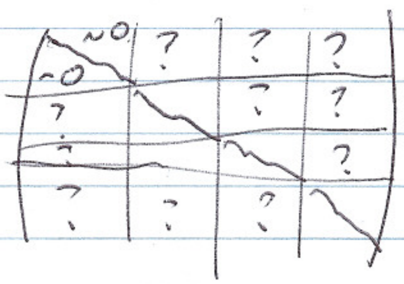


and calculated Fourier estimates at frequencies $\gg \frac{2\pi}{Q}$.

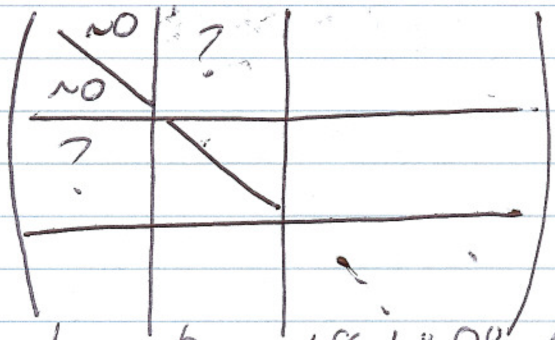
Looked at more abstractly, we are hoping to the full covariance matrix of X by a change of basis - "local Fourier components" -

"local" time-invariance means that within each block, the covariance matrix is diagonal.

Across blocks structure is not known -



What about across the blocks?



Could (i) reanalyze with a different "Q" to look for longer-time structure

(ii) "look at the results" - ROI's

dimensional reduction methods

(iii) spectrum of spectra

Rel. of spectrogram to spectrum -

With proper normalization (e.g, Chronux spectrogramc.m, etc.)

$$\int P_x(\omega,\tau)\,d\tau = P_x(\omega)$$

$$\int P_x(\omega,\tau)\,d\omega = \langle x(\tau)^2\rangle$$

The nice properties of spectra generalize: if X_1 & X_2 are independent

$$P_{X_1+X_2} = P_{X_1} + P_{X_2}$$

For linear filters that are slow on a timescale Q, $X \rightarrow \boxed{L} \rightarrow Y$

$$P_y(\omega,\tau) = |L(\omega)|^2 P_x(\omega,\tau)$$

& even $P_y(\omega,\tau) \approx |L(\omega,\tau)|^2 P_x(\omega,\tau)$ if L changes slowly

Things generalize to cross-spectra & coherence

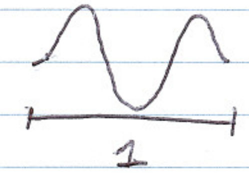
Cross spectrum: $P_{XY}(\omega) \cong \frac{1}{K}\sum_{j=1}^{k}\left(\int_0^T x(t)\,w_j(t,T)e^{-i\omega t}\,dt\right)\overline{\left(\int_0^T y(t)\,w_j(t,T)e^{-i\omega t}\,dt\right)}$

becomes

$$P_{XY}(\omega,\tau) \cong \frac{1}{K}\sum_{j=1}^{K}\left(\int_0^Q x(t+\tau)W_j(t,Q)e^{-i\omega t}\,dt\right)\overline{\left(\int_0^T y(t+\tau)W_j(t,Q)e^{-i\omega t}\,dt\right)}$$

- $\dfrac{P_{XY}(\omega,\tau)}{\sqrt{P_X(\omega,\tau)}\sqrt{P_Y(\omega,\tau)}}$ (coherency)

Alternative approaches

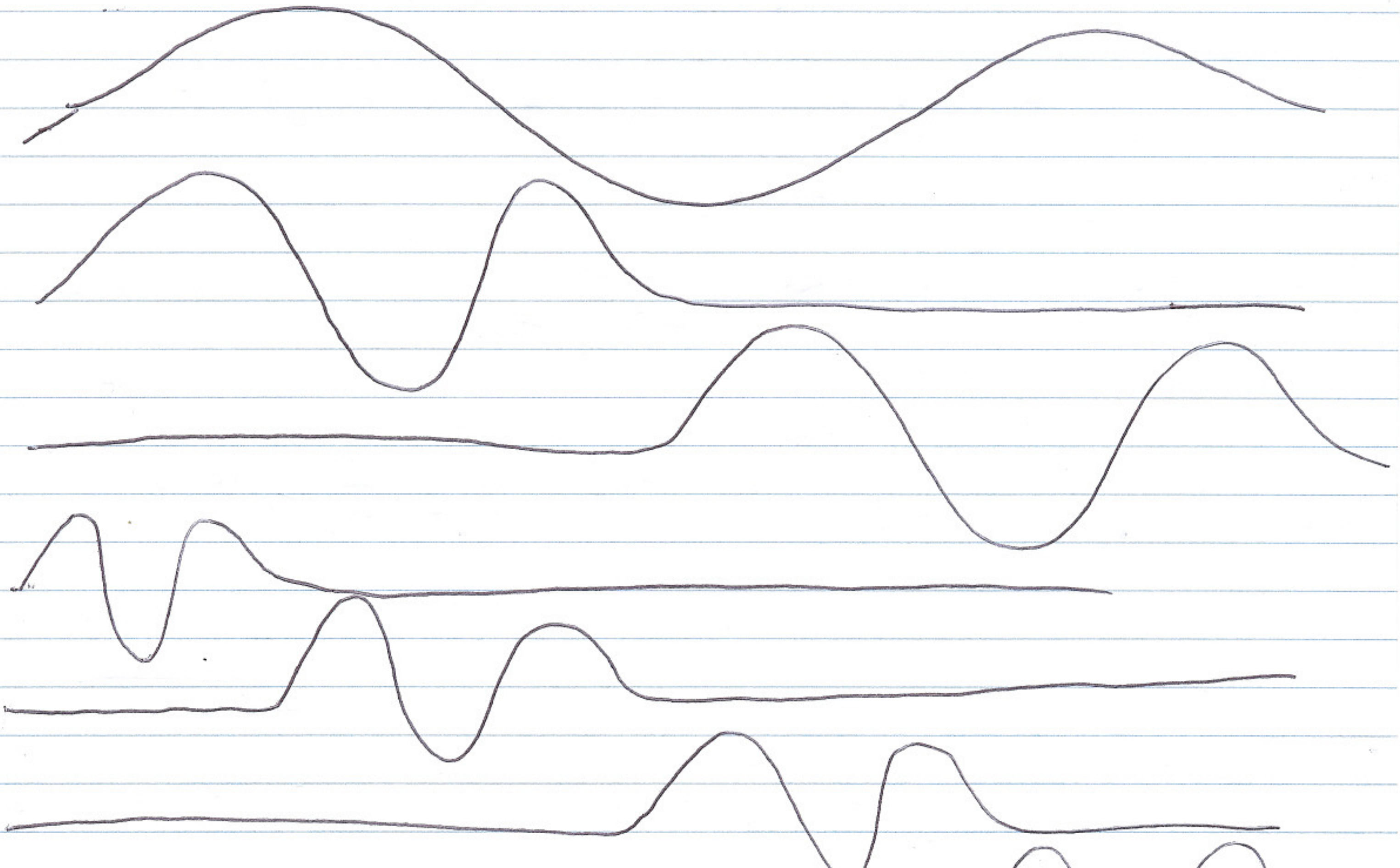
Wavelets Choose a "mother wavelet" $M(t)$, e.g. 

+ calculate

$$M_X(\tau, s) = \int X(t+\tau)\, M(t/s)\, dt = \int X(t)\, M\left(\frac{t-\tau}{s}\right) dt$$

for a range of values of s, typically in powers of 2, and for values of τ spaced by K/s. Highest $s = T$.

Essentially, projecting X onto

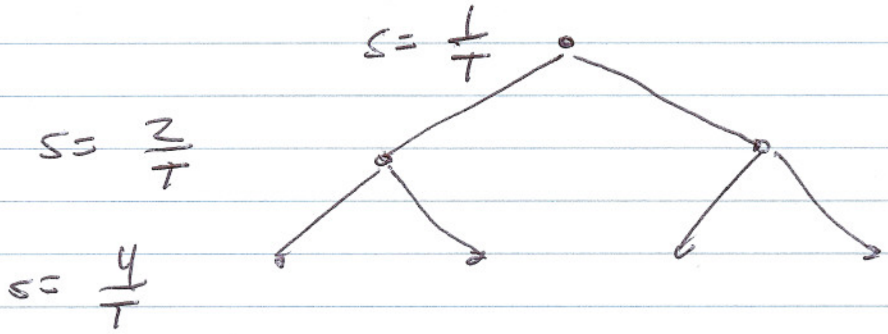


etc.

(40)

The wavelets form a basis, so

$M_X(\tau, s)$ has transformed X into another basis set



So $\langle |M_X(\tau, s)|^2 \rangle$ is analogous to $P_X(\tau, \frac{1}{s})$

• one can also calculate $\langle M_X(\tau, s)\, \overline{M_Y(\tau, s)} \rangle$, etc.

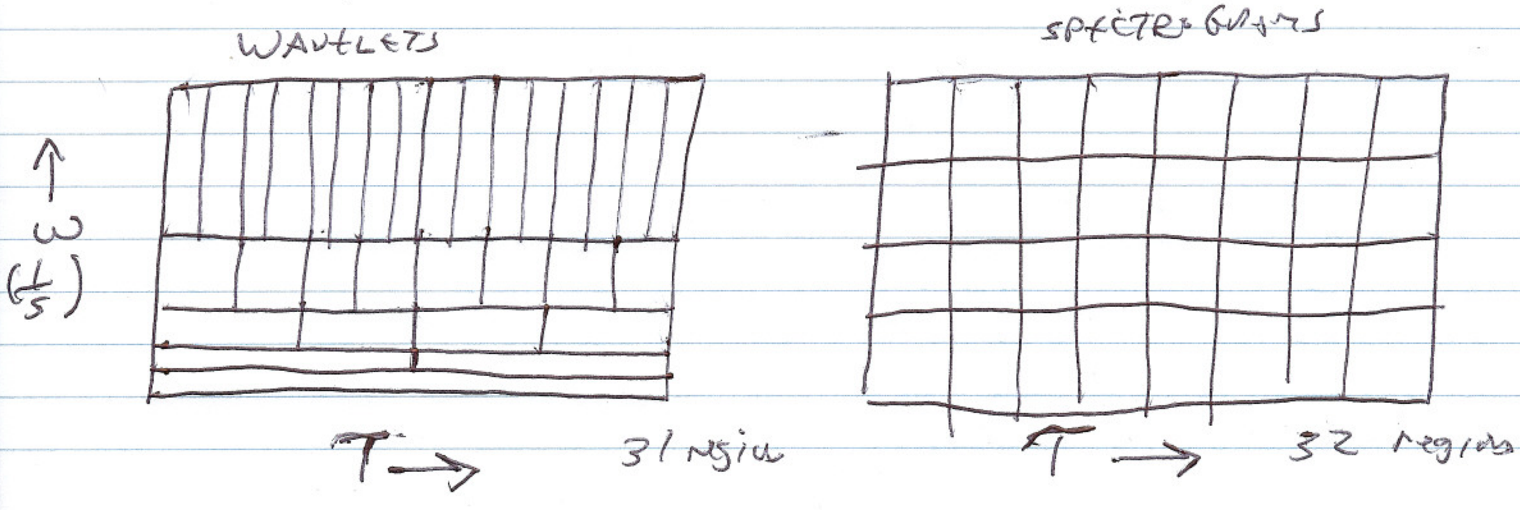
Compare & contrast:

The wavelets are not orthogonal. They are not quite as confined in frequency space.

↓ so ests are not independent

↓ so you could get fooled by oscillations.

But they are an alternative way to "tile" time-freq space



With wavelets, increasing freq resolution → decr time resolution, shape is controlled by "mother" wavelet – typically Gaussian × sine
× cos
× polynomial

No provision to exploit local time-invariance with wavelets – i.e., in MT method, values within each "column" on the time-freq. decomposition are ~ independent

However – if it is ~ true that shorter events have higher frequency, then wavelets will be more efficient representations.

E.g., natural scenes.

But ? natural sounds
? brain activity

$\downarrow$

* Group theoretic interpretation:
$s(t) \to s(t+\tau)$
$s(t) \to s(\alpha t)$
[two transformations]
* Look for a representation in which these transformations are "simple"
* $t \to t+\varphi \mapsto$ translate
$t \to \alpha t \mapsto$ move up the stack of wavelets

MT – wavelet ideas generalize to spatial signals

local translation invariance $\longrightarrow$ MT

+ approx. independence of spectra at long-range

+ scaling

$\downarrow$

Wavelet

Another approach, specific to time series

"Autoregressive Models"

(Stationary case first)

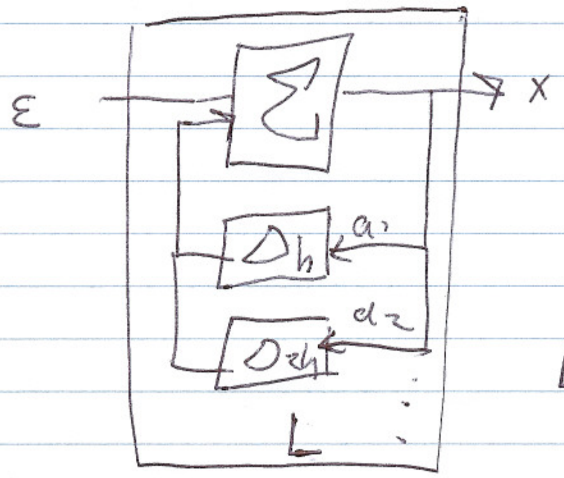
basic idea is to model the spectrum as the result of white noise acted on by a linear filter

$$Z \longrightarrow \boxed{L} \longrightarrow X$$

Z is unobservable. So we attempt to write

$$x(t) = a_1 x(t-h) + a_2 x(t-2h) + \cdots + a_r x(t-rh) + \varepsilon(t)$$

and minimize $\langle \varepsilon(t) \rangle^2$.



Power spectrum of $X = P_\varepsilon(\omega)\,|L|^2$

$P_\varepsilon(\omega)$ assumed to be flat, $= \langle \varepsilon^2 \rangle$.

$$L = \frac{1}{1 - (a_1 e^{-i\omega h} + a_2 e^{-2i\omega h} + \cdots)}$$

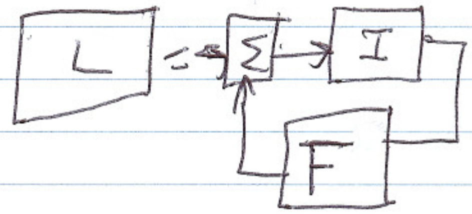
Put $\theta = e^{i\omega h}$

$$P_X(\omega) \approx \langle \varepsilon^2 \rangle \frac{1}{|1 - \Sigma a_k \theta^k|^2}$$

Note that the model samples x at intervals h.

So it will only capture frequencies $<< \frac{2\pi}{h}$. Equivalently, $L(\omega)=L(\omega+\frac{2\pi}{h})$.

Also, recall that for a feedback system,



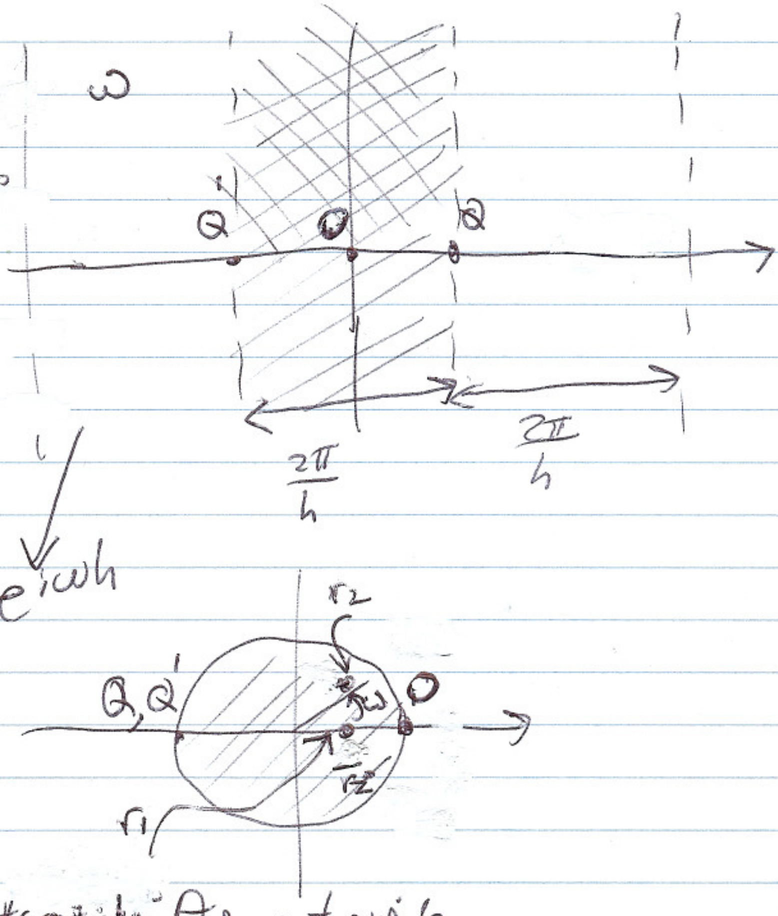
$$\tilde{L}(\omega)=\frac{1}{1-\tilde{F}(\omega)}$$

is stable only if $|\tilde{F}(\omega)|<1$ for all ω on real axis.

(More generally, on complex plane, $\tilde{L}(\omega)$ must be $<\infty$ for $\operatorname{Im}\omega<0$.)

$L(\omega)$ on the complex plane is periodic and must be bounded in lower half plane

$\theta=e^{i\omega h}$ maps the region marked by into the interior of the unit circle

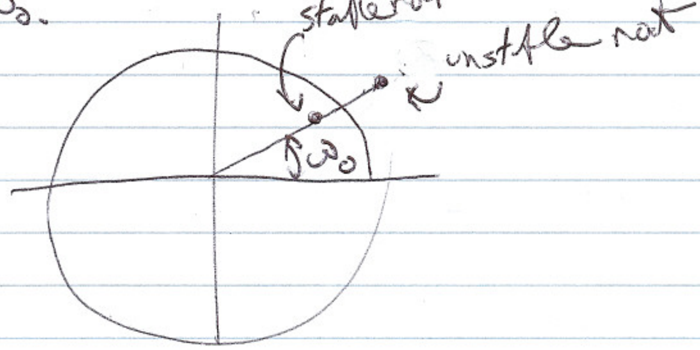


$$P_x(\omega)=\frac{K\varepsilon^2 h}{|1-A(\theta)|^2}$$

$A(\theta)$ is a polynomial, ∴

$A(\theta)=1$ can't have roots inside the unit circle.
Roots can come singly (as r_1) or in cc pairs $(r_2, \bar{r}_2)$.

A root near the unit circle corr. to a peak in the spectrum at the corresponding frequency ω_0.



Bottom line - only stereotyped spectra can be well-modeled for low-degree $A(\theta)$! peak width + peak height are coupled.

But if the spectrum is of the form $\dfrac{\langle\varepsilon^2\rangle}{|1-A(\theta)|^2}$,

then the procedure works well.

How to determine r and the a_r?

"Yule-Walker" equations: Minimize

$$R = \sum_t \left(x(t) - \left(a_1 x(t-h) + a_2 x(t-2h) + \dots + a_r x(t-rh)\right)\right)^2$$
$$\left(= \sum \varepsilon(t)^2\right)$$

This is a linear system: $x(t)$ is regressed against lags of $x(t)$

Criterion for adding another term: R will decrease when another term is added, but is this decrease more than one would expect if the $\varepsilon(t)$'s are independently drawn from a Gaussian?

(41)

Minimize R : put $\frac{\partial R}{\partial a_k} = 0$, this leads to

$$a_1 c_0 + a_2 c_1 + a_3 c_2 + \cdots + a_r c_{r-1} = c_1$$
$$a_1 c_1 + a_2 c_0 + a_3 c_1 + \cdots + a_r c_{r-2} = c_2$$
$$a_1 c_2 + a_2 c_1 + a_3 c_0 + \cdots + a_r c_{r-3} = c_3$$
$$\vdots$$
$$a_1 c_{r-1} + a_2 c_{r-2} + a_3 c_{r-3} + \cdots + a_r c_0 = c_{r-1}$$

Note diagonal structure.

Several strategies for assessing tradeoff of # of params and reducing R –

- reduced chi-squared
- Akaike criterion $2r + N \ln R$ (N points)

Nonstationary

fit spectral estimate to moving windows.

Multivariate

fit to multiple time series $x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_k(t) \end{pmatrix}$; a's are each $k \times k$ matrices

Autoregressive-moving average:

can add "moving average" terms to model, e.g.,

$$x(t) = a_1 x(t-h) + \cdots + a_r x(t-rh)$$
$$+ b_0 \varepsilon(t) + b_1 \varepsilon(t-h) + \cdots + b_s \varepsilon(t-sh)$$

46

"Granger Causality" is just multivariate autoregression --
how well can one predict X(t) from the past history of X and Y, vs how well can one predict X(t) from X(t) alone.