E4.13
AS2
SO15
ISE4.31

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2005

MSc and EEE/ISE PART IV: MEng and ACGI

SPECTRAL ESTIMATION AND ADAPTIVE SIGNAL PROCESSING

Tuesday, 26 April 10:00 am

Time allowed: 3:00 hours

Corrected Copy

There are FIVE questions on this paper.

Corrections Page 4 start of exam

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

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
Second Marker(s) : R. Nabar

© University of London 2005

1) Consider the problem of periodogram based spectrum estimation.

a) Write down the expression for the periodogram. [1]

i) Show how the periodogram can be computed from the discrete Fourier transform of a signal. [2]

ii) Compute the bias of the periodogram. Is the periodogram biased? [3]

iii) What is the variance of the periodogram proportional to? [2]

iv) Comment on the performance of the periodogram. [2]

b) A random process is known to consist of a single sinusoid with frequency ω_0 in white noise $w(n)$, given by

$$x(n) = Asin(n\omega_0 + \phi) + w(n)$$

where the variance of $w(n)$ is σ_w^2.

i) Suppose that the first three values of the autocorrelation sequence are estimated and found to be

$$r_x(0) = 1 \quad r_x(1) = \beta \quad r_x(2) = 0$$

Find and prepare a carefully labeled sketch of the spectrum estimate that is formed using the Blackman–Tukey method with a rectangular window. [4]

ii) Suppose that we compute the periodogram $\hat{P}_{per}(e^{j\omega})$ using N samples of $x(n)$ (Bartlett method). Find and prepare a carefully labeled sketch of the expected value of this spectrum estimate. Is this estimate biased? Is it consistent? [6]

2) Consider the problem of parametric spectrum estimation.

a) Explain the limitations of classical periodogram based methods and the need for parametric spectrum estimation techniques. [3]

b) Define the autoregressive (AR) spectrum estimation, and state the expression of a general AR power spectrum. [2]

i) Describe two typical AR modelling techniques. [2]

ii) What is the philosophy behind AR model order selection? [2]

c) Define the moving average (MA) spectrum estimation and state the expression for a general MA power spectrum. [2]

i) Explain the differences between the AR and MA spectrum. Which of the two is more suitable for estimating spectra with peaks? [2]

ii) By visual inspection or otherwise, conclude whether the spectrum given in the Figure below is most likely to be an AR, MA, or ARMA spectrum. Explain. [5]

iii) What are the limitations of parametric spectrum estimation techniques? [2]

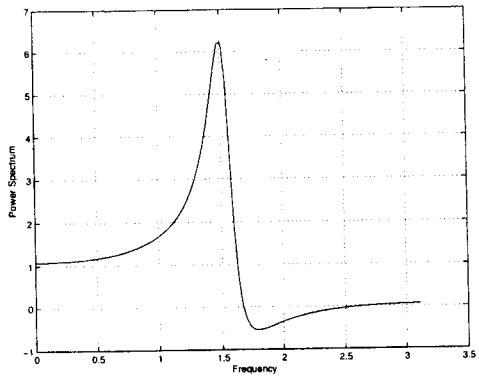


Figure 1: Power spectrum (in dB)

3) Consider adaptive filters with feedback.

a) Sketch the block diagram of an adaptive infinite impulse response (IIR) filter. [2]

b) For adaptive IIR filters, what is the difference between the equation error and output error mode of training? Does the equation error learning, strictly speaking, reflect the nature of this feedback structure? [3]

c) Derive the learning algorithm for an adaptive IIR filter in the output error mode. [7]

d) Explain the need for nonlinear feedback filters. [4]

i) Explain the difference between an IIR filter and recurrent perceptron. [2]

ii) Sketch the block diagram of a recurrent perceptron which realises the following difference equation

$$y(n) = \Phi\big(0.1y(n-1) + 0.8x(n-1) + 0.2x(n-2) + 0.9\big)$$

where Φ is some saturating nonlinear function. [2]

4)

a) Explain the notions of parametric, nonparametric and semi-parametric modelling. [3]

b) Explain the modes of operation of filtering, smooting and prediction. [3]

i) What operation is performed by a filter given by

$$y(n) = \big[x(n-2) + x(n-1) + x(n+1) + x(n+2)\big]/4$$

[2]

ii) What operation is performed by a filter given by

$$y(n) = \big[x(n) + x(n-1) + x(n-2) + x(n-3)\big]/4$$

[2]

c) State the problem of general sequential state estimation. Write down the equations that describe such a model, and explain the variables and functions involved. [2]

d) Write down the equation of the vector Gauss–Markov model. [3]

i) What is the form of the state transition matrix? [2] announced at start of exam

ii) Sketch the block diagram of Kalman filter. Explain the signal flow in such a structure. [2]

iii) What is the criterion of optimality for such a filter? [1]

5) The weight update of the Least Mean Square (LMS) algorithm for linear finite impulse response (FIR) adaptive filters is given by

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

where $\mathbf{w}(k)$ is the weight (coefficient) vector at time instant k, $\mathbf{x}(k)$ is the input vector, μ is the learning rate, and $e(k)$ is the instantaneous output error of the filter.

a) One modification of this algorithm is the so–called data reusing (DR) LMS. Write down the weight update for this algorithm, for $i = 1, \ldots, L$ data reusing iterations. (Hint: for a time instant k we reiterate weight updates $\mathbf{w}_i(k)$, where $\mathbf{w}_1(k) = \mathbf{w}(k)$ and $\mathbf{w}_{L+1}(k) = \mathbf{w}(k+1)$ for a fixed external input vector $\mathbf{x}(k)$). [4]

i) What is the motivation for using this algoritm? [2]

ii) Give a geometric explanation of the data reusing method. [2]

iii) Describe the relation between the DR algorithm and the NLMS algorithm. Where does $\mathbf{w}_{i+1}(k)$ converge when $i \to \infty$? [1]

iv) Calculate the total weight update for L data reusing iterations for a fixed time instant k. [1]

b) Derive the DR algorithm for a simple nonlinear FIR filter (dynamical perceptron) given in the Figure 2. Nonlinearity Φ is of a saturating type. [4]

i) In your own words, give a geometric interpretation in this case. [2]

ii) Is the algorithm more likely to converge for $\Phi' > 0$ or for $\Phi' < 0$? [2]

iii) Describe the relation between the step size μ and the slope of the nonlinearity Φ' so that such an algorithm converges. [2]

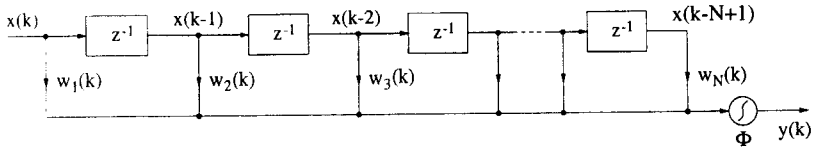


Figure 2: Nonlinear FIR filter

E4-13
Spectral Estimation & Adaptive Signal Processing ISE4 31
2005 SO15
AS2

Solutions:

i) a) $P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega}$ [1]

i) $x_N \rightarrow X_N(k) \rightarrow \frac{1}{N}|X_N(k)|^2 = \hat{P}_{per}(e^{j2\pi k/N})$ [2]

ii) $E\{r_x(k) = \frac{N-k}{N} r_x(k)\} \Rightarrow E\{\hat{P}_{per}(e^{j\omega}\} = \frac{2}{\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})$ [3]

where W_B is the Bartlett window. Therefore the periodogram is a biased estimated, but since W_B convergs to impulse as N goes to infinity, it is asymptotically unbiased.

iii) $var\{\hat{P}_{per}(e^{j\omega})\} = P_x^2(e^{j\omega})$ [2] + [2]

b) i) $\hat{P}_x(e^{j\omega}) = \sum_{k=-M}^{M} r_x(k)e^{-j\omega} = 1 + 2\beta cos\omega$ [4]

ii)

$$E\{\hat{P}_{per}(e^{j\omega}\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})$$

$$W_B(e^{j\omega}) = \frac{1}{N}\left[\frac{sin(N\omega/2)}{sin(\omega/2)}\right]^2$$

$$P_x(e^{j\omega}) = 1/2\pi A^2[u_0(\omega - \omega_0) + u_0(\omega + \omega_0)] + \sigma_w^2$$

$$\Rightarrow E\{\hat{P}_{per}(e^{j\omega}\} = \frac{2}{\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) =$$

$$= \sigma_w^2 + 1/4A^2[W_B(e^{j(\omega-\omega_0)}) + W_B(e^{j(\omega+\omega_0)}]$$

The periodogam is biased and since the variance does not go to zero as $N \rightarrow \infty$ it is not a consistent estimate ot the power spectrum. [6]

1/7

2) a) They are not consistent estimators, they give poor resolution and do not work well on short data records. They are also limited in their ability to resolve closely spaced narrowbad processes when the number of data saples is limited. An adantage: they do not make any assumptions or place any constraints on the process and can be used for any type of processes. [3]

b) ARMA spectrum is given by [2]

$$\hat{P}_x(e^{\jmath\omega}) = \frac{\left|\sum_{k=0}^{q} \hat{b}_q(k) e^{-\jmath k\omega}\right|^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_p(k) e^{-\jmath k\omega}\right|^2}$$

For the AR spectrum, we have only the denominator of the above expression, whereas for the MA spectrum the denominator of the above expression is constant.

i) The autocorrelation and covariance method. The difference is in the way they estimate the ACF. The artifact of the autocorrelation method is spectral line splitting. In the covariance method, no data windowing is needed. Produces higher resolution for short data record than the autocorrelation method. [2]

ii) To penalise for the model order. Tha MDL and AIC critera both have two terms [2]

$$\text{model order} = \min e^2 + \text{penalty for complexity}$$

c) See b) [2]

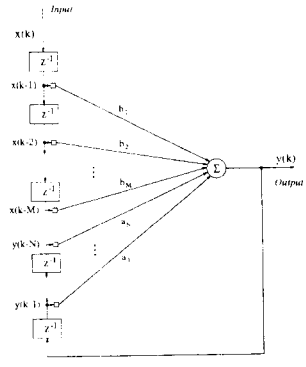
i) The AR spectrum is more suitable for modelling peaks in spectrum (has poles in transfer function). The MA spectrum is more suitable for modelling flat spectra and spectra with zeros in transfer function. [2]

ii) The spectrum in the Figure is most likely to be an ARMA spectrum, since it has a peak and a zero. In fact it is an ARMA(2,2) process with the zeros of $H(z)$ at $z = 0.95e^{\pm\jmath\pi/2}$ and the poles at $z = 0.9e^{\pm\jmath 2\pi/5}$. [5]

iii) The selection of model order is a critical step. It often requires some a priori knowledge. The problem is that is is of crucial importance that the model that is used is consistend with the process being analysed. [2]

7/7

3) a)



[2]

b) Two classes of adaptive learning algorithms

[3]

- **equation error** – the desired signal $d(k)$ is fed back
- **output error**– estimated outputs $\hat{y}(k)$ fed back.

c) Define the gradient $\nabla_{\Theta}(E(k))$ for cost function $E(k) = \frac{1}{2}e^2(k)$ as

[7]

$$\nabla_{\Theta}(E(k)) = \frac{\partial E(k)}{\partial \Theta(k)} = e_{OE}(k)\nabla_{\Theta} e_{OE}(k) = -e_{OE}(k)\nabla_{\Theta} y_{OE}(k) \qquad (1)$$

where $\Theta(k) = [b_1(k), \ldots, b_M(k), a_1(k), \ldots, a_N(k)]^T$. The gradient vector consists of partial derivatives of the output with respect to filter coefficients

$$\nabla_{\Theta} y_{OE}(k) = \left[\frac{\partial y_{OE}(k)}{\partial b_1(k)}, \ldots, \frac{\partial y_{OE}(k)}{\partial b_M(k)}, \frac{\partial y_{OE}(k)}{\partial a_1(k)}, \ldots, \frac{\partial y_{OE}(k)}{\partial a_N(k)}\right]^T \qquad (2)$$

Take the derivatives of both sides of (??) wrt $a_i(k)$ and $b_j(k)$

$$\begin{aligned} \frac{\partial y_{OE}(k)}{\partial a_i(k)} &= y_{OE}(k-i) + \sum_{m=1}^{N} a_m(k)\frac{\partial y_{OE}(k-m)}{\partial a_i(k)} \\ \frac{\partial y_{OE}(k)}{\partial b_j(k)} &= x(k-j) + \sum_{m=1}^{N} a_m(k)\frac{\partial y_{OE}(k-m)}{\partial b_j(k)} \end{aligned} \qquad (3)$$

- partial derivatives are wrt current values of $a_m(k)$ and $b_m(k)$, $\Rightarrow$ (3) non–recursive
- If Θ were independent of $\{y(k-i)\}$ calculation identical to the FIR case
- We have to use the *pseudolinear regression algorithm*

$$\Theta(k) \approx \Theta(k-1) \approx \cdots \approx \Theta(k-N) \qquad (4)$$

The previous approximation is particularly good for N small.

$$\begin{aligned}\frac{\partial y_{OE}(k)}{\partial a_i(k)} &\approx y_{OE}(k-i) + \sum_{m=1}^{N} a_m(k) \frac{\partial y_{OE}(k-m)}{\partial a_i(k-m)} \\ \frac{\partial y_{OE}(k)}{\partial b_j(k)} &\approx x(k-j) + \sum_{m=1}^{N} a_m(k) \frac{\partial y_{OE}(k-m)}{\partial b_j(k-m)} \end{aligned} \tag{5}$$

This admits computation in a recursive fashion. For compactness introduce the weight vector $\mathbf{w}(k)$ as

$$\mathbf{w}(k) = [b_1(k), b_2(k), \ldots, b_M(k), a_1(k), a_2(k), \ldots, a_N(k)]^T \tag{6}$$

and the IIR filter input vector $\mathbf{u}(k)$ as

$$\mathbf{u}(k) = [x(k-1), \ldots, x(k-M), y(k-1), \ldots, y(k-N)]^T \tag{7}$$

Thus, $w_2(k) = b_2(k)$, $w_{M+2}(k) = a_2(k)$, $u_M(k) = x(k-M)$, or $u_{M+1}(k) = y(k-1)$, and

$$\frac{\partial y_{OE}(k)}{\partial w_i(k)} \approx u_i(k) + \sum_{m=1}^{N} w_{m+M}(k) \frac{\partial y_{OE}(k-m)}{\partial w_i(k-m)} \tag{8}$$

Denote $\pi_i(k) = \frac{\partial y_{OE}(k)}{\partial w_i(k)}$, $i = 1, \ldots, M+N$, to yield

$$\pi_i(k) \approx u_i(k) + \sum_{m=1}^{N} w_{m+M}(k) \pi_i(k-m) \tag{9}$$

Finally :-) weight update equation for a linear IIR adaptive filter

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta(k) e(k) \boldsymbol{\pi}(k), \quad \boldsymbol{\pi}(k) = [\pi_1(k), \ldots, \pi_{M+N}(k)]^T \tag{10}$$

d) There are numerous situations in which the use of linear filters and models is limited:- [4]

- when trying to identify a saturation type nonlinearity, linear models will inevitably fail
- when separating signals with overlapping spectral components
- systems rich dynamical behaviour such as limit cycles, bifurcations and fixed points, cannot be captured by linear models
- communications channels, for instance, often need nonlinear equalizers to achieve acceptable performance

i) recurrent perceptron =IIR filter with an additional output saturating nonlinearity [2]

ii) As in a) with the feedback weight 0.1, feedforward weights 0.8 and 0.2 and bias weight 0.9. [2]

5/7

[3]

4) a)

- **Parametric** modelling assumes a fixed structure for the model. The model identification problem then simplifies to estimating a finite set of parameters of this fixed model. An example of this technique is the broad class of ARIMA/NARMA models.
- **Nonparametric** modelling seeks a particular model structure from the input data. The actual model is not known beforehand.
  We look for a model in the form of $y(k) = f(x(k))$ without knowing the function $f(\cdot)$.
- **Semiparametric** modelling is the combination of the above. Part of the model structure is completely specified and known beforehand, whereas the other part of the model is either not known or loosely specified.

[3]

b) Filtering: $\theta = s[n]$ to be estimated based on $x[m] = s[m]+w[m]$ $m = 0, 1, ..., n$. Filters signal from noise, based on current and past data, i.e. casual filtering.
Smoothing: $\theta = s[n]$ to be estimated based on entire dataset $\{x[0], x[1], ..., x[N-1]\}$. Requires all data to be collected.
Prediction: $\theta = x[N-1+l]$ for l a positive integer based on $\{x[0], x[1], ..., x[N-1]\}$ "l-step" forward prediction.

[1]

i) it is smoothing, $y(n)$ is estimated based on previous an future values of y.

ii) filtering

[2]

c)

[2]

$$\begin{aligned} x(k+1) &= f(x(k),k) + G(x(k),k)w(k) \\ y(k) &= h(x(k),k) + v(k) \end{aligned}$$

- $x(k), y(k)$ are N_x and N_y dimensional stochastic processes
- $d(k)$ and $v(k)$ are noise processes
- $f(\cdot)$, $G(\cdot)$, $h(\cdot)$ – known functions

[3]

d)

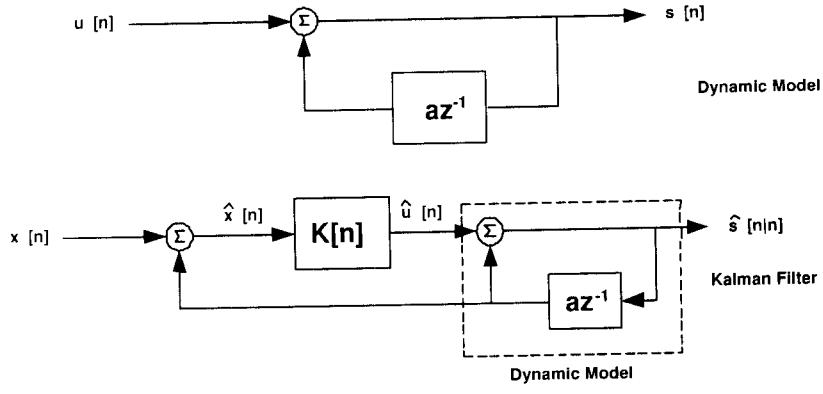
- State vector

$$\begin{bmatrix} s[n-p+1] \\ s[n-p+2] \\ \vdots \\ s[n-1] \\ s[n] \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \dots & 1 \\ -a(p) & -a(p-1) & \dots & -a(1) \end{bmatrix}}_{\textit{State transition matrix } A} \begin{bmatrix} s[n-p] \\ s[n-p+1] \\ \vdots \\ s[n-2] \\ s[n-1] \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{B} u[n]$$

- $u[n]$ can be a $r \times 1$ vector - models a vector signal output with vector input

i) it is the companion matrix given above [1]

ii)



[2]

iii) Criterion of optimality

$$\underbrace{E\left\{(s[n]-\hat{s}[n|m])^2\right\}}_{\text{with respect to } p(x[0],x[1],\ldots,x[n],s[n])}$$

[1]

7/7

5) a) [4]

$$e_i(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}_i(k)$$
$$\mathbf{w}_{i+1}(k) = \mathbf{w}_i(k) + \eta e_i(k)\mathbf{x}(k)$$

subject to

$$|e_{i+1}(k)| \leq \gamma |e_i(k)|, \ 0 < \gamma < 1, \ i = 1, \ldots, L$$

[2]

i) To "correct" the noisy gradients using a posteriori procedures [2]
ii) It iteratively updates the weights in the direction of the data vector. In terms of the error, it gradually decreases towards the NLMS error.
iii) DR algorithms approaches the NLMS algorithm when the number of DR iterations is large. [1]
iv) $\sum_{i=1}^{L} \Delta \mathbf{w}_i(n)$. [1]

b) [4]

$$\begin{aligned} e_i(k) &= y(k) - \Phi(\mathbf{x}^T(k)\mathbf{w}_i(k) \\ \mathbf{w}_{i+1}(k) &= \mathbf{w}_i(k) + \mu e_i(k)\Phi'(\mathbf{x}^T(k)\mathbf{w}_i(k))\mathbf{x}(k) \end{aligned}$$

[2]

i) similar to the interpretation for the linear function, but the slope Φ' also plays important role [2]
ii) For a linear algorithm, $\Phi' = 0$. If Φ' is large, then the algorithm may diverge (not contractive function), if $\Phi' < 0$ the algorithm is always stable. [2]
iii) The effective step size above becomes $\mu \times \Phi'$, and this should be taken into account when considering stability.