4F7 Dr S Godsill sdg@eng.cam.ac.uk

ENGINEERING TRIPOS PART IIB

Saturday ?? May 2006 9 to 10.30

Module 4F7 WORKED SOLUTIONS

DIGITAL FILTERS AND SPECTRUM ESTIMATION

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** number of marks allocated to each part of a question is indicated in the right margin.*

Answers to questions in each section should be tied together and handed in separately.

There are no attachments.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
Engineering Data Book
CUED approved calculator allowed

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

Version 1- 8 Feb 2006

1 In the standard adaptive filtering problem we have an input signal $U = \{u(n)\}_{n=0}^{\infty}$, a reference signal $D = \{d(n)\}_{n=0}^{\infty}$, and a Finite Impulse Response filter (FIR) filter $H = \{h(m)\}_{m=0}^{M}$. This filter is applied to the input signal to produce an output signal $Y = \{y(n)\}_{n=0}^{\infty}$, i.e.

$$y(n) = [Hu](n) = \sum_{m=0}^{M} h(m)u(n-m).$$

The aim is to design H such that the output signal Y is as "close" as possible to D.

(a) Describe how the general adaptive filtering scenario above may be formulated as a Wiener filtering problem. As part of this description you should define the error criterion used and explain (without detailed calculations) the steps involved in deriving a solution. [30%]

Solution:

Error criterion is the expected value of the mean-squared error between the reference signal D and the filtered estimate Y:

$$J = E\{|d(n) - y(n)|^2\}$$

In the Wiener filter this expected squared error is minimised with respect to the filter coefficients H to give an optimal solution. This solution can be obtained by differentiating J with respect to H, setting to zero and solving for the coefficients.

(b) For each of the following three cases, describe how the problem may be formulated as an adaptive filtering problem, defining (with the aid of diagrams) appropriate reference signals, input signals and error signals. Detail any assumptions or approximations involved.

(i) Let $X = \{x(n)\}_{n=0}^{\infty}$ be a sequence of independent and identically distributed random symbols such that

$$\Pr\{x(n) = 1\} = \Pr\{x(n) = -1\} = 0.5.$$

These symbols are transmitted through a communication channel H_{channel} which distorts the transmitted symbols to produce

$$z(n) = [H_{\text{channel}}x](n).$$

The aim is to design a FIR filter to recover the transmitted symbols as accurately as possible. [30%]

Solution:

•To solve this problem, we will need a set of training symbols. Call this the reference signal $\{d(n)\}$

•Call the signal coming out of the communication channel the input signal $\{u(n)\}$

•We will filter $\{u(n)\}$ to yield $\underline{h}^{\mathrm{T}}\underline{u}(n)$ where $\underline{u}(n) = (u(n), u(n-1), \ldots, u(n-M+1))^{\mathrm{T}}$

•Let the error signal be $e(n) = d(n) - \underline{h}^{\mathrm{T}}\underline{u}(n)$

•The minimization of $E\left\{e^2(n)\right\}$ with respect to $\underline{h}$ is the Wiener filtering problem. Call this solution $\underline{h}_{\mathrm{opt}}$.

•Once the training period is over, we will recover the transmitted symbols using $\underline{h}_{\mathrm{opt}}^{\mathrm{T}}\underline{u}(n)$

(ii) In an echo cancellation problem for telephony, the local (near) user has a hands-free unit comprising a microphone and loudspeaker. The voice of the far speaker, coming out of loudspeaker, is reflected by the room back to the microphone (the echo), and from there transmitted back to the far speaker. The echo is annoying to the far speaker and the aim is thus to cancel the echo. [20%]

Solution:

•Define

– $u(n)$ to be signal out of loudspeaker (speech signal of the far speaker)

– $\varepsilon(n)$ the signal of near speaker

– $H_{\mathrm{room}}[u](n)$ the echo

•Call the signal into the mic $d(n)$,

$$d(n) = \underbrace{H_{\mathrm{room}}[u](n)}_{\text{echo}} + \underbrace{\varepsilon(n)}_{\text{near speaker speech}}$$

•We will filter $\{u(n)\}$ to yield $\underline{h}^{\mathrm{T}}\underline{u}(n)$ where $\underline{u}(n) = (u(n), u(n-1), \ldots, u(n-M+1))^{\mathrm{T}}$

•Let the error signal be $e(n) = d(n) - \underline{h}^{\mathrm{T}}\underline{u}(n)$

•The minimization of $E\left\{e^2(n)\right\}$ with respect to $\underline{h}$ is the Wiener filtering problem

 (TURN OVER for continuation of Question 1

•If we have completely cancelled the echo then

$$\begin{aligned} e(n) &= \varepsilon(n) \\ &= \text{signal from microphone minus echoes} \\ &= \text{speech signal of interest.} \end{aligned}$$

(iii) A recording is made of a bird singing in background noise. A second recording is simultaneously made at a nearby location. It is assumed that the second recording is statistically correlated with the noise in the first recording. The aim is to remove the noise from the first recording which contains both noise and birdsong. [20%]

Solution:

•Define the following reference signal:

$$d(n) = \underbrace{s(n)}_{\substack{\text{signal of interest} \\ \text{(bird)}}} + \underbrace{v(n)}_{\text{noise}}$$

$s(n)$ and $v(n)$ statistically independent

•Call the recording of background noise alone the input signal $u(n)$

•Obviously $u(n) \neq v(n)$ but $u(n)$ and $v(n)$ are correlated

•Now **filter** recorded noise $u(n)$ to make it more like $v(n)$

•Recovered signal is $d(n) - \underline{h}^{\mathrm{T}}\underline{u}(n)$ where $\underline{u}(n) = (u(n), u(n-1), \ldots, u(n-M+1))^{\mathrm{T}}$

•Optimise the filter $\underline{h}$ by solving $\min_{\underline{h}} E\left\{\left(d(n) - \underline{h}^{\mathrm{T}}\underline{u}(n)\right)^2\right\}$

•The solution to this problem is the Wiener filter

2 Recall that the *inner product* of two vectors $\mathbf{x} = [x_1, \ldots, x_m]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, \ldots, y_m]^{\mathrm{T}}$ is defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{m} x_i y_i$ and the *norm* as $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

(a) Consider two random variables x and y. It is desired to find a linear estimator of x given y, i.e. $\widehat{x} = ay$ where scalar a is to be determined.

By defining an appropriate inner product and norm for random variables, express the Wiener solution to this problem in vector space form (i.e. in terms of the inner products and norms you have defined). State also the orthogonality property that is satisfied by this solution. [30%]

Solution:

Define the inner product as

$$\langle x, y \rangle = E\{xy\},$$

and the norm as

$$\|x\|^2 = \langle x, x \rangle = E\{xx\}.$$

The Wiener solution is found by minimising

$$\|x - ay\|^2$$

w.r.t. a. This is the Wiener criterion for the estimator. Expanding the norm using the inner product and upon differentiating yields the solution

$$a = \frac{\langle x, y \rangle}{\langle y, y \rangle}$$

So, $\widehat{x}$ is the projection of x onto y

The residue of the projection $x - \widehat{x}$ is orthogonal to y:

$$\langle x - \widehat{x}, y \rangle = \langle x, y \rangle - \langle \widehat{x}, y \rangle = 0.$$

(b) We now have a collection of random variables $\{y_1, \ldots, y_t\}$ and it is desired to find a linear estimator of x of the form $\widehat{x} = \sum_{i=1}^{t} a_i y_i$. If t is increased by 1, we would like to update the solution for $\{y_1, \ldots, y_t\}$ to $\{y_1, \ldots, y_t, y_{t+1}\}$ using a recursive update.

Describe the Gram-Schmidt procedure for solving this problem. [40%]

Solution:

- We know that the solution should be to project x onto the space spanned by $\{y_1, \ldots, y_t\}$

•We will use this interpretation to derive the recursive implementation

•We must first make $\{y_1,\ldots,y_t\}$ orthogonal. Set $\varepsilon_1 = y_1$. For $j > 1$,

$$\varepsilon_j = y_j - \sum_{i=1}^{j-1} \frac{E\left(y_j \varepsilon_i\right)}{E\left(\varepsilon_i^2\right)} \varepsilon_i$$

•Now project x onto $\{\varepsilon_1,\ldots,\varepsilon_t\}$,

$$\widehat{x} = \sum_{i=1}^{t} \frac{E\left(x \varepsilon_i\right)}{E\left(\varepsilon_i^2\right)} \varepsilon_i.$$

•Main point: to define ε_{t+1} we don't have to change $\{\varepsilon_1,\ldots,\varepsilon_t\}$, which gives the recursive implementation

(c) The standard formulation for the solution to the problem in part (b) is to find $\mathbf{a}_t$ by minimising

$$E\left\{\left(x - \sum_{i=1}^{t} a_i y_i\right)^2\right\} = E\left\{\left(x - \mathbf{a}_t^{\mathrm{T}} \mathbf{y}_t\right)^2\right\}$$

where vector $\mathbf{a}_t^{\mathrm{T}} = [a_1,\ldots,a_t]$ and $\mathbf{y}_t^{\mathrm{T}} = [y_1,\ldots,y_t]$. Derive an explicit solution to this problem. [30%]

This is a Wiener filtering problem and the solution is

$$\mathbf{a}_t = \left(E\left\{\mathbf{y}_t \mathbf{y}_t^{\mathrm{T}}\right\}\right)^{-1} E\left\{x \mathbf{y}_t\right\}$$

3 (a) Describe the parametric model-based approach to spectrum estimation, and compare its properties with non-parametric methods such as the periodogram. [30%]

Solution [More detailed than required:]

- •Periodogram-based methods can lead to biased estimators with large variance
- •If the physical process which generated the data is known or can be well approximated, then a parametric model can be constructed
- •Careful estimation of the parameters in the model can lead to power spectrum estimates with improved bias/variance.
- •Consider spectrum estimation for LTI systems driven by a white noise input sequence.
- •If a random process $\{X_n\}$ can be modelled as white noise exciting a filter with frequency response $H(e^{j\omega T})$ then the spectral density of the data can be expressed as:

$$\boxed{S_X(e^{j\omega T}) = \sigma_w^2 |H(e^{j\omega T})|^2}$$

where σ_w^2 is the variance of the white noise process. [It is usually assumed that $\sigma_w^2 = 1$ and the scaling is incorporated as gain in the frequency response]

- •We will study models in which the frequency response $H(e^{j\omega T})$ can be represented by a finite number of parameters which are estimated from the data.
- •Parametric models need to be chosen carefully - an inappropriate model for the data can give misleading results

ARMA Models A quite general representation is the autoregressive moving-average (ARMA) model:

- •The ARMA(P,Q) model difference equation representation is:

$$\boxed{x_n = -\sum_{p=1}^{P} a_p x_{n-p} + \sum_{q=0}^{Q} b_q w_{n-q}} \quad (1)$$

where:

a_p are the AR parameters,

b_q are the MA parameters

and $\{W_n\}$ is a zero-mean stationary white noise process with unit variance, $\sigma_w^2 = 1$.

•Clearly the ARMA model is a pole-zero IIR filter-based model with transfer function

$$H(z) = \frac{B(z)}{A(z)}$$

where:

$$A(z) = 1 + \sum_{p=1}^{P} a_p z^{-p}, \quad B(z) = \sum_{q=0}^{Q} b_q z^{-q}$$

•Unless otherwise stated we will always assume that the filter is stable, i.e. the poles (solutions of $A(z) = 0$) all lie *within* the unit circle (we say in this case that $A(z)$ is *minimum phase*). Otherwise the autocorrelation function is undefined and the process is technically *non-stationary*.

•Hence the power spectrum of the ARMA process is:

$$S_X(e^{j\omega T}) = \frac{|B(e^{j\omega T})|^2}{|A(e^{j\omega T})|^2}$$

The ARMA model is quite a flexible and general way to model a stationary random process:

•The poles model well the *peaks* in the spectrum (sharper peaks implies poles closer to the unit circle)

•The zeros model troughs in the spectrum

•Complex spectra can be approximated well by large model orders P and Q

(b) Suppose that a single frequency component having normalised frequency Ω and amplitude a is buried in independent, uncorrelated, complex Gaussian noise with variance σ_e^2:

$$x_n = a\exp(jn\Omega) + e_n = a\,g_n(\Omega) + e_n$$

We observe a sequence of N data points from this process, $\mathbf{x} = \{x_0, x_1, ..., x_{N-1}\}$. Show that the maximum likelihood solution for amplitude and frequency values can be obtained by minimising the total squared error:

$$J(a, \Omega) = \sum_{n=0}^{N-1} |x_n - a\exp(jn\Omega)|^2$$

[15%]

Solution:

As for the AR model, think of a change of variables from e_n to x_n:

$$x_n = ag_n(\omega) + e_n$$

with $ag_n(\omega)$ considered as a constant. This leads to:

$$\begin{aligned} p(x_n|a,\omega) &= \left.\frac{p(e_n)}{|J|}\right|_{e_n = x_n - ag_n(\omega)} \\ &= \frac{1}{2\pi\sigma_e^2} \exp\left(-\frac{1}{2\sigma_e^2}|e_n^2|\right) \end{aligned}$$

[$|J| = 1$ as for the AR model since $\frac{\partial x_n}{\partial e_n} = 1$.]

Now consider a vector of N observed data points (measurements):

$$\mathbf{x} = [x_0, x_2, ..., x_{N-1}]$$

and the corresponding vector of frequency terms:

$$\mathbf{g}(\omega) = [g_0(\omega), g_2(\omega), ..., g_{N-1}(\omega)]$$

Since the e_n are independent:

$$\begin{aligned} p(\mathbf{x}|a,\omega) &= \prod_{n=1}^{N} p(x_n|a,\omega) \\ &= \frac{1}{(2\pi\sigma_e^2)^N} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{n=0}^{N-1}|x_n - ag_n(\omega)|^2\right) \\ &= \frac{1}{(2\pi\sigma_e^2)^N} \exp\left(-\frac{1}{2\sigma_e^2}J(a,\omega)\right) \end{aligned}$$

Clearly maximising the likelihood is equivalent to minimising $J(a,\omega)$

(c) Hence show that the maximum likelihood solution for a is given by

$$a^{ML} = \frac{\mathbf{g}(\Omega)^H \mathbf{x}}{\mathbf{g}(\Omega)^H \mathbf{g}(\Omega)}$$

where $\mathbf{g}(\Omega) = [g_0(\Omega), g_1(\Omega), ..., g_{N-1}(\Omega)]^T$ [30%]

Solution:

$$\begin{aligned} J(a,\omega) &= \sum_{n=0}^{N-1} |x_n - a g_n(\omega)|^2 \\ &= |\mathbf{x} - a\mathbf{g}(\omega)|^2 \\ &= (\mathbf{x} - a\mathbf{g}(\omega))^H (\mathbf{x} - a\mathbf{g}(\omega)) \\ &= \mathbf{x}^H\mathbf{x} - 2\Re(\mathbf{x}^H \mathbf{g}(\omega) a) + |a\mathbf{g}(\omega)|^2 \end{aligned}$$

[recall that ^{H} is the complex conjugate of the transpose ('Hermitian transpose')]

First consider minimising $J(a, \omega)$ with respect to the amplitude. This could be done by differentiation wrt the real and imaginary parts of a. It's more straightforward however to rewrite the expression as:

$$\mathbf{x}^H\mathbf{x} - 2\Re(\mathbf{x}^H\mathbf{g}(\omega)a) + |a\mathbf{g}(\omega)|^2$$
$$= \mathbf{x}^H\mathbf{x} - \frac{|\mathbf{g}(\omega)^H\mathbf{x}|^2}{\mathbf{g}(\omega)^H\mathbf{g}(\omega)} + \mathbf{g}(\omega)^H\mathbf{g}(\omega)\left|(a - \frac{\mathbf{g}(\omega)^H\mathbf{x}}{\mathbf{g}(\omega)^H\mathbf{g}(\omega)})\right|^2$$

[verify this expression as an extra example question]

Now, the only term in this expression that depends on a is the final term, which is minimised (equals zero) when:

$$a^{ML} = \frac{\mathbf{g}(\omega)^H\mathbf{x}}{\mathbf{g}(\omega)^H\mathbf{g}(\omega)}$$

This is the ML solution for the amplitude.

(d) Finally, show that the maximum likelihood solution for Ω is found by minimising

$$J(a^{ML}, \Omega) = \mathbf{x}^H\mathbf{x} - \frac{|\mathbf{g}(\Omega)^H\mathbf{x}|^2}{\mathbf{g}(\Omega)^H\mathbf{g}(\Omega)}$$

Relate this final result to the standard periodogram frequency estimator, and comment on this relationship. [25%]

Solution:

At this value of a the final term in $J(a^{ML}, \omega)$ is zero, so:

$$J(a^{ML}, \omega) = \mathbf{x}^H\mathbf{x} - \frac{|\mathbf{g}(\omega)^H\mathbf{x}|^2}{\mathbf{g}(\omega)^H\mathbf{g}(\omega)}$$

To solve for ω^{ML} would involve a search of $J(a^{ML}, \omega)$ for all values of ω to find the minimum.

 (TURN OVER for continuation of Question 3

Note, however, that $\mathbf{x}^H\mathbf{x}$ does not depend upon ω while the second term has a very special (and familiar) form: The dot product $\mathbf{g}(\omega)^H\mathbf{x}$ is the DTFT of $\mathbf{x}$ evaluated at ω, since

$$\mathbf{g}(\omega)^H\mathbf{x} = \sum_{n=0}^{N-1} x_n \exp(-i\omega nT)$$

Moreover, the modulus-squared of this term divided by $\mathbf{g}(\omega)^H\mathbf{g}(\omega) = N$ is the periodogram:

$$\begin{aligned} J(a^{ML}, \omega) &= \text{const.} - \frac{|\sum_{n=0}^{N-1} x_n \exp(-i\omega nT)|^2}{N} \\ &= \text{const.} - \text{Periodogram}(\omega) \end{aligned}$$

(see Eq. (**??**)).

Thus, to find the ML frequency estimate simply find the frequency at the maximum value of the periodogram power spectrum estimate. This gives another strong justification for the use of the periodogram in the presence of deterministic sinusoidal components in noise.

4 From measurements of a stationary random process $\{x_n\}$, the autocorrelation sequence is estimated as

$$\hat{R}_{XX}[l] = 0.9^{|l|}$$

(a) Briefly describe the principal methods for improving the performance of the periodogram estimator of power spectra and state their properties. [20%]

Solution:

Principal methods are Bartlett, Blackman-Tukey and Welsh.

Bartlett:

- •Bartlett method averages periodogram estimates from successive sub-frames of the data:The Bartlett estimate is then given by:

$$\boxed{\hat{S}_X^B(e^{j\omega T}) = \frac{1}{K}\sum_{k=0}^{K-1} \hat{S}_X^{(k)}(e^{j\omega T})} \quad (2)$$

- •If the data subsequences are uncorrelated with one another the Bartlett procedure reduces the variance by a factor of K, by less if they are correlated.
- •Bartlett allows a trade-off between frequency resolution ($\propto N$) and variance of the estimate ($\propto 1/K$).
- •Reduction in variance is at the expense of requiring more data for the same resolution.

Blackman-Tukey

- •The Blackman-Tukey method applies a window function of length $2L+1$ to the estimated autocorrelation function:

$$\boxed{\hat{S}_X^{BT}(e^{j\omega T}) = \sum_{-L}^{L} w_l \hat{R}_{XX}[l] \exp(-j\omega T)} \quad (3)$$

where $L < N$ and w_l is any suitable window function, e.g. Hamming, Hanning, Bartlett,...

- •It is clear that the resulting spectrum can be written as a frequency domain convolution:

$$\hat{S}_X^{BT}(e^{j\omega T}) = \frac{1}{2\pi} W(e^{j\omega t}) * \hat{S}_X(e^{j\omega t})$$

where $W(.)$ is the DTFT of the window function and $\hat{S}_X(.)$ is the Periodogram.

 (TURN OVER for continuation of Question 4

- The B-T method can reduce the variance of the periodogram estimate at the expense of some frequency resolution. A special case is the correlogram considered also in the course

Welsh procedure:

- The Welch procedure performs averaging over frames as in the Bartlett method
- However, the periodograms are *modified* to incorporate a window function on the data:

$$\boxed{\hat{S}'^{(k)}(e^{j\omega T}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} w_n x_n^{(k)} e^{-j\omega nT}\right|^2}$$

with $1/N\sum_{n=0}^{N-1} w_n^2 = 1$.

- As for the Bartlett method, averaging is then performed over K frames:

$$\boxed{\hat{S}_X^W(e^{j\omega T}) = \frac{1}{K}\sum_{k=0}^{K-1} \hat{S}_X'^{(k)}(e^{j\omega T})} \qquad (4)$$

- The expected value of this spectral estimate can be shown to be:

$$E[\hat{S}_X^W(e^{j\omega T})] = \frac{1}{2\pi}V(e^{j\omega T}) * S_X(e^{j\omega T})$$

where $W(e^{j\omega T})$ is the DTFT of the window and $V(e^{j\omega T}) = \frac{1}{N}|W(e^{j\omega T})|^2$.

- When the segments are non-overlapping the variance is approximately that of the Bartlett estimate.

(b) Calculate the first order ($P = 1$) autoregressive model which corresponds to this data and show that the corresponding estimated power spectrum is:

$$S_X(e^{j\Omega}) = \frac{1-0.9^2}{|1-0.9e^{-j\Omega}|^2}$$

Solution:

Use Yule-Walker equations for P=1:

$$R_{XX}[0]a_1 = -R_{XX}[1]$$

 (cont.

So:

$$a_1 = -0.9$$

Also (gain term),

$$b_0^2 = R_{XX}[0] + R_{XX}[1]a_1 = 1 - 0.9^2 = 0.19$$

Hence, power spectrum is:

$$S_X(e^j) = \frac{b_0^2}{|1 + a_1 e^{-j}|^2} = \frac{1 - 0.81}{|1 - 0.9e^{-j}|^2}$$

[30%]

(c) Obtain an expression for the correlogram estimate of the power spectrum from the same data, using $2L+1$ autocorrelation values (i.e. using correlation lags $-L$ to $+L$). [Recall that the correlogram estimate uses only the central part of the estimated autocorrelation function in estimating the power spectrum.] [30%]

$$\begin{aligned}
\hat{S}_X(e^{j\omega T}) &= \sum_{k=-L}^{L} \hat{R}_{XX}[k]\, e^{-jk\omega T}, \\
&= \sum_{k=-L}^{L} 0.9^{|l|}\, e^{-jk\omega T} \\
&= e^{j0} + \sum_{k=1}^{L} 0.9^{|l|}(e^{-jk\omega T} + e^{+jk\omega T}) \\
&= e^{j0} + 0.9\frac{1 - 0.9e^{-jL\omega T}}{1 - 0.9e^{-j\omega T}} + 0.9\frac{1 - 0.9e^{jL\omega T}}{1 - 0.9e^{j\omega T}}
\end{aligned}$$

(d) What is the relationship between the two power spectrum estimates in b) and c) (you should consider both the case of medium and very large values of L)? [20%]

For medium L the correlogram estimate is the AR estimate convolved with the spectrum of a rectangular window function. For large L the effect of this convolution becomes negligible and the two estimates should be identical. (Note - this applies because the estimated R_{XX} is precisely that of a first order AR process.

END OF PAPER

Version 1- 8 Feb 2006