E3.08
ISE3.17

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2004

EEE/ISE PART III/IV: MEng, BEng and ACGI

ADVANCED SIGNAL PROCESSING

Monday, 26 April 10:00 am

Time allowed: 3:00 hours

There are FIVE questions on this paper.

Answer TWO of the questions 1, 2, 3 and ONE of the questions 4, 5.

All questions carry equal marks

Corrected Copy

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : D.P. Mandic
Second Marker(s) : D.B. Ward

1) Consider the problem of parametric autoregressive moving average (ARMA) modelling. Write down the equation of a general AR(p) model. [1]

a) Consider the first order autoregressive (Markov) process.

i) Derive the expression for the autocorrelation function for this process. [3]

ii) Write down and plot the autocorrelation function for an AR(1) process for the cases when the parameter $a = 0.9$ and $a = -0.9$. [3]

iii) What are the variance and spectrum of such a process? What can we say about the spectrum of an AR(1) process for a negative value of the parameter a? [3]

iv) Define the partial autocorrelation function and explain how the partial autocorrelation coefficients are calculated. Can the values of partial autocorrelation coefficents suggest the order of the AR model of a given process? [4]

b) Consider a general moving average (MA) process, MA(q).

i) What is the expression for the variance of this process? [2]

ii) Is the autocorrelation function finite or infinite in duration? [1]

iii) Consider the MA(1) process given by

$$z[n] = 0.8w[n-1] + w[n]$$

where w denotes the driving white noise sequence. Write down the expression for the spectrum of this process. Is this process invertible? [3]

2) Consider the problem of least squares (LS) estimation.

a) Sketch the block diagram of the data model and state the optimisation problem of the least squares estimation. [4]

b) Consider an optimisation problem, where measurements $\{x(n)\}$ are given, and the task is to find the best finite impulse response (FIR) model to filter signal $\{x(n)\}$ to yield desired signal $\{d(n)\}$. The output of the filter is given by $y(n) = \sum_{r=0}^{N} h(r)x(n-r) = \mathbf{h}^T(n)\mathbf{x}(n)$, where the output error is given by $e(n) = d(n) - y(n)$ and $\{h(n)\}$ are the filter coefficients.

i) Define the deterministic and stochastic error function J (in terms of the output error $e(n)$. [2]

ii) By using the method of least squares, derive the expression for the coefficient vector $\mathbf{h}$ which minimises the absolute squared error. [6]

iii) Define the orthogonality condition and explain its geometrical interpretation. [3]

c) Given a wide sense stationary random process $x(n)$, design a "linear" predictor that will predict the value $x(n+1)$ using a linear combination of $x(n)$ and $x(n-1)$. Thus the predictor for $x(n+1)$ is of the form

$$\hat{x}(n+1) = ax(n) + bx(n-1)$$

where a and b are constants. Assume that the process has zero mean $E\{x(n)\} = 0$ and that we want to minimise the mean square error

$$\xi = E\left\{[x(n) - \hat{x}(n)]^2\right\}$$

i) With $r_x(k)$ the autocorrelation of $x(n)$, determine the optimum predictor of $x(n)$ by finding the values of a and b that minimise the mean square error. [3]

ii) If $x(n+1)$ is uncorrelated with $x(n)$, what form does the predictor take? If $x(n+1)$ is uncorrelated with both $x(n)$ and $x(n-1)$, what form does the predictor take? [2]

3) Consider the problem of maximum likelihood estimation (MLE).

a) State the definition of the likelihood function. [2]

b) Derive the Cramer–Rao lower bound for the estimation of a DC level in white Gaussian noise ($w(n) \in \mathcal{N}(0, \sigma^2)$), given by

$$x[n] = A + w[n], \quad n = 0, 1, \ldots, N-1$$

for which the probability density function is given by

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right]$$ [8]

c) Derive the maximum likelihood estimator for the problem from b). [7]

d) Is the MLE estimator from c) unbiased? Is it efficient? Does such an MLE estimator attain the Cramer–Rao lower bound? [3]

4)

a) State the problem of sequential least squares. [2]

b) Explain the difference between the method of least squares and sequential least squares. [2]

c) What are the advantages of using a sequential estimator? [3]

d) A least squares estimator for signal $\{x\}$ given by

$$x(n) = A + w(n)$$

that is, DC level A in white Gaussian noise $\{w\}$, is given by

$$\hat{A}(N-1) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$$

i) Derive the sequential least squares estimator for this case. [6]

ii) Describe each of the terms that comprise the sequential estimator in i). [2]

iii) Derive the minimum least square error for the estimator in i). [2]

e) Consider a filter given by

$$y(n) = \frac{1}{4}\left[x(n) + x(n-1) + x(n-2) + x(n-3)\right]$$

Can such a filter be used to estimate a DC level in noise? Explain the difference between this estimator and the sequential least squares estimator? [3]

5) Adaptive linear prediction is at the core of adaptive filtering.

a) Sketch the block diagram for the adaptive prediction configuration. [4]

b) A finite impulse response (FIR) adaptive filter is employed within the adaptive prediction configuration. Derive the least mean square (LMS) algorithm for this filter, which minimises the cost function

$$J(n) = \frac{1}{2}e^2(n)$$

[6]

c) A sign–sign algorithm is used to reduce the computational complexity of the LMS algorithm. Write down the weight update of the sign–sign algorithm. What are the benefits and drawbacks of using this algorithm as compared with LMS? [4]

d) Let $x(n)$ be a second order autoregressive (AR) process that is generated according to the difference equation

$$x(n) = 1.2x(n-1) - 0.8x(n-2) + w(n)$$

where $w(n)$ is unit variance white noise. An adaptive FIR predictor is used to predict process $\{x(n)\}$.

i) Using the result from b) write down the equations for the LMS weight updates for such an adaptive predictor (A second order adaptive FIR filter). [2]

ii) Describe the bound on the step size which ensures the convergence of such a filter. What effect does the value of the step size have on the convergence trajectory on the error performance surface? [2]

iii) For the AR process above, what is the minimum mean square error achievable by using an adaptive FIR predictor? [2]

✓ E3.08

ADVANCED SIGNAL PROCESSING

2014

Solutions:

1)

$$z[n] = a_1 z[n-1] + a_2 z[n-2] + \cdots + a_p z[n-p] + w[n]$$

where $a_1, \ldots, a_p$ are the model parameters and $\{w[n]\}$ is the driving white noise.

a) i) The first order Markov process is given by [1]

$$z[n] = az[n-1] + w[n]$$

By applying the expectation operator $E\{\cdot\}$ to [3]

$$z[n-k]z[n]$$

we have

$$\rho(k) = a\rho(k-1) \quad or \quad \rho(k) = a^k, \quad k \geq 0$$

where $\rho(0) = 1$ and $\rho(1) = a$.

ii) The ACF for $a = \pm 0.9$ is $\rho(k) = (\pm 0.9)^k, k \geq 0$. The plots are a decaying function with or without alternating the sign (for a negative a). [3]

iii) For $k = 0$ the variance becomes

$$\sigma_z^2 = \frac{\sigma_w^2}{1 - a\rho(1)} = \frac{\sigma_w^2}{1 - a^2}$$

The spectrum of an AR(1) process is given by [3]

$$S(f) = \frac{2\sigma_w^2}{|1 - ae^{-j2\pi f}|^2} = \frac{2\sigma_w^2}{1 + a^2 - 2a\cos(2\pi f)}$$

For a negative a this represents a high–pass filter.

iv) Initially we may not know which order of autoregressive process to fit to an observed time series. This problem is analogous to deciding on the number of independent variables to be included in a multiple regression.
The partial autocorrelation function is a device which exploits the fact that whereas an AR(p) process has an autocorrelation function which is infinite in extent, it can by its very nature be described in terms of p nonzero functions of autocorrelations. Denote by a_{kj} the jth coefficient in an autoregressive representation of order k, so that a_{kk} is the last coefficient. The a_{kj} satisfy the set of equations

$$\rho(j) = a_{k1}\rho(j-1) + \cdots + a_{kk}\rho(j-k) \quad j = 1, 2, \ldots, k$$

leading to the Yule–Walker equations. The quantity a_{kk}, regarded as a function of lag k is called the partial autocorrelation function. The large values of the partial autocorrelation function may therefore indicate undermodelling. [4]

b) i) For the MA(q) process

$$z[n] = b_1 w[n-1] + \cdots + b_q w[n-q] + w[n] \qquad [2]$$

the variance is given by

$$var(MA(q)) = \left(1 + b_1^2 + \cdots + b_q^2\right) \sigma_w^2$$

ii) The ACF is finite in duration and has a length q. [1]

iii) The spectrum of an MA(q) process is given by

$$S(f) = 2\sigma_w^2 \left|1 - b_1 e^{-j2\pi f} - \cdots - b_q e^{-j2\pi q f}\right|^2 \qquad [3]$$

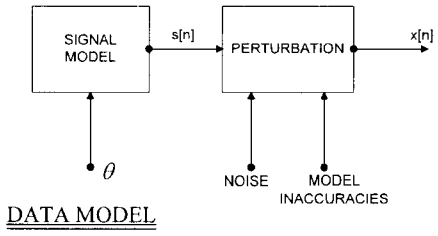
Theferore, for the given MA(1) we have

$$S(f) = 2\sigma_w^2[1 + 0.8^2 - 2 * 0.8\cos(2\pi f)], \quad 0 \leq f \leq 0.5$$

The value of $b = 0.8 \Rightarrow |b| < 1$ satisfies the invertibility condition.

2) a)

• Signal, $s[n]$ is assumed to be generated by the signal model which is a function



[4]

Figure 1: Data model for least squares estimation

of θ

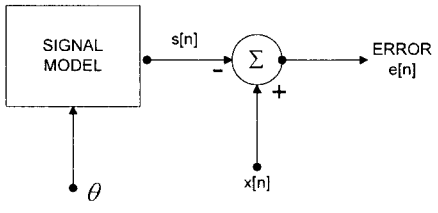
• The observation noise/model inaccuracies perturb s[n] to yield the measurement $x[n]$

• The Least Squares Estimator of θ chooses the value that makes $s[n]$ closest to the observed data $x[n]$, where closeness is measured by the LS error criterion

$$J(\theta) = \sum_{n=0}^{N-1} \underbrace{(x[n] - s[n])^2}_{e[n]}$$

LSE: $\min_\theta$ J(θ)

Note, no probabilistic assumptions have been made about the data $x[n]$



b)

i) We draw a distinction between stochastic and deterministic measures

(a) Stochastic

$$J = \min_{\mathbf{h}} E\{|e(n)|^p\}$$

[2]

See page 3

(b) Deterministic

$$J = \min_{\mathbf{h}} \sum_{n} |e(n)|^p$$

With $p = 2$, Problem (a) is known as the Wiener filtering problem, whereas Problem(b) is known as the Least Squares problem. These problems are also analytically easily tractable.

ii) **The absolute squared error**

$$|e(n)|^2 = |d(n) - \mathbf{h}^T\mathbf{x}||d(n) - \mathbf{h}^T\mathbf{x}| \qquad [6]$$

Or

$$|e(n)|^2 = |d(n)|^2 - \mathbf{g}^T\mathbf{h} - \mathbf{h}^T\mathbf{g} + \mathbf{h}^T\mathbf{\Phi}\mathbf{h}$$

where for the stochastic case

$$\mathbf{g} = E\{d(n)x(n-j)\}$$
$$\mathbf{\Phi} = E\{x(n-k)x(n-j)\} = \{\Phi_{kj}\}$$

while for the *deterministic* case, we have the same expressions, but **expectations** are replaced by **summations**.

In both cases
$\mathbf{g}$ is the crosscorrelation between the data and the desired signal, $\mathbf{\Phi}$ is the autocorrelation matrix of the data

Differentiating $|e(n)|^2$ wrt $\mathbf{h}$ and setting the result to zero, we obtain

$$0 = \mathbf{g} + \mathbf{\Phi}\mathbf{h}$$

or

$$\mathbf{h} = -|\mathbf{\Phi}^T|^{-1}\mathbf{g}$$

Differentiating again yields the autocorrelation matix, which is positive definite and hence we have a **minimum**.

iii)

$$e(n)\mathbf{h}^T = [d(n) - \mathbf{h}^T\mathbf{x}]\mathbf{x}^T = d(n)\mathbf{x}^T - \mathbf{h}^T\mathbf{x}\mathbf{x}^T \qquad [3]$$

On taking the expectation we obtain

$$E\{e(n)\mathbf{h}^T\} = E\{d(n)\mathbf{x}^T - \mathbf{h}^T\mathbf{x}\mathbf{x}^T\} = \mathbf{g} - \mathbf{h}^T\mathbf{\Phi} = 0$$

This is known as the **orthogonality** condition

at the optimum the error vector is orthogonal to the data

c) The MSE we want to minimise is

$$\xi = E\{[x(n+1) - \hat{x}(n+1]^2\} = E\{x^2(n+1) - 2x(n+1)\hat{x}(n+1) + \hat{x}^2(n+1)\}$$

i) Since the estimate of $x(n+1)$ is

$$\hat{x}(n+1) = ax(n) + bx(n-1)$$

then setting the derivative of ξ wrt a and b equal to zero, we have

$$\frac{\partial \xi}{\partial a} = -2E\{x(n+1)x(n) + E\{2\hat{x}(n+1)x(n)\}\} = 0$$

$$\frac{\partial \xi}{\partial b} = -2E\{x(n+1)x(n-1) + E\{2\hat{x}(n+1)x(n-1)\}\} = 0$$

Dividing by 2 and substituting for $\hat{x}(n+1)$ gives

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(0) \end{bmatrix}$$

Solving for a and b we find

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{r_x^2(0) - r_x^2(1)} \begin{bmatrix} r_x(0)r_x(1) - r_x(1)r_x(2) \\ r_x(0)r_x(2) - r_x^2(1) \end{bmatrix}$$

ii)
If $x(n)$ and $x(n+1)$ are uncorrelated, then $r_x(1) = 0$ and the values for a and b become

$$a = 0 \qquad b = r_x(2)/r_x(0)$$

In this case, the linear predictor is

$$\hat{x}(n+1) = \frac{r_x(2)}{r_x(0)} x(n-1)$$

Similarly, if $x(n+1)$ is uncorrelated with both $x(n)$ ands $x(n-1)$, then the values for a and b are

$$a = b = 0$$

and the linear predictor os

$$\hat{x}(n+1) = 0$$

which is equal to the expected value of $x(n+1)$

$$\hat{x}(n+1) = E\{x(n+1)\}$$

3) a) The likelihood function is a probability density function defined in terms of the unknown parameter to be estimated, that is $p(x, \theta)$.

[2]

b) The CRLB for A

$$p(\underline{x};\theta) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

Taking the first derivative

$$\frac{\partial ln\ p(\underline{x};A)}{\partial A} = \frac{\partial}{\partial A}\left[-ln\left[(2\pi\sigma^2]^{N/2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]\right.$$

$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)$$

$$= \frac{N}{\sigma^2}(\bar{x}-A)$$

where $\bar{x}$ is the sample mean.

Differentiating again

$$\frac{\partial^2 ln\ p(\underline{x};A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

[3]

Therefore $Var(\hat{A}) \geq \frac{\sigma^2}{N}$ is the CRLB.

c) DC level in WGN

$$x[n] = A + w[n] \qquad n=0,1,...,N\text{-}1$$

$\uparrow$
A to be estimated

$$w[n] \sim N(0,\sigma^2)$$

[7]

PDF $p(\underline{x};A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$

Take the derivative of the log-likelihood function

$$\frac{\partial \ln p(\underline{x};A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)$$

Set the result to zero to yield the MLE

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

d) In this case the MLE is clearly the MVU estimator which yields the CRLB, hence it is efficient.
If an efficient estimator exists the maximum likelihood procedure will produce it. The MLE has the desirable feature that it yields "an asymptotically efficient" estimator - namely one, that for sufficiently large datasets, that is unbiased and it attains the CRLB.

[3]

4) a) The problem: Given we know the LSE $\underline{\theta}$ based on $\{x[0]x[1]\dots x[N-1]\}$, and then we observe $x[N]$, can we update $\underline{\theta}$ (in time) without having to solve the normal equations? [2]

b) Data can be collected sequentially, namely one point at a time. We can either wait until all the data points (samples) are collected and make an estimate of the unknown parameter, namely the block-based approach or least squares, or refine our estimate in time as each new sample arrives, the sequential approach or sequen tial least squares. [2]

c) Such an estimator is computationally much less demanding and can be run on line.

d) i) [6]

$$\hat{A}[N-1] = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

If we now observe, $x[N]$, we can rewrite the LSE

$$\hat{A}[N] = \frac{1}{N+1}\sum_{n=0}^{N} x[n] = \frac{1}{N+1}\Big(\sum_{n=0}^{N-1} x[n] + x[N]\Big)$$

$$\Rightarrow \hat{A}[N] = \frac{N}{N+1}\hat{A}[N-1] + \frac{1}{N+1}x[N]$$

Clearly $\hat{A}[N]$ can be calculated from $\hat{A}[N-1]$ together with the new observation $x[N]$.

[2]

ii) Equation I) can be rewritten as

$$\underbrace{\hat{A}[N]}_{\text{New estimate}} = \underbrace{\hat{A}[N-1]}_{\text{Old estimate}} + \frac{1}{N+1}\Big(\underbrace{x[N] - \hat{A}[N-1]}_{\text{correction term - error in prediction x[n] by the previous sample}}\Big)$$

iii) The minimum LS error may also be computed recursively

$$J_{\min}[N] = J_{\min}[N-1] + \frac{N}{N+1}(x[N] - \hat{A}[N-1])^2$$ [2]

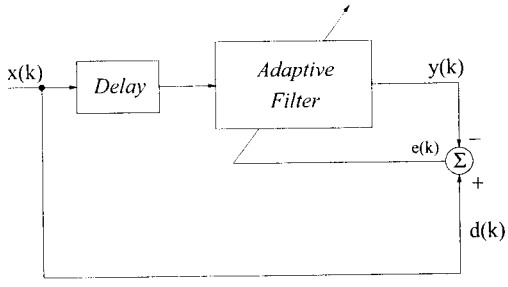
e) This filter is a moving average filter, working on only four signal samples at a time. It is fast to estimate the unknown parameter value but is more prone to noise since it is not adaptive and is very short.
In fact it is a low pass filter, with [3]

$$H(z) = \frac{1-z^{-4}}{1-z^{-1}}$$

which filters out the low pass component of the additive WGN.

Therefore, both the least squares and sequential least squares provide a better estimate of a DC level in WGN and they are asymptotically unbiased and consistent.

5) a)



[4]

Figure 2: Adaptive prediction configuration

b) The cost function for the LMS algorithm is given by

$$J(k) = \frac{1}{2}\, e^2(k)$$

which is based on the instantaneous output error $e(k) = d(k) - y(k)$. The instantaneous gradient is

$$\frac{\partial J(k)}{\partial \mathbf{w}(k)} = e(k)\, \frac{\partial e(k)}{\partial \mathbf{w}(k)}$$

Following a general gradient descent procedure algorithm

$$\frac{\partial e(k)}{\partial \mathbf{w}(k)} = -\,\mathbf{x}(k)$$

and

$$\frac{\partial J(k)}{\partial \mathbf{w}(k)} = -\,e(k)\mathbf{x}(k)$$

The set of equations that describes the LMS algorithm is then given by

$$\begin{aligned} y(k) &= \sum_{i=1}^{N} x_i(k) w_i(k) \;=\; \mathbf{x}^T(k)\mathbf{w}(k) \\ e(k) &= d(k) - y(k) \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \eta\, e(k)\mathbf{x}(k) \end{aligned}$$

[6]

c) The sign–sign algorithm combines the sign–error and sign–regressor algorithms, and is given by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta sign(e(n)) sign(\mathbf{x}(n))$$

It is very convenient to implement in hardware, especially if the step size is chosen to be a multiple of 2. In that case it is extremely fast. The drawback is an increased mean square error as compared to LMS. [4]

d) i)

$$w_1(n+1) = w_1(n) + \mu e(n)x(n-1)$$
$$w_2(n+1) = w_2(n) + \mu e(n)x(n-2)$$

[2]

where μ is the learning rate.

ii) The step size of the LMS is bounded by the reciprocal of the maximum eigenvalue of the input autoccorelation matrix. In this case, $r_x(0) = 5.7523, r_x(1) = 4.0450$ (not expected to calculate in the exam). The step size bounds now become

$$0 < \mu < 0.204$$

[2]

iii) The minimum mean square error for this example is

$$\xi_{min} = \sigma_w^2 = 1$$

Namely after the filter coefficients converge to the true values of the AR parameters, there is still the "unpredictable" part due to the error term in the AR model, which in this case has unit variance. [2]