E3.08
ISE3.17

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2006

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

Corrected Copy

Q2

ADVANCED SIGNAL PROCESSING

Friday, 28 April 10:00 am

Time allowed: 3:00 hours

There are FIVE questions on this paper.

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

All questions carry equal marks

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

Examiners responsible First Marker(s) : D.P. Mandic, D.P. Mandic
Second Marker(s) : M. Petrou, M. Petrou

1) Consider the problem of estimating the value of an unknown scalar parameter, θ, from a sequence of random variables $x[n], \quad n = 1, 2, \ldots, N$. The estimate $\hat{\theta}$ is a function of N random variables, and will be denoted by $\hat{\theta}_N$.

a) Define the notion of bias B in parameter estimation. Define an unbiased estimate. [2]

b) Define a minimum variance unbiased (MVU), asymptotycally unbiased, and consistent estimator. [6]

i) Define the term "mean square convergence". [2]

ii) Is the sample mean estimate $\hat{m}_x = \frac{1}{N}\sum_{n=1}^{N} x[n]$ unbiased and consistent? [2]

c) Let x be the random variable defined on a coin flipping experiment, with $x = 1$ if the outcome is heads and $x = -1$ if the outcome is tails. The coin is unfair so that the probability of flipping *heads* is p and the probability of flipping *tails* is $(1 - p)$.

i) Find the mean of x. [2]

ii) Suppose the value for p is unknown and that the mean of x is to be estimated. Flipping the coin N times and denoting the resulting values for x by $x[i], \quad i = 1, \ldots, N$, consider the following estimator for m_x

$$\hat{m}_x = x[N]$$

Is this estimator unbiased? [3]

iii) Does the accuracy of this estimator improve as the number of observations N increases? [3]

2) Consider the problem of mixed autoregressive moving average (ARMA) modelling.

a) State the equation of a general ARMA(p, q) process. [2]

i) Derive the expression for the autocorrelation function $r_{zz}[k]$ of this process. What is the expression for the autocorrelation function $r_{zz}[k]$ for $k \geq q + 1$? [6]

ii) Discuss the stationarity conditions for this process. [2]

iii) State and explain the equation for the power spectrum of a general ARMA(p, q) process. [3]

b) Consider the process

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

where $w[n]$ is white noise.

i) Is the process $z[n]$ is stationary and invertible? [3]

ii) What are the values of the first two correlation coefficients $\rho[0]$ and $\rho[1]$? [2]

iii) Write down the expression for the sequence of autocorrelation coefficients $\rho[k]$ of this process. [2]

3) Consider the problem of Minimum Variance Unbiased (MVU) estimation.

a) Define the likelihood function for a random signal x. State the likelihood function for a single random sample $x[0] = A + w[0]$, where A is the DC level, and $w \sim \mathcal{N}(0, \sigma^2)$. [4]

b) Define the curvature of the log–likelihood function. What does the curvature give information about? In your own words explain why it is convenient to use the log–likelihood function. [6]

c) Define and discuss the Cramer–Rao Lower Bound (CRLB) for the scalar parameter. [5]

d) Consider the estimation of a DC level in white Gaussian noise (WGN), and assume N observations

$$x[n] = A + w[n] \qquad n = 0, 1, 2, \ldots, N-1$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$. Determine the CRLB for A. [5]

$$\left(\text{Hint :} \quad \mathrm{p}(\underline{\mathrm{x}}; \mathrm{A}) = \frac{1}{(2\pi\sigma^2)^{\mathrm{N}/2}} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{\mathrm{n}=0}^{\mathrm{N}-1}(\mathrm{x}[\mathrm{n}]-\mathrm{A})^2} \right)$$

4) Consider the method of least squares (LS).

a) State the least squares optimisation problem for the estimation of a vector parameter. [4]

b) Explain the role of the observation matrix $\mathbf{H}$. In your own words comment on the physical meaning of the vectors making up the columns of the observation matrix. [6]

c) We would like to build a predictor of digital waveforms. Such a system forms an estimate of a later sample (say n_0 samples later) by observing p consecutive data samples, and is given by

$$\hat{x}[n+n_0] = \sum_{k=1}^{p} a_p[k]x[n-k]$$

The predictor coefficients $a_p[k]$ are to be chosen to minimize

$$E_p = \sum_{n=0}^{\infty} \left(x[n+n_0] - \hat{x}[n+n_0]\right)^2$$

Derive the equations that define the optimum set of coefficients $a_p[k]$. [7]

d) Discuss the advantages of using the method of least squares. [3]

5) Explain the need for adaptive filtering. Comment on the advantages and disadvantages of adaptive filters as compared to the Wiener filter (suitability for a non–stationary operating environment, mode of operation, accuracy). [3]

a) Draw a block diagram of the inverse system modelling configuration and explain its application for channel equalisation in non–stationary environments. Is there any delay in the input–output operation of this configuration? [4]

b) Explain the need for an adaptive step size within the Least Mean Square (LMS) algorithm. Describe in your own words the behaviour of an ideal adaptive step size. [2]

c) Write down the expression for the learning rate of the Normalised LMS (NLMS) algorithm. [1]

i) Derive the learning rate of the NLMS algorithm by expanding the output error $e(k+1)$ of the LMS algorithm using Taylor series around $e(k)$ and setting $e(k+1) = 0$. [4]

ii) Give the physical justification for the use of η_{NLMS}. [1]

d) An $AR(3)$ process $x(n)$ is generated by the difference equation

$$x(n) = 1.3x(n-1) - 0.75x(n-2) + 0.1x(n-3) + w(n), \quad w(n) \sim \mathcal{N}(0,1)$$

i) Write down the output $\hat{x}(n)$ of a three–coefficient LMS–type adaptive predictor for this process. [2]

ii) Write down the expression for the NLMS weight updates of such an adaptive predictor. [3]

ADVANCED SIGNAL PROCESSING E3.08

ISE 3.17

2006 Master copy - 10.4.06

1/9

Solutions:

1) a) Bias $B = \theta - E\{\hat{\theta}_N\}$.
If the bias is zero, then the expected value of the estimate is equal to the true value, i.e. $E\{\hat{\theta}_N\} = \theta$ and the estimator is said to be unbiased. [**bookwork**]

b) An MVU estimator is unbiased and attains the Cramer Rao Lower Bound.

An estimator is asymptotically unbiased if an estimate is biased but the bias goes to zero as the number of observations, N goes to infinity, that is

$$\lim_{N\to\infty} E\{\hat{\theta}_N\} = \theta$$

An estimator is consistent if it is unbiased in the mean squared error sense. [**bookwork**]

i) An estimate $\hat{\theta}_N$ is said to converge to θ in the mean square sense if [**bookwork**]

$$\lim_{N\to\infty} E\{|\hat{\theta}_N - \theta|^2\} = 0$$

ii) The sample mean estimate is unbiased, since $E\{\hat{m}_x\} = \frac{1}{N}\sum_{n=1}^{N} E\{x[n]\} = m_x$. Since the variance of the sample mean estimate is

$$var\{\hat{m}_x\} = \frac{1}{N^2}\sum_{n=1}^{N} var\{x[n]\} = \frac{\sigma_x^2}{N}$$

which goes to zero as $N \to \infty$, it follows that the sample mean is a consistent estimator. [**application of theory**]

c) i) $m_x = E\{x\} = p \cdot 1 + (-1)\cdot(1-p) = 2p - 1$. [**new computed example**]

ii) For an estimator $\hat{m}_x = x[N]$, the mean is

$$E\{\hat{m}_x\} = E\{x[N]\} = 2p - 1$$

and the estimator is unbiased. However, $\hat{m}_x = x[N]$ is not a good estimator of the mean. [**new computed example**]

iii) The estimate of the mean, $\hat{m}_x$ will either be equal to one, with probability p or it will be equal to minus one, with a probability of $(1-p)$. Therefore the accuracy of the estimate $E\{\hat{m}_x\} = x[N]$ does not improve as the number of observations N increases. The variance of the estimate

$$var\{\hat{m}_x\} = var\{x[N]\} = 4p(1-p)$$

does not decrease with N. The estimator does not converge in the mean square sense and is therefore not consistent. [**combination of theory and a new example**]

2/9

2)

a) For an ARMA(p,q) random processes $z[n]$, the process $z[n]$ and driving noise $w[n]$ are related by a linear constant coefficient equation [**bookwork**]

$$z[n] = \sum_{l=1}^{p} a_p(l) z[n-l] + \sum_{l=0}^{q} b_q(l) w[n-l]$$

i) The autocorrelation function of $z[n]$ and crosscorrelation between $z[n]$ and $w[n]$ follow the same functional expression as that of an ARMA(p,q) model above. Multiply both sides of the above equation by $z[n-k]$ and apply the statistical expectation operator, to yield [**combination of bookwork and worked example**]

$$r_{zz}[k] = \sum_{l=1}^{p} a_p(l) r_{xx}[k-l] + \sum_{l=0}^{q} b_q(l) r_{zw}[k-l]$$

Since for $k \geq q+1$, there is no correlation between $z[n]$ and $w[n]$, the ACF follows the AR part of the above equation, that is

$$r_{zz}(k) = \sum_{l=1}^{p} a_p(l) r_{xx}[k-l] \quad for \quad k \geq q$$

ii) One representation of a stochatic process is as an output from a linear filter, whose input is white noise $w[n]$, given by

$$z[n] = w[n] + a_1 w[n-1] a_2 w[n-2] + \cdots = w[n] + \sum_{j=1}^{\infty} a_j w[n-j]$$

(weighted sum of past inputs $w[n]$).

For this process to be a valid stationary process, the coefficients must be absolutely summable, that is, $\sum_{j=0}^{\infty} |a_j| < \infty$.

Under some mild conditions, $z[n]$ can also be represented as a weighted sum of its past values, and an added random "shock" $w[n]$, given by

$$z[n] = b_1 z[n-1] + b_2 z[n-2] + \cdots + w[n]$$

This process is *stationary* if $\sum_{j=0}^{\infty} |a_j| < \infty$, and is *invertible* if $\sum_{j=0}^{\infty} |b_j| < \infty$. [**bookwork**]

iii) From the $\mathcal{Z}$–domain representation of an ARMA(p,q) process, we have

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}}$$

Assuming that the filter is stable, the output process $z[n]$ will be wide–sense stationary and with $P_w = \sigma_w^2$, the power spectrum of $z[n]$ will be

$$P_z(z) = \sigma_w^2 \frac{B_q(z) B_q(z^{-1})}{A_p(z) A_p(z^{-1})}$$

or in terms of frequency θ

$$P_z(e^{\jmath\theta}) = \sigma_w^2 \frac{\left|B_q(e^{\jmath\theta})\right|^2}{\left|A_p(e^{\jmath\theta})\right|^2}$$

[**bookwork**]

b)

i) This process is invertible since $|a_1| = 0.8 < 1$, and is stationary due to the stationarity of $w[n]$. [**application of theory**]

ii) This is clearly an AR(1) (Markov) process. Hence $\rho[0] = 1, \rho[1] = 0.8$. [**application of theory**]

iii) For this AR(1) process, we have $\rho[k] = 0.8^k, \quad k \geq 0$. [**application of theory**]

4/9

3)[**parts a), b), c, bookwork**]

a) When the PDF is viewed as a function of the unknown parameter (with the dataset x fixed) it is term the "likelihood function".
For the random variable $x[0] = A + w[0]$ we have

$$ln\ p(x[0]; A) = -ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x[0] - A)^2$$

b) The "sharpness" of the likelihood function determines the accuracy with which the unknown parameter may be estimated. This sharpness is effectively measured by the negative of the second derivative of the logarithm of the likelihood function at its peak - the "curvature" of the log-likelihood function. Generally, the second derivative does depend upon $x[0]$, and hence a more appropriate measure of curvature is

$$-E\left[\frac{\partial^2 ln\ p(x[0]; A)}{\partial A^2}\right]$$

which measure the average curvature of the log-likelihood function.
Applying the logarithm to the likelihood function helps with mathematical tractability, especially for Gaussian signals, since the products are converted into sums and also the exponentials are avoided.

c) Cramer-Rao Lower Bound-Scalar Parameter

Under the assumption that the PDF $p(\underline{x}; \theta)$ satisfies the "regularity" condition

$$E\left[\frac{\partial ln\ p(\underline{x}; \theta)}{\partial \theta}\right] = 0 \qquad \forall\theta$$

where the expectation is taken with respect to $p(\underline{x}; \theta)$, then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$Var(\hat{\theta}) \geq \frac{1}{-E\left\{\frac{\partial^2 ln\ p(\underline{x};\theta)}{\partial \theta^2}\right\}}$$

where the derivative is evaluated at the true value of θ.

Moreover, an unbiased estimator may be found that attains the bound for all θ. If and only if

$$\frac{\partial ln\ p(\underline{x}; \theta)}{\partial \theta} = I(\theta)(g(\underline{x}) - \theta)$$

for some functions g and I.

That estimator is the MVU estimator, with $\hat{\theta} = g(\underline{x})$ and the minimum variance $\frac{1}{I(\theta)}$.

d) [**application of theory**]

To determine the CRLB for the estimation of a DC level in WGN, assume N observations

$$x[n] = A + w[n] \qquad\qquad n = 0, 1, 2, \ldots, N-1$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$, and

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

Take the first derivative of the log–likelihood function to yield

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

where $\bar{x}$ is the sample mean.
Differentiate once again to have

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

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

4) a) [**bookwork**] LSE is found by minimising

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

$$= (\underline{\boldsymbol{x}} - H\underline{\theta})^T(\underline{\boldsymbol{x}} - H\underline{\theta})$$

$$= \underline{\boldsymbol{x}}^T\underline{\boldsymbol{x}} - 2\underline{\boldsymbol{x}}^T H\underline{\theta} + \underline{\theta}^T H^T H\underline{\theta}$$

$$\frac{\partial J(\theta)}{\partial \underline{\theta}} = -2H^T\underline{\boldsymbol{x}} + 2H^T H\underline{\theta}$$

set the result to zero to yield

$$\underline{\hat{\theta}} = (H^T H)^{-1}H^T\underline{\boldsymbol{x}}$$

$$\Rightarrow (H^T H)\underline{\hat{\theta}} = H^T\underline{\boldsymbol{x}} \qquad \text{The "normal" equations}$$

b) Geometrical interpretations [**combination of bookwork and application of theory**]

Given $\underline{\boldsymbol{s}} = H\underline{\theta}$

$$= [\underline{\boldsymbol{h}}_1\underline{\boldsymbol{h}}_2 ... \underline{\boldsymbol{h}}_p]\begin{bmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_p\end{bmatrix}$$

$$= \sum_{i=1}^{p}\theta_i\underline{\boldsymbol{h}}_i$$

Signal vector model is the linear combination of the "signal" vectors $\{\underline{\boldsymbol{h}}_1\underline{\boldsymbol{h}}_2 ... \underline{\boldsymbol{h}}_p\}$ Therefore, to determine **s**, we use the so-called orthogonality condition

$$(\underline{\boldsymbol{x}} - \underline{\boldsymbol{s}}) \perp S^2$$

$$\Rightarrow \quad (\underline{\boldsymbol{x}} - \underline{\boldsymbol{s}}) \perp \underline{\boldsymbol{h}}_1 \qquad \rightarrow (\underline{\boldsymbol{x}} - \underline{\boldsymbol{s}})^T \cdot \underline{\boldsymbol{h}}_1 = 0 \qquad \text{-A}$$

$$(\underline{\boldsymbol{x}} - \underline{\boldsymbol{s}}) \perp \underline{\boldsymbol{h}}_2 \qquad \rightarrow (\underline{\boldsymbol{x}} - \underline{\boldsymbol{s}})^T \cdot \underline{\boldsymbol{h}}_2 = 0 \qquad \text{-B}$$

to yield the LSE

$$\underline{\hat{\theta}} = (H^T H)^{-1}H^T\underline{\boldsymbol{x}}$$

Noting that

$$\underline{\varepsilon} = \underline{\boldsymbol{x}} - H\underline{\theta}$$

is the error vector the LSE is found by

7/9

the error vector must be orthogonal to
$\underline{\varepsilon}^T \mathrm{H} = \underline{\boldsymbol{0}}^T \rightarrow$ the columns of H - this is the "Orthogonality Principle"

The error represents that part of $\underline{\boldsymbol{x}}$ which is not described by the signal model.

c) [**new example**]

We want to find the predictor coefficients $a_p[k]$ that minimise the linear prediction error $E_p = \sum_{n=0}^{\infty}(e[n])^2$.
To find these coefficients, differentiate E_p with respect to $a_p[k]$ and set the derivatives equal to zero as follows

$$\frac{\partial E_p}{\partial a_p[k]} = -\sum_{n=0}^{\infty} 2e[n] \frac{\partial \hat{x}[n+n_0]}{\partial a_p[k]} = 0$$

From $\hat{x}[n+n_0] = \sum_{k=1}^{p} a_p[k]x[n-k] \quad \Rightarrow \quad \frac{\partial \hat{x}[n+n_0]}{\partial a_p[k]} = x[n-k]$
Divide by two, and substitute for $e[n]$ to have,

$$\sum_{n=0}^{\infty} \left\{ x[n+n_0] - \sum_{l=1}^{p} a_p[l]x[n-l] \right\} x[n-k] = 0 \quad ; \quad k = 1, 2, \ldots, p$$

Therefore we obtain the normal equations

$$\sum_{l=1}^{p} a_p[l] r_x[k,l] = r_z[k, -n_0] \quad where \quad r_z[k,l] = \sum_{n=0}^{\infty} x[n-l]x[n-k]$$

d) [**bookwork and above example**]

No probability assumptions are made about the data; only a signal model is assumed. Usually easy to implement, either in a block based or sequential manner, amounts to the minimisation of a least squares criteria. In the (LS) approach we attempt to minimise the squared difference between the observed data and the assumed signal or noiseless data.

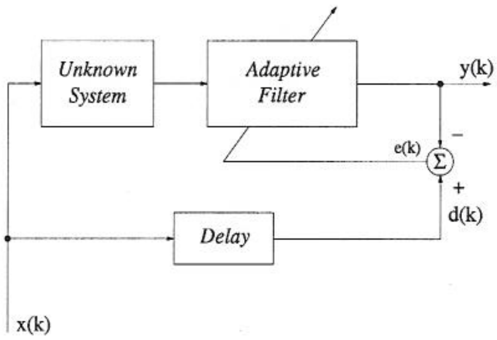
8/9

5) [**bookwork**]

Suitable for filtering of nonstationary data and sequential mode of operation. Due to the approximations in the derivation, in the steady state they are not as accurate as Wiener filters, but are much less complex and capable of an on–line mode of operation.

a) [**bookwork and intuitive reasoning**]

The adaptive filter is in cascade with the unknown channel and aims at estimating the inverse of the channel model. Application: adaptive channel equalisation in telecommunications, wher an adaptive system tries to compensate for the possibly time–varying communication channel, so that the transfer function from the input to the output (Figure below) approximates a pure delay. We need a delay



in the system, since we are dealing with sampled data systems and need time to propagate signals through filters.

b) [**bookwork and intuitive reasoning**]

In order to cope with the nonstationarity of a signal and changing signal dynamics we need adaptive step sizes. Ideally, a step size would be large in the beginning of adaptation and small when approaching the optimal Wiener solution.

c) [**bookwork**] $\eta_{NLMS} = \frac{1}{\|\mathbf{x}(n)\|_2^2}$

i) [**worked example**]

$$e(n+1) = e(n) + \sum_{k=1}^{p} \frac{\partial e(k)}{\partial w_k(n)} \Delta w_k(n) + \text{Higher Order Terms}$$

Inserting the partial derivatives from the above, we arrive at

$$e(k+1) = e(k)\left[1 - \eta \parallel \mathbf{x}(n) \parallel_2^2\right]$$

9/9

From there the NLMS step size which minimizes the error is

$$\eta_{NLMS} = \frac{1}{\| \mathbf{x}(n) \|_2^2}$$

ii) [**bookwork and intuitive reasoning**]

Normalisation of the learning rate by the tap input power helps with the conditioning of the error performance surface, and hence faster adaptation.

d) [**new example**]

i)

$$\hat{x}(n) = w_n(1)x(n-1) + w_n(2)x(n-2) + w_n(3)x(n-3)$$

ii)

$$w_{n+1}(k) = w_n(k) + \frac{1}{x^2(n-1) + x^2(n-2) + x^2(n-3)} e(n)x(n-k), \quad k = 1, 2, 3$$