Periodogram lecture

Proakis & Manolakis 14.1 + 14.2

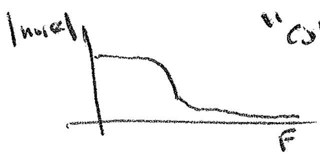
Outline:

1) Big picture / motivation
2) Math preliminaries
3) Simple approach – take DFT of a windowed section (just like for deterministic signal), adjust amplitude
   – problem w/ variance ~~amplitude~~
4) Periodogram averaging (Bartlett)
   – variance reduction
   – tradeoff: stability vs. resolution
5) Welch method
6) other methods

---

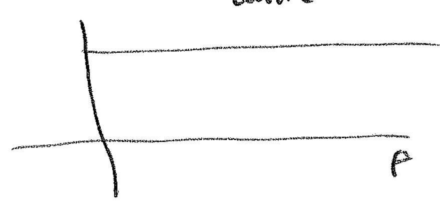
1) Big picture / motivation

Let's say we have a random/noise process with a particular spectrum:

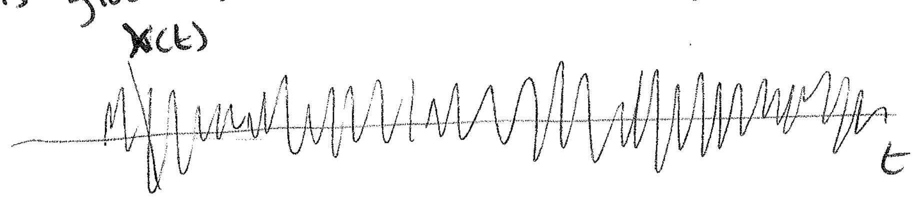


"colored"

"white"



this gives us a random time series:



We may want to estimate the spectrum magnitude given the time series.

Why?
– figure out where noise is, for frequency-selective filtering
– system identification (next page)

System ID:

$$x \to \boxed{h} \to y$$

$$Y(\omega) = H(\omega)\, X(\omega)$$

so,

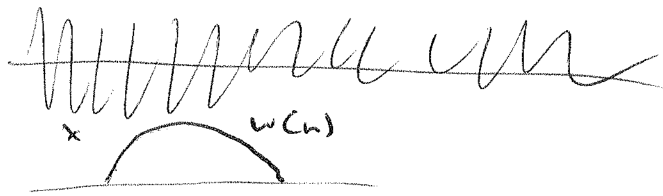
$$|Y(\omega)|^2 = Y Y^*$$
$$= H(\omega) H^*(\omega)\, X X^*$$
$$= |H(\omega)|^2\, |X(\omega)|^2$$

so if we can estimate $|X(\omega)|^2$ + $|Y(\omega)|^2$ we can find $|H(\omega)|^2$

how could we do this?

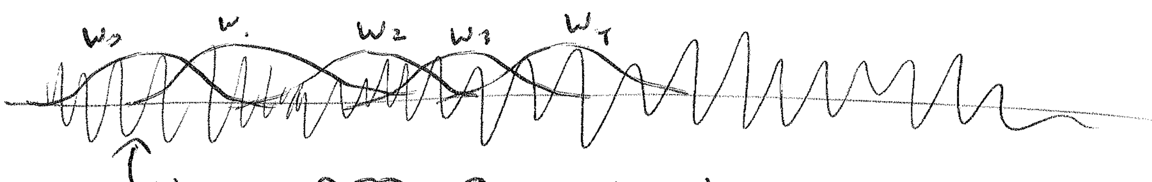
approach 1: do the same thing we did for deterministic ~~Pm~~ signals.

pick a time window, apply $w(n)$, take DFT



problem: variance is very high

approach 2: break signal into many windows, possibly overlapped, possibly non-rectangular weights, then average the |DFTs| of each



take DFT from w_0, w_1, etc.
average all the magnitudes of each DFT

More math preliminaries

<u>Random processes</u> used to model very complex or noise-like signals.

characterized by statistical properties

average
variance
autocorrelation
power density spectrum

examples:

Sample mean: $\hat{m}_x = \frac{1}{L} \sum_{n=0}^{L-1} x(n)$

sample variance: $\hat{\sigma}_x^2 = \frac{1}{L} \sum_{n=0}^{L-1} (x(n) - \hat{m}_x)^2$

do on board

we would like these estimators to be

<u>unbiased</u>: $\mathcal{E}$(estimator) = true value
↳ expected value of

weaker: <u>asymptotically unbiased</u>: $\mathcal{E}$(estimator) → true value as $L \to \infty$

~~const~~ <u>consistent</u>: variance(estimator) → 0 as $L \to \infty$

<u>Expected value</u> means we repeat the estimate many times and average results

classic results for estimating mean variance of Gaussian variable w/ mean μ, variance σ^2, L points

$\mathcal{E}(\hat{m}_x) = \mu$ (so sample mean is unbiased)

$\mathcal{E}(\mathrm{var}(\hat{m}_x)) = \sigma^2/L$ (so " " is consistent auto-correlation (1)

new quantity: auto-correlation + how it relates to spectra.

(1) def: the $\boxed{\text{cross-correlation}}$ of 2 signals is

$$\underset{\text{delay}}{r_{xy}(\ell)} = \sum_{n=-\infty}^{\infty} x(n)\,y(n-\ell) = \sum_{n=-\infty}^{\infty} x(n+\ell)\,y(n)$$

shift the signals by ℓ samples, multiply, sum

Closely related to correlation

$$r_{xy}(\ell) = \sum_{n=-\infty}^{\infty} x(n)\,y(n-\ell) \quad \Big| \quad x * y = \sum_{n=-\infty}^{\infty} x(n)\,y(\ell-n)$$

by inspection, if we flip argument of y, they match.

So, $r_{xy}(\ell) = x(n) * y(-n)$

autocorrelation is correlation of signal w/ itself:

$$r_{xx}(\ell) = \sum_{n=-\infty}^{\infty} x(n)\,y(n-\ell) = x(n) * x(-n)$$

sometimes we may want to compute time-averaged autocorrelation (P+M eq 14.1.24)

$$r'_{xx}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n)\,x(n-m)$$

(basically, average by # pts)

(2) Link between autocorrelation and spectra:

$$|X(\omega)|^2 = X(\omega)\, X^*(\omega)$$
$$= X(\omega)\, X(-\omega) \quad \text{(if conj. symmetric)}$$

thus $|X(\omega)|^2 = X(\omega)\,X(-\omega) \xLeftrightarrow[\text{DTFT}]{} x(n) * x(-n)$

by time reversal property of DTFT

but, $x(n) * x(-n) = r_{xx}$!

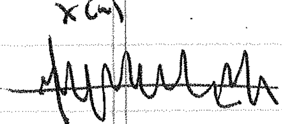
so $|X(\omega)|^2 \Longleftrightarrow r_{xx}(\ell)$

ie., mag-squared spectrum is F.T. of the autocorrelation

(3) Why is white noise "white" (all frequencies?)

Consider a zero-mean signal, each time sample unrelated to any others.

$x(n)$

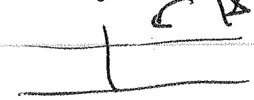


then, $r_{xx}(\ell) = \begin{cases} \sum x(n)^2, & \ell = 0 \\ 0 & \text{else} \end{cases}$, on average

so $\mathcal{E}(r_{xx}(\ell)) = N\sigma^2\,\delta(\ell)$

$\Updownarrow$

$|X(\omega)|^2 = N\sigma^2$



ω

if we use time-averaged r'_{xx}, get $\mathcal{E}(r'_{xx}(\ell)) = \sigma^2\delta(\ell)$

2) math preliminaries

some notation (Chapter 3 2.1.2)

an energy signal is one w/ finite energy

$$E = \sum |x(n)|^2$$

(for example, finite-length signals)

a power signal

has finite power: $P = \lim_{N \to \infty} \frac{1}{2N+1} E_N$ is $< \infty$

which is for example, an infinitely long signal that is well-behaved.

Energy spectral density is mag^2 of an energy signal

$S_{xx}(F) = |X(F)|^2$ ← shows how energy is distributed across

Power spectral density shows how energy in a power signal is distributed

$\Gamma_{xx}(F)$. mathematically, can define as F.T. of autocorrelation

In 14.1.1 P&M discuss energy signals

in 14.1.2, P&M discuss power signals / spectra

for random processes, we need to talk about PSD

P&M notation

	ESD	PSD
true spectrum	$S_{xx}(f)$	$\Gamma_{xx}(f)$
estimate	$S_{\tilde{x}\tilde{x}}(f)$	$P_{xx}(f)$

3) Simple approach (which doesn't work)

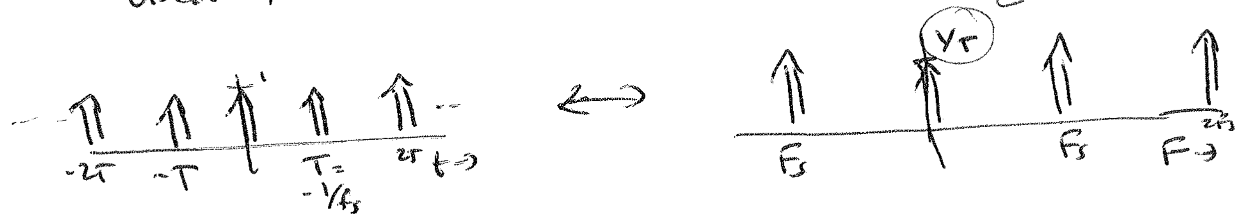
a) First, let's establish the link between the <u>continuous-time spectrum</u> and the <u>sampled-time spectrum</u>

~~because~~

$$x_a(t) \rightarrow \boxed{ADC} \rightarrow x(n), \quad \uparrow T$$

<u>Sampling theory refresher</u>:

earlier, we saw that when going from $x_a(t)$ to $x(n)$, we used the transform pair:



this gave us the sampled spectrum:

$$X(\omega) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_a(F - kF_s)$$

where $f = F/F_s$

~~but~~ $\omega = 2\pi f$, so above is same as P&M eq 14.1.6

assuming there is <u>no aliasing</u>, $X(\omega) = \frac{1}{T} X_a(F)$, $|\omega| < \pi$

so the sampled spectrum

$$S_{xx}(\omega) = |X(\omega)|^2 = \left(\frac{1}{T}\right)^2 |X_a(F)|^2$$

so the sampled spectrum is just a scaled version of the true spectrum - thus it's <u>useful</u>

skip during lecture

b) now, let's look at actually calculating the spectrum. A possible flow chart is:

APPROACH 1

$$x(n) \rightarrow \otimes \xrightarrow{v(n)} \boxed{DFT} \xrightarrow{V(k)} \boxed{|\ |^2} \rightarrow |V(k)|^2 = |V(\omega)|^2\Big|_{\omega = \frac{2\pi k}{L}}$$

$\uparrow$ $w(n)$, length L

under approach 1, $V(\omega) = X(\omega) * W(\omega)$

so window smooths the true spectrum. note if we use an infinitely long rectangular window, $W(\omega) \Rightarrow \delta(\omega)$

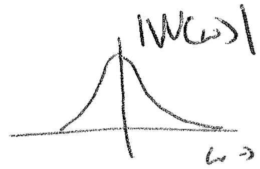
so $V(\omega) \to X(\omega)$. ← $\boxed{\text{asymptotically unbiased}}$

discussion, P+M end of 14.1.1

how about <u>amplitude</u>? we may want absolute values (not relative) in ~~some~~ many applications.

ex: if estimating vibration spectrum, we care about the values! Don't want levels to change if $w(n)$ changes

we have $V = X * W$



✱ to keep the window from changing the energy, we want to <u>normalize</u> it's energy.

By Parseval, $\frac{1}{2\pi}\int_{-\pi}^{\pi} |W(\omega)|^2 = \sum_{n=0}^{N-1} w^2(n) = E_w$

rectangular window:

$$\sum_{n=0}^{N-1} w^2(n) = \sum_{n=0}^{N-1} 1^2 = N$$

in general, define

$$UN = \sum_{n=0}^{N-1} w^2(n), \quad \text{or} \quad U = \frac{1}{N}\sum_{n=0}^{N-1} w^2(n)$$

then we can recover the true level by

$$P_{\tilde{x}\tilde{x}}(\omega) = \frac{1}{UN}\left|\sum_{n=0}^{N-1} (w(n)x(n))\, e^{-j\omega kn/N}\right|^2$$

the <u>periodogram</u> ↗

↑ $\tilde{\ }$ means x is distorted by window

$U=1$ for rectangular
otherwise, calculate it.

Approach 2: autocorrelation

we can write the mag-squared of a signal $x(n)$ as

$$|X(\omega)|^2 = X(\omega)X^*(\omega) = X(\omega)X(-\omega)$$

real signals

but this Fourier transforms to:

$$X(\omega)X(-\omega) \longleftrightarrow x(n) * x(-n)$$

$$= \sum_{n=0}^{K-1} x(n)x(n+m)$$

(if x has K samples)

$$= r_{xx}(m) \quad \leftarrow \text{auto-correlation!}$$

by analogy, if we operate on $v(n) = w(n)x(n)$,

$$P_{xx}(\omega) = \frac{1}{UN} \text{DFT}\{r_{vv}(n)\}$$

approach 1 (DFT-based) → Bartlett, Welch
approach 2 (r_{xx} based) → Blackman/Tukey

Q. How does periodogram work, as described above?

A. not very well at all. example:

periodogram_no_average.m

Consider the statistics (P&M 14.1.2; hard to follow)

Average: can show the periodogram is asymptotically unbiased

i.e. if we define $B(\omega) = |W(\omega)|^2$

$$\mathcal{E}(P_{\tilde{x}\tilde{x}}(\omega)) = \frac{1}{N}\Gamma_{xx}(\omega) * B(\omega)$$

for finite N, $\mathcal{E}(P_{\tilde{x}\tilde{x}}) \neq \Gamma_{xx}$, so estimate is biased

as $N \to \infty$, $\frac{1}{N}B(\omega) \to \delta(\omega)$, so $\mathcal{E}(P_{\tilde{x}\tilde{x}}) \to \Gamma_{xx}$

asymptotically unbiased; longer window helps.

short ver:

$\mathcal{E}(P_{\tilde{x}\tilde{x}}(\omega)) \to \Gamma_{xx}$

as $N \to \infty$

Variance: hard to show, but

$$\mathrm{var}(P_{\tilde{x}\tilde{x}}(\omega)) \approx \Gamma_{xx}^{2}(\omega)$$

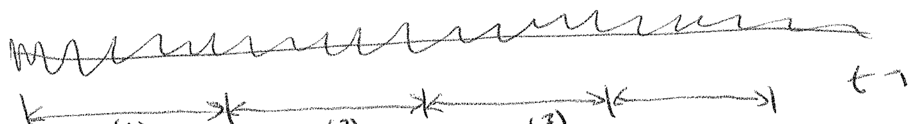
↑ true value

correction: variance ~ true value ^2

longer N does not help!

4) Periodogram averaging (Bartlett method) P+M 14.2.1

This is the simplest method that works:
Break up data into chunks, calculate periodogram for each, and average. Rectangular window



$P^{(1)}_{xx}(\omega)$ $P^{(2)}_{xx}(\omega)$ $P^{(3)}_{xx}(\omega)$

$$P_{xx}(\omega) = \frac{1}{K}\sum_{i=0}^{K-1} P^{(i)}_{xx}(\omega)$$

↑ K windows averaged

stats: since each window is the same, average is unchanged from single periodogram

variance improves:
$$\mathrm{var}(P_{xx}(\omega)) = \frac{1}{K}\Gamma_{xx}^{2}(\omega)$$

so as $K\uparrow$, var $\downarrow$

TRADEOFF
If we have N total points, we now break them into K windows, length $M = N/K$

(bad) resolution drops by factor of K (main lobe bigger)
(good) variance drops " " " K

Bartlett method — Quality factor

$$Q = \frac{E(P_{xx})^2}{var(P_{xx})}$$

for Bartlet, this is $\frac{\cancel{\left(\frac{1}{N}\Gamma_{xx}(\omega) * B(\omega)\right)^2}}{1/K} \approx \frac{\Gamma_{xx}^2}{1/K\,\Gamma_{xx}^2} = K$

if no averaging,
$$Q = \frac{\Gamma_{xx}^2}{\Gamma_{xx}^2} = 1$$

Matlab examples : bartlett_tradeoffs.m

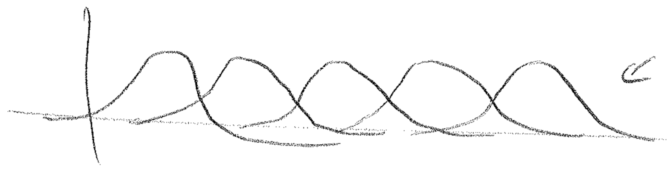
4) welch method (14.2.2 in P+M)

two basic modifications to above:
- a) apply windows other than rectangular
- b) allow windows to overlap

what happens when we overlap the windows?
- → more terms to average — lower variance
- → but overlapped windows are correlated. Thus benefit is not simple $\frac{1}{\#\text{windows}}$

A result: calculation show for 50% overlap, triangle or hanning window, variance reduction is $\sim 2x$
for 75% overlap, reduction $< 4x$

$$Q \text{ factor} = \begin{cases} K, & \text{no overlap (Bartlett} \\ \frac{8K}{9}, & 50\% \text{ overlap, triangle.} \end{cases}$$



← 50% overlapped windows

5) Other methods

→ Blackman Tukey - estimate cross-correlation, smooth it, Fourier transform.

Generally, pretty similar to Welch; no pros or cons.

→ Parametric methods - big topic

~~Best~~ Approaches so far are nonparametric - no assumptions on spectrum shape.

alternative: assume the ~~process~~ spectrum has a particular mathematical form, (all-pole, etc.) & estimate coefficients of the model.

much more high-res, less stable.

homework 14.8
14.9 a)