IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2011

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

ADVANCED SIGNAL PROCESSING

Thursday, 19 May 10:00 am

Corrected Copy

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) :	T-K. Kim, T-K. Kim

1) Consider the problem of minimum variance unbiased (MVU) estimation. The estimates obtained from N data points shall be denoted by subscript N, that is, μ_N is the estimate of signal mean obtained from N data points.

a) Define the notions of bias $B(\hat{\theta})$ and variance $\mathrm{var}(\hat{\theta})$ in parameter estimation. When do we say that an estimator is unbiased and with minimum variance? [3]

b) The sample mean

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

is used as an estimator of the mean value of an independent and identically distributed discrete time random sequence $\{x[0], \ldots, x[N-1]\}$ with zero mean and variance σ^2. Derive the expressions for the bias and variance of such an estimator and show whether this is a minimum variance unbiased (MVU) estimator. [5]

c) The Mean Squared Error (MSE) criterion, given by

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\Big\{ \big(\hat{\theta} - \theta\big)^2 \Big\}$$

measures the average mean squared deviation of the estimate $\hat{\theta}$ from the true value of the parameter θ. Derive the so called "bias-variance" tradeoff formula, that is

$$\mathrm{MSE}\big(\hat{\theta}\big) = \mathrm{var}\big(\hat{\theta}\big) + \mathrm{B}^2\big(\hat{\theta}\big)$$

and explain in your own words whether this criterion guarantees an MVU estimator. [6]

d) The variance of the random signal from part b) is estimated using the estimator

$$\hat{\sigma}^2 = \frac{\alpha}{N}\sum_{n=0}^{N-1} x^2[n]$$

For $\alpha = 1$ this is an MVU estimator with the variance

$$\mathrm{var}(\hat{\sigma}^2) = \frac{2\sigma^4}{\mathrm{N}}$$

and thus achieves the Cramer-Rao lower bound (CLRB).

Using the result in part c) show that the MSE estimator obtains the minimum value of MSE for $\alpha_{min} = \frac{N}{N+2}$, reaching

$$\min \mathrm{MSE}(\hat{\sigma}^2) = \frac{2\sigma^4}{\mathrm{N}+2}$$

Is this estimator unbiased? Give reasons for this estimator to achieve lower variance than that given by CLRB. [6]

2) Assume that a dataset $x[n]$ was generated by an autoregressive (AR) model of order p driven by white Gaussian noise $w[n]$ with zero mean and variance σ_w^2.

a) Write down the general mathematical form for an $AR(p)$ model and derive the expressions for the autocorrelation function and power spectrum of this process. Explain how the power spectrum can be obtained from the autocorrelation function of an $AR(p)$ process. [6]

b) Figure 2.1 shows the power spectral density corresponding to an unknown ARMA(p,q) process.

i) Explain whether this power spectrum is better modelled by an AR or MA model. [3]

ii) Explain what would be the minimum order of the $AR(p)$ process which can generate such a power spectrum. [3]

iii) Indicate what changes you would observe in the shape of this power spectrum in case of overmodelling by a factor of two. [2]

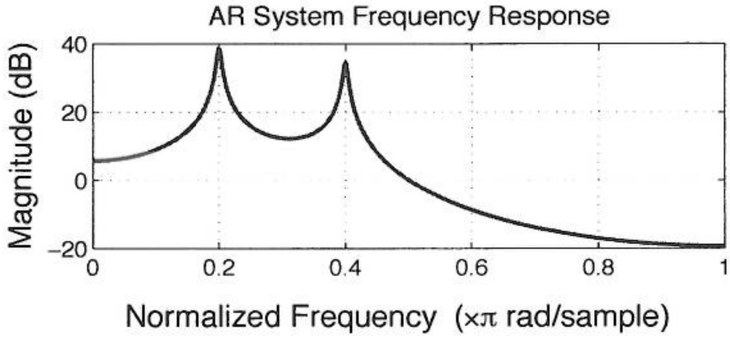


Figure 2.1: Frequency response of an autoregressive process

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

i) Write down the mathematical expressions for this process and for its variance. Is the autocorrelation function finite or infinite in duration? [2]

ii) Consider the MA(1) process given by

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

where w denotes the driving white noise sequence. Write down the expression for the spectrum of this process and comment on whether this process is invertible. [4]

3) Consider the problem of Minimum Variance Unbiased (MVU) estimation based on a linear data model given by

$$x[n] = A + Bn + w[n], \qquad n = 0, 1, \ldots, N-1, \quad w[n] \sim \mathcal{N}(0, \sigma^2)$$

a) For a known B ($B = 0$) the problem reduces to that of estimating a DC level A in white Gaussian noise.

   i) Define the likelihood function for the random signal x. In your own words explain why it is convenient to use the log–likelihood function. [3]

   ii) Define the curvature of the log–likelihood function. What does the curvature give information about? [3]

   iii) State the Cramer-Rao lower bound theorem for a scalar parameter. [4]

b) In the DC level estimation from noisy observation in part a), in addition to the DC level A the noise variance σ^2 is also unknown. The unknown parameter vector has the form

$$\boldsymbol{\theta} = [A, \sigma^2]^T$$

Show that the Fisher information matrix for this case becomes [4]

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2}\right] & -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A \partial \sigma^2}\right] \\ -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \sigma^2 \partial A}\right] & -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial {\sigma^2}^2}\right] \end{bmatrix} = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

c) For the linear data model, for which the unknown parameter vector is $\boldsymbol{\theta} = [A, B]^T$, show that the Fisher information matrix is given by [4]

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} N & \frac{N(N-1)}{2} \\ \frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

(Hint: $\sum_{n=0}^{N-1} n = N(N-1)/2$, $\sum_{n=0}^{N-1} n^2 = N(N-1)(2N-1)/6$)

d) It can be shown that the Cramer-Rao lower bounds (CRLB) for parameters A and B from part c) are given by

$$\text{CRLB(A)} \approx \frac{4\sigma^2}{\text{N}} \qquad \text{CRLB(B)} \approx \frac{12\sigma^2}{\text{N}^3}$$

Compare with the CRLB for $B = 0$ obtained in a) and comment on the increase in the minimum achievable variance when B is unknown. [2]

4) It is often important in practical applications to employ least squares type of estimation to model a signal $x(n)$ that we believe is quasiperiodic. Suppose the following autocorrelations are known

$$r_x(0) = 1.0, r_x(1) = 0.4, r_x(2) = 0.4, r_x(3) = 0.3, r_x(4) = 0.2, r_x(5) = 0.9, r_x(6) = 0.4$$

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

b) Use the least squares approach to find the unknown coefficients which minimize the modelling error for the signal described above, when the system model is given by:

i) An autoregressive model [4]

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

ii) An all-pole model of the form

$$H(z) = \frac{1}{1 + a_N z^{-N}}$$

where both a_N and N are considered to be the model parameters. [4]
(Hint: use either the Yule-Walker equations or the least squares method)

c) Let $x[n]$ be a process that is generated according to the difference equation

$$x[n+1] = a_1 x[n] + w[n]$$

where a_1 is a parameter, and $w[n] \sim \mathcal{N}(0, 1)$. An adaptive finite impulse response (FIR) filter is used for the prediction of this process.

i) Derive the expression for the Least Mean Square (LMS) update of a single coefficient adaptive FIR filter. [4]

ii) State the bound on the step size which ensures the convergence of such a filter, and the minimum mean square error achievable by using this FIR predictor. [4]

5) Comment on the advantages and disadvantages of adaptive filters as compared to the Wiener filter (with fixed filter weights) and explain the need for adaptive filtering in real world applications in statistically nonstationary environments. [3]

a) State real world applications of the inverse system modelling adaptive filtering configuration and draw a block diagram of this configuration. Comment on any delay introduced by this configuration and explain how this delay influences the so calculated inverse channel transfer function. [4]

b) Explain the problems arising in adaptive filters due to a fixed stepsize (both small and large). Describe in your own words the behaviour of an ideal adaptive step size in the Least Mean Square (LMS) setting. [2]

c) Consider linear adaptive Finite Impulse Response (FIR) filters with a time-varying learning rate (step size).

i) Derive the time-varying stepsize η_{NLMS} of the normalised LMS (NLMS) algorithm by expanding the output error $e(k+1)$ of the LMS algorithm using Taylor series expansion around $e(k)$. [5]

ii) Give physical interpretation of the behaviour of η_{NLMS}, and justify its use for signals with large time-varying dynamics. [2]

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

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + a_3 x(n-3) + w(n), \quad w(n) \sim \mathcal{N}(0,1)$$

Write down the output $\hat{x}(n)$ of a three–coefficient LMS–type adaptive predictor for this process and the expression for the NLMS weight updates of such an adaptive predictor. [4]

E E 3 - 08

Advanced Signal Processing

Solutions 2011

1) **[Bookwork and practical application of bookwork]**

a) Bias $B = E\{\hat{\theta}\} - \theta$.

If the bias is zero, then the expected value of the estimate is equal to the true value, i.e. $E\{\hat{\theta}\} = \theta$ and the estimator is said to be unbiased.

For an estimate to be meaningful, it is necessary that $var \rightarrow 0 \quad as \quad N \rightarrow \infty$ or in other words

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

If the estimator $\hat{\theta}_N$ is unbiased, that is $E\{\hat{\theta}_N\} = \theta$, and it satisfies the Cramer-Rao lower bound (CRLB) then it is a Minimum Variance Unbiased (MVU) estimator.

b) **[bookwork and new example]**

$$\mu_N = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = \frac{1}{N}\sum_{n=0}^{N-1} \underbrace{E\{x[n]\}}_{\mu} = \frac{1}{N}N\mu = \mu$$

For a zero mean sequence

$$Var\{\mu_N\} \;=\; Var\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = \frac{1}{N^2}\sum_{n=0}^{N-1} \underbrace{Var\{x[n]\}}_{\sigma^2} = \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}$$

Notice the variance $\rightarrow 0$ and N $\rightarrow \infty$ $\quad\Rightarrow$ consistent estimator.

c) **[bookwork and new example]**

Mean Squared Error (MSE) is given by $MSE(\hat{\theta}) = E\left\{\left(\hat{\theta} - \theta\right)^2\right\}$, and measures the average mean squared deviation of the estimator from the true value.

This criterion leads, however, to unrealisable estimators - namely, ones which are not solely a function of the data.

$$\begin{aligned} MSE(\hat{\theta}) &= E\left\{\left[\left(\hat{\theta} - E(\hat{\theta})\right) + \left(E(\hat{\theta}) - \theta\right)\right]^2\right\} \\ &= Var(\hat{\theta}) + E\left\{E(\hat{\theta}) - \theta\right\}^2 = Var(\hat{\theta}) + B^2(\hat{\theta}) \end{aligned}$$

$\Rightarrow$ MSE = VARIANCE OF THE ESTIMATOR + SQUARED BIAS

This criterion guarantees no optimality and an MVU estimator is in general not achieved, as there is a squared bias term contributing to the total error power.

d) **[new example]**

From

$$E\{\sigma^2\} = \alpha\sigma^2$$

we have

$$\begin{aligned}
\text{MSE}(\hat{\sigma}^2) &= E\{(\hat{\sigma}^2 - \sigma^2)^2\} = E\{\hat{\sigma}^4\} + \sigma^4(1 - 2\alpha) \\
&= \frac{\alpha^2}{N^2} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} E\{x^2[t]x^2[s]\} + \sigma^4(1 - 2\alpha) \\
&= \frac{\alpha^2}{N^2} \left(N^2\sigma^4 + 2N\sigma^4\right) + \sigma^4(1 - 2\alpha) \\
&= \sigma^4 \left[\alpha^2(1 + 2/N) + (1 - 2\alpha)\right]
\end{aligned}$$

It is straightforward to show that the minimum value is obtained for

$$\alpha_{min} = \frac{N}{N+2}$$

giving the minimum MSE in the form

$$\min \text{MSE}(\hat{\sigma}^2) = \frac{2\sigma^4}{\text{N} + 2}$$

This estimator is BIASED (see lecture notes), however it achieves a lower variance than the CRLB (which guarantees an MVU estimator). This is a good example of the bias-variance dilemma, where we can allow for a certain amount of bias if we desire low variance of an estimator. The class of MSE estimators in general does not guarantee any optimality.

2) a) **[bookwork and new examples]**
An autoregressive (AR) process of order p, that is $AR(p)$ is given by $x[n] = \sum_{i=1}^{p} a_i x[n-i] + w[n]$

where $x[n]$ is the output of the model and $w[n]$ are samples of zero mean white Gaussian noise with variance σ^2 ($w[n] \sim \mathcal{N}(0, \sigma^2)$).

Since the autocorrelation function is a function of correlation lag k, we need to calculate $E\{x[n]x[n-k]\}$. TO achieve this, first evaluate the product

$$\begin{aligned}
x[n-k]x[n] = a_1 x[n-k]x[n-1] + a_2 x[n-k]x[n-2] + \cdots \\
+ a_p x[n-k]x[n-p] + x[n-k]w[n]
\end{aligned}$$

Notice that $E\{x[n-k]w[n]\}$ vanishes when $k > 0$, since the driving WGN is not correlated with $x[n]$ for $k > 0$. For the correlation function function we therefore have

$$\begin{aligned}
r_{xx}(k) &= a_1 r_{xx}(k-1) + a_2 r_{xx}(k-2) + \cdots + a_p r_{xx}(k-p) \quad k > 0 \quad \textit{and} \\
r_{xx}(0) &= a_1 r_{xx}(1) + a_2 r_{xx}(2) + \cdots + a_p r_{xx}(p) + \sigma_w^2 \quad \text{for k} = 0
\end{aligned}$$

The general expression for the power spectrum of ARMA models is (follows by applying the z transform to the time domain expression for ARMA models)

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

From this expression and for $B = 1$ we obtain the expression for the power spectrum of an $AR(p)$ process

$$P_{xx}(f) = \frac{2\sigma_w^2}{|1 - a_1 e^{-j2\pi f} - \cdots - a_p e^{-j2\pi p f}|^2} \qquad 0 \leq f \leq 1/2$$

b) i) .ii) and iii) [**bookwork, coursework and intuitive reasoning**]
This is an $AR(4)$ model, since the spectrum has two peaks, and every pair of conjugate complex poles of the transfer function of an AR model gives rise to one peak in spectrum. Since we have 4 peaks, these peaks are generated by 2 conjugate complex pairs of poles. This cannot be the frequency response of an MA spectrum since there are no finite zeros in the spectrum, and the MA model gives rise to zeros and cannot generate sharp peaks. If we overmodelled by the factor of 2, we would assume an AR(8) process to generate this spectrum. This is in principle perfectly acceptable, however, it is likely that we would have some spurious peaks in the spectrum due to the excess number of pole pairs. For relatively short data sequences and imprecise autocorrelation function values, we may experience so called spectral line splitting, where ever peak would have another 'ghost' peak next to it.

c) [**new example and bookwork**]
i) For the MA(q) process

$$x[n] = b_1 w[n-1] + \cdots + b_q w[n-q] + w[n]$$

the variance is given by

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

The ACF is finite in duration and has a length q. It is calculated from

$$r_k = E\left[\left(w[n] + b_1 w[n-1] + \cdots + b_q w[n-q]\right)\left(w[n-k] + b_1 w[n-k-1] + \cdots + b_q w[n-k-q]\right)\right]$$

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

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

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.

3) a) [**bookwork**]
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

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

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.

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 $\frac{1}{I(\theta)}$.

b) **[new example]**
From

$$lnp(\mathbf{x};\boldsymbol{\theta}) = -\frac{N}{2}ln2\pi - \frac{N}{2}ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

we have

$$\begin{aligned}
\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial A} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) \\
\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}(x[n]-A)^2 \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2} &= -\frac{N}{\sigma^2} \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial A \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{n=0}^{N-1}(x[n]-A) \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial {\sigma^2}^2} &= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{n=0}^{N-1}(x[n]-A)^2
\end{aligned}$$

Upon taking the negative expectations, we arrive at the desired result.
c) **[new example]**
From

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}}e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A-Bn)^2}$$

we have

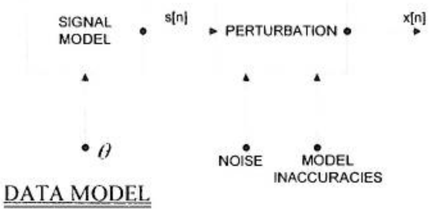
$$\begin{aligned}
\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial A} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A-Bn) \\
\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial B} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A-Bn)n \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2} &= -\frac{N}{\sigma^2} \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial A \partial B} &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n \\
\frac{\partial^2 p(\mathbf{x};\boldsymbol{\theta})}{\partial B^2} &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n^2
\end{aligned}$$

We can now use the identities given in the Hint to obtain the desired Fisher information matrix.

d) **[new example]**

The CRLB for A when $B = 0$ is σ^2/N, this smaller than in the case of estimating jointly A and B. In general the CRLB always increases as we estimate more parameters.

4) a) **[bookwork]** • Signal, $s[n]$ is assumed to be generated by the signal model



DATA MODEL

Figure 1: Data model for least squares estimation

which is a function of θ

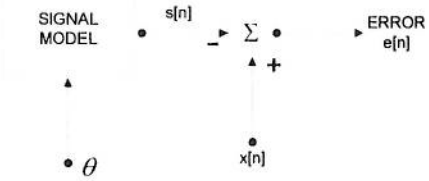
• The observation noise/model inaccuracies perturb s[n] to yield the measurement $x[n]$

• 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) and ii) **new examples**

i) Note that we normally use the autoregressive model of the form

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + w(n) \qquad \Rightarrow \qquad H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

so our usual AR coefficients are the same in the original transfer function, just with the negative sign. The Yule-Walker equations give

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \end{bmatrix}$$

and by inserting the values of the autocorrelation we have

$$\begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$$

and thus $a_1 = a_2 = -2/7$, giving the modelling error

$$\varepsilon = r_x(0) - a_1 r_x(1) - a_2 r_x(2) = 1.2286$$

ii) In a similar way as in i), we from the solution for a_N (see also the AR(1) model) is

$$a_N = \frac{r_x(N)}{r_x(0)}$$

and the minimum mean square error (MMSE) becomes

$$\varepsilon_{min} = \frac{r_x^2(0) - r_x^2(N)}{r_x(0)}$$

and has a minimum for $N = 5$. In other words, to minimise the error we want to find the value of N for which $x(n)$ and $x(n+N)$ have the highest correlation.

c) i) [**application of bookwork and coursework**]

i) Here, we need to find adaptively the value of the unknown coefficient a which generates this process. This will be achieved if the instantaneous output error of the adaptive filter is white (in this case the "error" is $w[n]$), and we minimise the instantaneous estimate of the error power $E(n) = \frac{1}{2}e^2(n)$. Following the standard LMS update

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

where $\mathbf{w}$ are filter coefficients, μ learning rate, $e(n)$ is the instantaneous output error and $\mathbf{x}(n)$ the input signal in filter memory, we have

$$\begin{aligned} \hat{x}[n+1] &= a_1[n].\hat{x}[n] \\ a_1[n+1] &= a_1[n] + \mu e[n] x[n] \end{aligned}$$

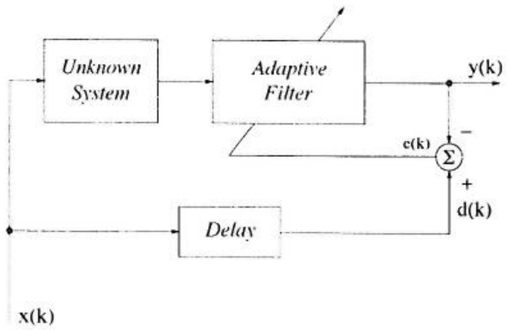
ii) **[applied bookwork]** Same as for standard LMS with length $N = 1$.

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, where 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]**

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

9

Inserting the partial derivatives from the above, we arrive at

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

From there the NLMS step size which minimizes the error is

$$\eta_{NLMS} = \frac{1}{\parallel \mathbf{x}(n) \parallel_2^2}$$

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**]

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

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