**Spectral Estimation**

Spectral Analysis - determine, from finite set of samples, power in each harmonic of signals. For non-periodic signals, use Power Spectral Density.

Bank of Band Pass Filters

Spectral Resolution - measure of how close in frequency 2 sinusoids can be before they merge.

Spectral quality - measure of how good spectral estimation is in terms of mean and variance: 

\[ Q = \frac{(E[S_{xx}(f)])^2}{\text{var}[S_{xx}(f)]} \]

It can be improved by increasing the number of samples, M.

Both resolution and Quality can’t be improved at same time for given M. If resolution is improved by reducing BPF bandwith, then BPF impulse response gets longer so less unbiased samples available for estimation.
Spectral Estimators: Parametric or Non-parametric

Non-parametric: require no assumptions about the data except (quasi-) stationarity.

Parametric: based on model of data so their application is more restrictive. Advantage: when applicable, they give more accurate spectral estimate and for fixed M, can reduce bias and variance due to a-priori knowledge in model.

Power Spectral Density:

\[ S_{xx}(\omega) \text{ is defined as DFT of the auto-correlation, } R_{xx}(m) = E[x(k)x(k+m)] \]

\[ S_{xx}(f) = \sum_{m=-\infty}^{\infty} R_{xx}(m)e^{-j\omega mT} - \text{Weiner Khintchine theorem} \quad (1) \]

\( R_{xx} \) can be got from \( S_{xx} \) by finding its Fourier coefficients:

\[ R_{xx}(m) = \frac{T}{2\pi} \int_{0}^{2\pi T} S_{xx}(\omega)e^{j\omega mT}d\omega \]
Classic Spectral Analysis

Eq (1) requires infinite set $R_{xx}(m)$ to calculate $S_{xx}(\omega)$ but only set of M samples available

One approach is to form M-point DFT:

$$X_M(\omega) = \sum_{k=0}^{M-1} x(k)e^{-j\omega kT}$$

An estimate of PSD is then formed by dividing energy at each frequency by no. of data points:

$$S_{xx}(\omega) = \frac{|X_M(\omega)|^2}{M}$$

- Periodogram:
Leakage and Windowing

In practice only \( x(k), \ 0 < k < N \) available so

\[
\tilde{x}(k) = x(k)w(k), \ w(k) = 1 \ for \ 0 < k < N, \ else = 0
\]

\[
\hat{X}(f) = X(f) * W(f) = \int_{-\frac{N}{2}}^{\frac{N}{2}} X(\alpha)W(f - \alpha)d\alpha
\]

Sidelobes of \( W(f) \) cause leakage

Eg: Signal with \( X(f) = 1 \ for \ |f| < 0.1, \ else = 0, \) is windowed by rectangular window of length \( N=61 \)

\[
W(f) = \sum_{k=0}^{N-1} x(k)e^{-j2\pi fk} = \frac{1 - e^{-j\omega N}}{1 - e^{-j\omega}} = e^{-j\omega(n-1)/2} \frac{\sin(\omega N/2)}{\sin(\omega/2)}
\]

Width of main lobe \( \Delta \omega = 4\pi/61 \) or \( \Delta f = 2/61 << widthX(f) \)
Power Spectral Density

For signal $x(t)$, average power is:

$$P_x = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^2(t) dt = \int_{-\infty}^{\infty} S_{xx}(f) df$$

$$S_{xx} = F[R_{xx}(\tau)] = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f \tau} d\tau$$

Properties of P.S.D:
1. $S_{xx}$ is real & +ve
2. Average power in $x(t) = R_{xx}(0) = \int_{-\infty}^{\infty} S_{xx}(f) df$
3. For $x(t)$ real, $S_{xx}(-f) = S(f)$

P.S.D of Random Sequences:

$$S_{xx} = \sum_{-\infty}^{\infty} R_{xx}(k)e^{-j2\pi fk}, \quad -\frac{1}{2} < f < \frac{1}{2} \text{ if } T_s = 1$$

E.g: random binary waveform

$$R_{xx}(\tau) = 1 - |\tau|/T, \quad |\tau| < T, = 0 \text{ elsewhere}$$

$$S_{xx} = F[R_{xx}(\tau)] = T \left[ \frac{\sin \pi f T}{\pi f T} \right]^2$$
M.F. Spectral Estimation

\[ \hat{S}_{xx} = \sum_{k=-\infty}^{N-1} \hat{R}_{xx}(k)e^{-j2\pi fk}, \quad -\frac{1}{2} < f < \frac{1}{2} \text{ for } T_s = 1 \]

where \( \hat{R}_{xx}(k) = \frac{1}{N-k} \sum_{i=0}^{N-k-1} x(i)x(i-k), \quad k = 0, 1, \ldots, N - 1 \)

1. With N data points we can estimate \( R_{xx} \) only for \( |k| < N \).
2. As \( k \to N \), we use fewer points to obtain \( \hat{R}_{xx} \).

So Variance increases as \( k \to N \)

Mean = \( E[\hat{R}_{xx}] = R_{xx}(k), \quad k < N \)

So estimator is unbiased if \( k < N \).

\[ \text{Var}[\hat{R}_{xx}] = \frac{1}{N} \sigma_x^4 \] for Gaussian process whose variance is \( \sigma_x^2 \).

As \( k \to N \), var[\( \hat{R}_{xx} \)] increases.

Estimate \( \hat{S}_{xx} \) has 2 main disadvantages:
1. Large computation needed for \( \hat{R}_{xx} \)
2. Variance[\( \hat{S}_{xx} \)] can be large.
Periodogram

Another estimate for $R_{xx}$ is

$$\hat{R}_{xx}(k) = \frac{1}{N} \sum_{i=0}^{N-k-1} x(i)x(i+k), \ k = 0,1,...N-1$$

$$= \frac{N-k}{N} \hat{R}_{xx}(k)$$

Define $d(n) = 1$, $n = 0...N-1$, $= 0$, else

$$\hat{R}_{xx}(k) = \frac{1}{N} \sum_{n=\infty}^{\infty} [d(n)x(n)] [d(n+k)x(n+k)]$$

$$\hat{S}_{xx}(f) = F[\hat{R}_{xx}(k)] = \sum_{n=\infty}^{\infty} \hat{R}_{xx}(k)e^{-j2\pi kf}$$

$$= \sum_{n=\infty}^{\infty} \left[ \frac{1}{N} \sum_{n=\infty}^{\infty} [d(n)x(n)] [d(n+k)x(n+k)]e^{-j2\pi kf} \right]$$

$$= \frac{1}{N} \left[ \sum_{n=\infty}^{\infty} [d(n)x(n)e^{-j2\pi nf}] [\sum_{n=\infty}^{\infty} [d(n+k)x(n+k)e^{-j2\pi (k+n)f}] \right]$$

$$= \frac{1}{N} X(f)X(f) = \frac{1}{N} |X(f)|^2 \ - \text{Periodogram}$$

36
Bias of Periodogram:

\[
E[\hat{S}_{xx}(f)] = E\left[ \sum_{n=-\infty}^{\infty} \hat{R}_{xx}(k)e^{-j2\pi kf} \right] = \sum_{n=-\infty}^{\infty} E[\hat{R}_{xx}(k)]e^{-j2\pi kf}
\]

\[
= \sum_{n=-\infty}^{\infty} \frac{1}{N} \left( \sum_{n=-\infty}^{\infty} [d(n)x(n)][d(n+k)x(n+k)] \right)e^{-j2\pi kf}
\]

\[
= \sum_{n=-\infty}^{\infty} R_{xx}(k)e^{-j2\pi fk} \left( \sum_{n=-\infty}^{\infty} d(n)d(n+k) \right)/N
\]

\[
= \sum_{n=-\infty}^{\infty} w_N(k)R_{xx}(k)e^{-j2\pi fk}, \quad w_N(k) = \sum_{n=-\infty}^{\infty} d(n)d(n+k)]/N = 1 - |k|/N, \quad |k| < N
\]

\[
= F[w_N(k).R_{xx}(k)]
\]

\[
= \int_{-1/2}^{1/2} S_{xx}(\alpha)W_N(f - \alpha)d\alpha
\]

where \( W_N(f) = F[w_N(k)] = \frac{1}{N} \left[ \frac{\sin(\pi fN)}{\sin(\pi f)} \right]^2 \)

Variance of Periogram:

\[
\text{var}[\hat{S}_{xx}(f)] = S_{xx}^2(f)[1 + \left( \frac{\sin(\pi fN)}{\sin(\pi f)} \right)^2] \xrightarrow{N \rightarrow \infty} S_{xx}^2 \neq 0
\]

Quality Factor of Periodogram:

\[
Q_P = \frac{[E[\hat{S}_{xx}(f)]]^2}{\text{var}[\hat{S}_{xx}(f)]} = \frac{S_{xx}^2}{S_{xx}^2} = 1
\]
Variance of Periodogram (contd)..

For zero mean, white Gaussian sequence with variance $\sigma^2$:

$$\text{var}[\hat{S}_{xx}(f)] = \sigma^4$$

$\Rightarrow$ normalised standard error, $\epsilon = \frac{\sqrt{\text{var}[\hat{S}_{xx}(f)]}}{S_{xx}} = \frac{\sigma}{\sigma} = \frac{\sigma^2}{\sigma}$

i.e $\text{var}[\hat{S}_{xx}(f)]$ does NOT $\to 0$ as $N \to \infty$

**Trade off between Bias and Variance:**

$$\hat{S}_{xx}(f) = F[\hat{R}_{xx}(k)]$$ so examine $\hat{R}_{xx}(k)$ which estimates $R_{xx}(k)$ for $|k| = 0, 1, ..., M$.

When $M << N$ we get good estimates of $R_{xx}(k)$ for $|k| = 0, 1, ..., M$. But bias of $\hat{S}_{xx}(f)$ is larger since $R_{xx}$ is truncated for $k > M$.

As $M \to N$, Bias of $\hat{S}_{xx}(f)$ decreases (less truncation of $\hat{R}_{xx}(k)$) but $\text{var}[\hat{R}_{xx}(k)]$ increases when $k \to N$ as fewer points used in estimator.
Smoothing of Spectral Estimates

1. Bartlett Method - Averaging Periodograms: take $N(>>1)$ samples, $x(0)........x(N-1)$, divide into $n$ sections, each of $N/n$ points, form $n$ different $\hat{S}_{xx}(f)$ and average the $n$ different estimators to form averaged spectral estimators

$$\bar{S}_{xx}(f) = \frac{1}{n} \sum_{k=1}^{n} \hat{S}_{xx}(f)_k$$

If $\hat{S}_{xx}(f)_k$ are independent, then $\text{var}[\bar{S}_{xx}(f)]$ will be reduced by factor $n$. But since fewer points used for $\hat{S}_{xx}(f)_k$, $W_{N/n}(f)$ will be $n$ times wider than $W_N(f)$ so bias will be larger.

$$\text{var}[\bar{S}_{xx}(f)] = \frac{1}{n^2} \sum_{k=1}^{n} \text{var}[\hat{S}_{xx}(f)_k] = \frac{1}{n} \text{var}[\hat{S}_{xx}(f)]$$

$$Q_B = \frac{[E[\bar{S}_{xx}(f)]]^2}{\text{var}[\bar{S}_{xx}(f)]} = n, \quad N \to \infty$$

Freq resolution of Bartlett estimate $\Delta f = .9/M, \quad M = N/n$

$$\Rightarrow Q_B = n = N/M = \frac{N}{.9/\Delta f} = 1.1N\Delta f$$
2. Welch Method: Average Modified Periodograms

Two modifications to Bartlett method:

1. Overlap data segments, \( x_i(k) = x(k + iD) \), \( k = 0, 1, \ldots, M - 1 \); \( i = 0, 1, \ldots, L - 1 \)

2. Data segments windowed prior to computing \( P' \)gram to give modified \( P' \)gram:

\[
\tilde{S}_{xx}(f) = \frac{1}{MU} \left| \sum_{k=0}^{M-1} x_i(k) w(k) e^{-j2\pi kf} \right|^2
\]

Where \( U \) is normalisation factor for power in window, \( U = \frac{1}{M} \sum_{k=0}^{M-1} w^2(k) \)

Welch estimate is average of these periodograms

\[
S^W_{xx}(f) = \frac{1}{L} \sum_{i=0}^{L-1} \tilde{S}_{xx}(f)_i
\]

Mean:

\[
E[S^W_{xx}(f)] = \frac{1}{L} \sum_{i=0}^{L-1} E[\tilde{S}_{xx}(f)_i] = E[\tilde{S}_{xx}(f)_i]
\]

\[
= \int_{-1/2}^{1/2} S_{xx}(\alpha) W(f - \alpha) d\alpha, \quad W(f) = \frac{1}{MU} \left| \sum_{k=0}^{M-1} w(k) e^{-j2\pi kf} \right|^2
\]

As \( N \to \infty \) and \( M \to \infty \), \( E[S^W_{xx}(f)] \to S_{xx}(f) \)

40
Variance of Welch Estimate

\[
var[S_{xx}^W(f)] = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} E[\tilde{S}_{xx}(f)_i \tilde{S}_{xx}(f)_j] - \left\{ E[S_{xx}^W(f)] \right\}^2
\]

For no overlap (\(D=M\), \(L=n\)):

\[
var[S_{xx}^W(f)] = \frac{1}{L} \text{var}[\tilde{S}_{xx}(f)_i] \approx \frac{1}{L} S_{xx}^2(f)
\]

\[Q_W = L = N/M\]

For 50% overlap (\(D=M/2\), \(L=2n\)) with triangular window:

\[
var[S_{xx}^W(f)] = \frac{9}{8L} S_{xx}^2(f)
\]

\[Q_W = \frac{8L}{9} = \frac{16N}{9M}\]

Spectral width of the triangular window at the 3dB points is \(\Delta f = 1.28/M\)

\[Q_W = 0.78N\Delta f\] for no overlap

\[= 1.39N\Delta f\] for 50% overlap + triangular window

41
Blackman Tukey Estimate

Window autocorrelation sequence, take Fourier tfm:

\[
S_{xx}^B(f) = \sum_{m=(M-1)}^{M-1} r_{xx}(m)w(m)e^{-j2\pi mf}
\]

\[
= \int_{-1/2}^{1/2} \hat{S}_{xx}(\alpha)W_M(f-\alpha)d\alpha
\]

Effect of windowing \( r_{xx}(m) \) is to smooth \( S_{xx} \), thus decreasing variance but reducing resolution.

\[
E[S_{xx}^B(f)] = \int_{-1/2}^{1/2} E[\hat{S}_{xx}(\alpha)]W_M(f-\alpha)d\alpha
\]

\[
= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} S_{xx}(\theta)W_B(\alpha-\theta)W_M(f-\alpha)d\theta d\alpha
\]

\[
\approx \int_{-1/2}^{1/2} S_{xx}(\theta)W_M(f-\theta)d\theta \text{ if } M << N
\]

\[
\text{var}[S_{xx}^B(f)] \approx \frac{1}{N} \int_{-1/2}^{1/2} S_{xx}^2(\alpha)W_M^2(f-\alpha)d\alpha
\]

\[
\approx S_{xx}^2(f)\left[\frac{1}{N} \int_{-1/2}^{1/2} W_M^2(f-\theta)d\theta\right]
\]

if \( W_M(f) \) narrow c.f. \( S_{xx}(f) \)
Multiplying by $w(k)$ and taking Fourier tfm:

$$E[\tilde{S}_{xx}(f)] = \int_{-1/2}^{1/2} S_{xx}(\alpha) W_M(f - \alpha) d\alpha, \quad W_M = F[w(k)]$$

Rectangular: $w(k) = 1$ for $|k| \leq M$, $= 0$ for $|k| > M$

$$W_M(f) = \frac{\sin[(2M + 1)\pi f]}{\sin\pi f}$$

Bartlett: $w(k) = 1 - |k|/M$ for $|k| \leq M$, $= 0$ for $|k| > M$

$$W_M(f) = \left[\frac{1}{M}\right] \frac{\sin(M\pi f)}{\sin\pi f}^2$$

Blackman-Tukey: $w(k) = .5(1 + \cos(\pi k/M))$ for $|k| \leq M$, $= 0$ for $|k| > M$

$$W_M(f) = .25[D_M(2\pi f - \pi/M) + D_M(2\pi f + \pi/M)] + D_M(2\pi f)$$
Quality of B-T Estimate

Mean, \( E[S_{xx}^{BT}(f)] \approx \int_{-1/2}^{1/2} S_{xx}(\theta) W_M(f - \theta) d\theta \rightarrow S_{xx}(f) \) as \( N \rightarrow \infty \), ie asymptotically unbiased

\[
\text{var}[S_{xx}^{B}(f)] \approx S_{xx}^2(f)[\frac{1}{N} \sum_{m=(M-1)}^{M-1} w^2(m)]
\]

\( = S_{xx}^2(f) \frac{2M}{N} \) for rect. window
\( = S_{xx}^2(f) \frac{2M}{3N} \) for triang. window

\( \Rightarrow Q_{BT} = 1.5N/M \) for triangular window

As window length is \( 2M - 1 \), \( \Delta f = 1.28/2M = .64/M \)
\( \Rightarrow Q_{BT} = \frac{1.5}{0.64} N \Delta f = 2.34N\Delta f \)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Quality Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett</td>
<td>1.11N\Delta f</td>
</tr>
<tr>
<td>Welch (50% overlap)</td>
<td>1.39N\Delta f</td>
</tr>
<tr>
<td>Blackman–Tukey</td>
<td>2.34N\Delta f</td>
</tr>
</tbody>
</table>

44
Parametric Spectral Estimation

Model-free methods: simple, efficiently computed via FFT,
- but need long data for good frequency response and suffer from spectral leakage.
- Basic Limitation: \( r_{xx} = 0 \) for \( m \geq N \) from

\[
S_{xx}(f) = \sum_{m=-(N-1)}^{N-1} r_{xx}(k)e^{-j2\pi fm}
\]

Model-based methods don’t need this assumption so avoid spectral leakage and give better resolution with less data if model fits well.

Based on modelling data \( x(k) \) as o/p of

\[
H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{q} b_{i}z^{-i}}{1 + \sum_{i=1}^{p} a_{i}z^{-i}} \ldots (1)
\]

\[
x(k) = \sum_{i=0}^{q} b_{i}w(k - i) - \sum_{i=1}^{p} a_{i}x(k - i) \ldots (2)
\]

Then PSD is \( S_{xx}(f) = |H(f)|^2 S_{ww}(f) \). Assume \( w(k) \) is white noise with \( r_{ww}(m) = \sigma_w^2 \delta(m) \). Then

\[
S_{xx}(f) = \sigma_w^2 |H(f)|^2 = \sigma_w^2 \frac{|B(f)|^2}{|A(f)|^2} \ldots (3)
\]

To find \( S_{xx}(f) \):
1. From \( x(k) \), estimate \( \{a_i\} \) and \( \{b_i\} \) of model
2. From these estimates, compute \( S_{xx}(f) \) from (3).
Types of Model

$x(k)$ generated by pole-zero model (1) is called an ARMA process of order $(p,q)$.

If $q = 0$, $H(z) = b_0/A(z)$ and $x(k)$ is an AR process.

If $A(z) = 1$, $H(z) = B(z)$ and $x(k)$ is a MA process. Any ARMA, AR, or MA process can be represented by AR or MA model.

AR Spectral Analysis

\[ S_{xx}(f) = \sigma_e^2 = |H^{-1}(f)|^2 S_{xx}(f), \quad H^{-1} = 1 - \sum_{i=1}^{p} a_i z^{-i} |_{z=e^{-j2\pi fT}} \]

The whitening filter $H^{-1}(z)$ is a linear predictor:
\[ a_i \text{ are chosen to minimise the MSE, } E[e^2(k)] \text{ which lead to ACN Equations:} \]
\[ \Phi_{xx} a = \phi_{xx} \text{ where} \]
\[ \Phi_{xx} = E[x(k-1)x^T(k-1), \phi_{xx} = E[x(k-1)x^T(k-1)] \]
\[ \text{Min MSE, } \sigma_e^2 = E[x^2(k)] - a^T \phi_{xx} \]

To perform AR analysis directly on the data, use sum of squared error cost fn:
\[ \rho = \sum_k [x(k) - a^T x(k-1)]^2 \]

Set of \( a \) which minimise \( \rho \) is provided by the deterministic form of ACN eqn:
\[ R_x a = r_x \]

Choosing the limits for \( k \) makes assumptions about \( x(k) \) before \( x(0) \) and after \( x(N-1) \), which effects the accuracy of \( a \). Eg assume \( k \) starts at 0. Then \( e(0) = x(0) - a^T x(-1) = x(0) \) which gives no information about \( a \).

Yet \( e(0) \) is weighted equally with errors from the middle of \( x(k) \).

Similarly \( e(1) = x(1) - a^T x(0) = x(1) - a_1 x(0) \) which gives information only about \( a_1 \).

It is only when \( e(p) \) is considered that no assumptions about the data have to be made.

47
AR Windows

Errors at the end of the data cause similar effect for possible windowing variations which lead to four possible estimates of $a$:

<table>
<thead>
<tr>
<th>Lower Limit $\sum_{k=0}^{N-p}$</th>
<th>Upper Limit $\sum_{k=p}^{N-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>autocorrelation</td>
<td>$\rho$ pre-windowed covariance</td>
</tr>
<tr>
<td>post-windowed covariance</td>
<td></td>
</tr>
</tbody>
</table>

Despite its inaccuracy, auto-correlation form is popular in real-time systems as $R_{xx}$ is Toeplitz. The minimum sq error $\rho$ is used to get an estimate, $\hat{\sigma}_e^2$ of the MSE by dividing $\rho$ by the no. of errors.

For covariance form:

$$\hat{\sigma}_e^2 = \frac{\sum_{k=p}^{N-1} x^2(k) - r_x^T a}{N - 1 - p}$$

The AR(p) spectral estimate is

$$S_{xx}(f) = \frac{\hat{\sigma}_e^2}{|1 + \sum_{i=1}^{p} a_i e^{j\omega T}|^2}$$

19
AR Model Order

If \( p \) is too low, spectrum is very smoothed and can’t separate close peaks
If \( p \) is too high, spectrum may have spurious peaks and is inefficient

\( \sigma^2_e \) decreases as the order of AR model increases.

Rate of decrease can be monitored and \( p \) chosen when rate of decrease becomes slow but this is imprecise.

Other methods for selecting \( p \):

1. Final Prediction Error: order is selected to minimise

\[
FPE(p) = \hat{\sigma}_e^2 (N + p + 1)/(N - p - 1)
\]

2. Akaike Information Criterion: order is selected to minimise

\[
AIC(p) = \ln \hat{\sigma}_e^2 + 2p/N
\]

3. Criterion Autoregressive Transfer: order is selected to minimise

\[
CAT(p) = (1/N) \sum_{j=1}^{p} ((N - j)/\hat{\sigma}_e^2) - ((N - p)/\hat{\sigma}_e^2)
\]
MA Model

\[
S_{xx}(z) = \sigma_w^2 H(z)H(z^{-1}) = \sigma_w^2 \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})} = \sigma_w^2 \frac{D(z)}{C(z)}
\]

\[
= \sigma_w^2 \sum_{m=-q}^{q} \frac{d_m z^{-m}}{\sum_{m=-p}^{p} c_m z^{-m}}
\]

where \(d_m = \sum_{k=0}^{q-|m|} b_k b_{k+m}, \ |m| \leq q\)

When \(a_1 = a_2 = ....a_p = 0\) it can be shown that

\[
r_{xx}(m) = \begin{cases} \sigma_w^2 d_m & |m| \leq q \\ 0 & |m| > q \end{cases}
\]

and PSD for MA(q) process is:

\[
S_{xx}^{MA}(f) = \sum_{m=-q}^{q} r_{xx}(m)e^{-2j\pi fm}
\]

= BT modified PERIODGRAM

i.e we dont need to solve for \(b_i\) to get \(S_{xx}^{MA}(f)\) which is identical to non-parametric PSD estimate.

Since \(r_{xx}(m) = 0\) for \(|m| > q\), the order of MA model may be determined by noting if \(r_{xx}(m) \approx 0\) for large lags. If this is not the case, MA model will give poor frequency resolution.
ARMA Model

Mpy Eq (2) by $x^*(k - m)$ and take expected values:

$$E[x(k)x^*(k-m)] = \sum_{n=0}^{q} b_n E[w(k-n)x^*(k-m)] - \sum_{n=1}^{p} a_n E[x(k-n)x^*(k-m)]$$

$$r_{xx}(m) = -\sum_{n=1}^{p} a_n r_{xx}(m-n) + \sum_{n=0}^{q} b_n r_{wx}(m-n)$$

$$r_{wx} = \left\{ \begin{array}{ll}
\sigma^2_w h(-m) & |m| \leq 0 \\
0 & |m| > 0
\end{array} \right.$$

$$\Rightarrow r_{xx}(m) = \left\{ \begin{array}{ll}
-\sum_{n=1}^{p} a_n r_{xx}(m-n) & m > q \\
-\sum_{n=1}^{p} a_n r_{xx}(m-n) + \sigma^2_w \sum_{n=0}^{q-m} b_{n+m} h(m) & 0 \leq m \leq q \\
r_{xx}^*(-m) & |m| > 0
\end{array} \right.$$

Coeffs $a_n$ for $m > q$ found as in AR case.

Then $A(z) = 1 + \sum_{n=1}^{p} a_n z^{-n}$, which can filter $x(k)$ to give $v(k) = x(k) + \sum_{n=1}^{p} a_n x(k-n)$

Then apply MA method to $v(k)$ to obtain:

$$S_{vv}^{MA}(f) = \sum_{m=-q}^{q} r_{vv}(m)e^{-2\pi fm}$$

$$\Rightarrow S_{xx}^{ARMA}(f) = \frac{S_{vv}^{MA}(f)}{|1 + \sum_{n=1}^{p} a_n e^{-2\pi fm}|^2}$$
Maximum Likelihood Spectral Estimation

Model Spectrum Analyser as tunable filter. To find power at frequency, ω, tune filter to ω and measure o/p power.

Filter must meet 2 criteria:
1. If signal consists of nothing but sinusoid at freq, ω, o/p power = i/p power.
2. Any signal at other freqs must make the min possible contribution to o/p power (Min Variance).

Filter o/p, \( y(n) = \sum_{k=0}^{N-1} a_k x(n - k) \) and assume i/p consists of complex sinusoid of freq, \( \omega \), and other components:

\[
x(n) = Ae^{j\omega n} + p(n)
\]

MV design criteria may be written as:
1. If \( p(n) = 0 \), then \( y(n) = x(n) \)
2. Otherwise, o/p power of the filter is to be a minimum.
Consider 1. It means that

\[ \sum_{k=0}^{N-1} a_k A e^{j\omega(n-k)} = A e^{j\omega n} \]

Divide by \( Ae^{j\omega n} \):

\[ \sum_{k=0}^{N-1} a_k e^{j\omega k} = 1 \] ...(3)

Use Vectors:

\[ a = [a_0, a_1, \ldots, a_{N-1}]^T \]

\[ e = [1, e^{j\omega}, e^{2j\omega}, \ldots, e^{j(N-1)\omega}]^T \]

Then (3) becomes \( e^*a = 1 \)

Consider 2.

Variance, \( \sigma^2 = E[y^2(n)] \)

\[ = E\left[ \sum_{i=0}^{N-1} a_i^* x(n-i) \sum_{k=0}^{N-1} a_k x(n-k) \right] \]

Interchanging order of summations and expected values:

\[ \sigma^2 = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_i^* a_k E[x(n-i)x(n-k)] \]

\[ = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_i^* a_k \delta_{i-k} = a^* Ra \]

54b
Need to min $\sigma^2 = a^* R a$ subject to $e^* a = 1$

Using Lagrange’s method

$$F(a) = a^* R a - \lambda e^* a$$

$$\frac{dF}{da} = [2a^* R - \lambda e^*] = 0 \Rightarrow a = .5 \lambda R^{-1} e$$

Use constraint to find $\lambda$:

$$e^* a = .5 \lambda e^* R^{-1} e = 1 \Rightarrow \lambda = 2/(e^* R^{-1} e)$$

$$\Rightarrow a = \frac{R^{-1} e}{e^* R^{-1} e} = \frac{R^{-1} e}{e^* R^{-1} e}$$

For these $a$, $\sigma^2$ is estimate of spectral power at $\omega$

$$\sigma^2 = a^* R a = \frac{e^* R^{-1} R R^{-1} e}{(e^* R^{-1} e)^2} = \frac{1}{e^* R^{-1} e} = P_{ML}(\omega)$$

$e^* R^{-1} e$ evaluated by multiplying elements along any diagonal of $R^{-1}$ by the same $e^{-j\omega}e^{j\omega} = e^{j(k-i)\omega}$. If $n = k-i$, then $e^* R^{-1} e = \sum_{n=-\infty}^{N-1} e^{j\omega} \rho_n$, $\rho_n = \text{sum of all elements of } R^{-1} \text{ along diagonal } n$.

This can be computed with DFT, so we don’t have to evaluate $\frac{1}{e^* R^{-1} e}$ for every freq of interest.

54c
Maximum Entropy Spectral Est

Traditional methods not suited to finite data

Windowing assumes data outside window is 0 - "an affront to common sense" [Burg]

Burg’s ME method starts with known data and gets everything needed from that.

Entropy is measure of uncertainty. Since we want to leave signal free outside window, we maximise entropy of power spectrum.

Start with expression of entropy of spectrum and find spectrum that maximises this, subject to condition that spectrum be consistent with known autocorrelations.
**Entropy:** Let \( x \) be a discrete random variable and \( p(x_i) \) be the probability that \( x \) takes value \( x_i \).

Entropy of \( x \) is \( H(x) = - \sum x_i p(x_i) \log p(x_i) \)

E.g. if \( x \) can take only 2 values \( x_0 \) and \( x_1 \) then:

\[
H(x) = -p \log p - (1-p) \log(1-p)
\]

where \( p = p(x_0) \)

\[\text{H(x)}\]

\[\begin{array}{cc}
0 & 1 \\
H(x) & \\
0.5 & 0.5 \\
ap & 1.0 \\
\end{array}\]

\( H(x) \) has the following properties:

1. \( H(x) \) is a continuous function of \( p(x_i) \)
2. \( H(x) \) is non-negative and is 0 only for cases where \( p(x_i) = 0 \) for all the values of \( x_i \) but one.
3. \( H(x) \) is max when all values equally likely
4. \( H(x) \) increases with \( N \), if \( N \) can take on \( N \) different values, all equally likely
5. \( H(x) \) is additive in following sense: suppose we partition possible values of \( x \) into 2 sets \( A \) and \( B \). Then finding the value of \( x \) can be divided into 2 steps: determine whether \( x \) is in \( A \) or \( B \); determine actual value of \( x \) given that it is known to be in \( A \) (or \( B \)). The entropies of these 2 steps will sum to \( H(x) \)
If x is cont random variable with pdf = f(x), then entropy is

$$H(x) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx \ldots (2)$$

Base of log gives unit of entropy.

For discrete r.v., base = 2 and unit is bit

For cont. r.v., use natural logs and unit is nat
- for c.r.v with std dev $\sigma$, pdf with max entropy is Gaussian

For zero-mean Gaussian pdf with std dev $\sigma$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x^2/2\sigma^2)}$$

From (2): $H(x) = \ln(\sqrt{2\pi e\sigma})$ nats

For n-dimensional, 0-mean, multivariate Gaussian pdf with covariance matrix $C$:

$$f(x) = \left[(2\pi)^{n/2} |C|\right]^{-1/2} e^{-\frac{1}{2}(x^T C^{-1} x)}$$

$$H'(x) = \ln(\sqrt{(2\pi e)^n |C|}) \text{nats} \ldots (3)$$

Avoid problem of non-convergence of $H(x)$ as signal length increases by using entropy rate:

$$H'(x) = \lim_{n \to \infty} \frac{1}{n} \ln(\sqrt{(2\pi e)^n |C|}) = \ln(\sqrt{2\pi e}) + \frac{1}{2} \lim_{n \to \infty} \ln(|C|)/n$$

If signal is stationary, $C$ is Toeplitz and:

$$\lim_{n \to \infty} \ln(|C|)/n = \frac{1}{2W} \int_{-W}^{W} \ln S(f) df$$

$$\Rightarrow H'(x) = \ln(\sqrt{2\pi e}) + \frac{1}{4W} \int_{-W}^{W} \ln S(f) df \ldots (4)$$
Application of ME to Spectral Analysis

Assume we know $r_n$ out to max lag $p$, none beyond $p$. If we knew all $r_n$, we could find PSD from:

$$S(f) = \frac{1}{2W} \sum_{n=-\infty}^{\infty} r_ne^{-j2\pi fnT}, \quad T = 1/(2W)$$

But few known $r_n$ are not enough; there are an infinite number of spectra, depending on unknown $r_n$.

**ME Method:** Find $S(f)$ which maximises $H'(x)$ subject to

$$r_n = \frac{1}{2W} \int_{-W}^{W} S(f)e^{j2\pi fnT} df, \quad |n| \leq p \quad (5)$$

So (4) becomes:

$$H'(x) = \ln(\sqrt{2\pi e}) + \frac{1}{4W} \int_{-W}^{W} \ln[ \frac{1}{2W} \sum_{n=-\infty}^{\infty} r_ne^{-j2\pi fnT}] df$$

$$\frac{\partial H'(x)}{\partial r_n} = \frac{1}{4W} \int_{-W}^{W} \frac{1}{2W} \sum_{n=-\infty}^{\infty} r_ne^{-j2\pi fnT}]^{-1}e^{-j2\pi fnT} df$$

$$= \frac{1}{4W} \int_{-W}^{W} \frac{1}{S(f)}e^{-j2\pi fnT} df$$

Constraint (5) means derivatives = 0 only w.r.t. $r_n$, $n > p$

But these derivatives are Fourier coeffs of $1/S(f)$. So, ME criterion means that Fourier expansion of $1/S(f)$ must have finite no. of terms. So $S(f)$ must be of form:

$$S(f) = [ \sum_{n=-p}^{p} q_ne^{-j2\pi fnT}]^{-1} \quad (6)$$
To find coefficients \( \{q_n\} \): Write (6) as z-tfm:

\[ S(f) = S_D(z)|_{z=e^{j2\pi fT}} , \text{ where } S_D(z) = 1/Q(z) \]

\[ S_D(z) = P/[A(z)A^*(z^{-1})] , \text{ where } P = 1/q_0 \text{ and } A(z) = \sum_{i=0}^{p} a_i z^{-i} \]

If \( R(z) \) is z-tfm of \( r_n \), then (5) becomes:

\[ R(z) = P[A(z)A^*(z^{-1})] \]

\[ R(z)A(z) = P/A^*(z^{-1}) \ldots (7) \]

As \( P/A(z) \) is z-tfm of impulse response \( h(n) \) of casual system, \( P/A^*(z^{-1}) \) is z-tfm of \( h^*(-n) \). So:

\[ r_n * a_n = h^*(-n) \Rightarrow \sum_{i=-p}^{p} a_i r_{n-i} = h^*(-n) \ldots (8) \]

Equation (8) to be satisfied for \( n=0 \) to \( p \) (to match \( r_n \) out to \( r_p \) ) But if \( h(n) \) is causal, \( h^*(-n) = 0 \) for +ve \( n \). Hence

\[ \sum_{i=-p}^{p} a_i r_{n-i} = \begin{cases} h^*(0) , & n = 0 \\ 0 , & 1 \leq n \leq p \end{cases} \ldots (9) \]

These equations are similar to ACNE for linear predictor. Once we get \( \{a\} \) we can get power spectrum: use

\[ S(f) = \frac{P}{|A(z)|^2} \bigg|_{z=e^{j2\pi fT}} \ldots \ldots (9a) \]

and obtain \( S_{ME}(f) \) from the reciprocal of the absolute-value-squared Fourier tfm of \( \{a_0, a_1, ..., a_p, 0, ..., 0\} \)
Linear Prediction and Burg’s Method

Similarity of equation (9) to ACNE for Linear predictor can be used to develop Burg’s Method. Recursive solution:

\[ a_0(0) = 1 \]

\[ e_0 = r_0 \]

\[ a_{n+1}(j) = a_n(j) + K_{n+1}a_n(n + 1 - j) \quad ... \quad (10)a \]

\[ E_{n+1} = E_n(1 - K_{n+1}^2) \quad ... \quad (10)b \]

where

\[ K_{n+1} = -\frac{1}{E_n} \sum_{i=0}^{n} a_n(i)r_{n+1-i} \quad ... \quad (11) \]

So far we have assumed we know \( r_n \) to find \( \{a\} \)
But real \( r_n \) may not be known; how to find \( \{k\} \) and \( \{a\} \)?

**Burg’s Method:** Suppose we need a way of minimizing \( E_n \) that
1. is consistent with recursion (10)
2. gives us \( \{k\} \) directly,
3. does not rely on \( r_n \)

There is such a way based on recurrence relations among individual forward and backward prediction errors for order-p linear predictor:

\[ f_p(n) = f_{p-1}(n) + K_pb_{p-1}(n) \quad ... \quad (12a) \]

\[ b_p(n) = b_{p-1}(n - 1) + K_pf_{p-1}(n - 1) \quad ... \quad (12b) \]
On each recursion step $n$, use Eq (12) to find value of $K_n$ that minimises sum of mean-squared forward and backward prediction errors $E$ over sequence of length $N$:

$$E = \frac{1}{2(N-m)} \sum_{j=n}^{N-1} f_n^2(j) + b_n^2(j + 1)$$

$$= \frac{1}{2(N-m)} \sum_{j=n}^{N-1} [f_{n-1}(j) + K_n b_{n-1}(j)]^2 + [b_{n-1}(j) + K_n f_{n-1}(j)]^2$$

$$\frac{\partial E}{\partial K_n} = \frac{1}{(N-m)} \sum_{j=n}^{N-1} K_n [f_{n-1}^2(j) + b_{n-1}^2(j)] + 2 f_{n-1}(j)b_{n-1}(j)$$

Setting this derivative to 0 gives:

$$K_n = -2P/Q \quad \text{(13)}$$

where

$$P = \sum_{j=n}^{N-1} f_{n-1}(j)b_{n-1}(j)$$

$$Q = \sum_{j=n}^{N-1} f_{n-1}^2(j) + b_{n-1}^2(j)$$

(13) is expression for $K_n$, using errors known from previous step and doesn't need $r$. Having found new $K_n$, we can find new $\{a\}$ from (10) and new prediction errors from (12).

To start the process, we observe that for $n = 0$, predicted values are all 0, so errors are simply the values of the sequence. Burg recursion then proceeds as follows:

$$61$$
Burg’s Recursion:

1. For $n = 0$:
   \[ f_0(j) = y(j), \quad b_0(j) = y(j - 1), \quad j = 1 \text{ to } N, \quad a_0(0) = 1 \]

2. For $n = 1$ to $p$:
   
   (a) $K_n = -2P/Q$ from (13)
   
   (b) $a_n(n) = K_n$ and $a_n(0) = 1$
   
   (c) From (10a), for $i = 1$ to $n - 1$
   \[ a_n(i) = a_{n-1}(i) + K_n a_{n-1}(n - i) \]

   (d) From (12), for $j = n$ to $n - 1$
   \[ f_n(j) = f_{n-1}(j) + K_n b_{n-1}(j) \]
   \[ b_n(j) = b_{n-1}(j - 1) + K_n f_{n-1}(j - 1) \]

3. Using \{a\}, obtain $S(f)$ from (9a)

   Step 3 is carried out by finding the reciprocal of the squared magnitude of the DFT of \{a_0, a_1, ..., a_p, 0, ..., 0\}
The Burg method doesn’t need the calculation of \( \{r\} \) but if they are required, they can be got as a by-product of the Burg recursion. To see how this can be done, refer to Eq (11) rewritten as:

\[
K_n = -\frac{1}{E_{n-1}} \sum_{i=0}^{n-1} a_{n-1}(i)r_n - i
\]

\[
\Rightarrow r_n = -E_{n-1}K_n - \sum_{i=1}^{n-1} a_{n-1}(i)r_n - i
\]

\[
= -K_n \sum_{i=1}^{n} a_{n-1}(n - i)r_n - i - \sum_{i=1}^{n-1} a_{n-1}(i)r_n - i
\]

\[
= \sum_{i=1}^{n} a_n(i)r_{n-i}
\]

So we can include computation of \( r \) into Burg recursion by adding to step 1:

\[
r_0 = \sum_{j=1}^{N} y^2(j)
\]

and to step 2:

\[
r_n = \sum_{i=1}^{n} a_n(i)r_{n-i}
\]
Performance of ME Spectral Estimate

E.g for 3 sinusoids at 764, 955 and 1019 Hz
Figure 2.17 Computer simulation results for test case data: (a) True power spectral density; (b) percpaper; (c) Blackman-Tukey; (d) MVDE; (e) autocorrelation; (f) Burg; (g) covariance; (h) modified covariance; (i) Dolph; (j) MTWE; (k) LSMYWE.