ENGINEERING TRIPOS PART IIA 2004

Solutions to Module 3F3
Signal and Pattern Processing
Principal Assessor: Dr S J Godsill
Second Assessor: Professor W J Fitzgerald

1 (a) Use the matched z-transform to convert the analogue filter with transfer function

$$H(s) = \frac{s+0.2}{(s+0.2)^2+4}$$

into a digital IIR filter. Select $T = 0.2$ and compare the location of the zeros in $H(z)$ with the locations of the zeros obtained by applying the impulse invariance method in the conversion of $H(s)$. [40%]

Answer. $H(s)$ has one zero $z_1 = -0.2$ and two poles $p_{1,2} = -0.2 \pm 2j$. The matched-z transform maps these into

$$\begin{aligned} \widetilde{z}_1 &= e^{-0.2T} = e^{-0.04} = 0.9608, \\ \widetilde{p}_1 &= e^{(-0.2+2j)T} = 0.9608e^{j0.4}, \\ \widetilde{p}_2 &= 0.9608e^{-j0.4}. \end{aligned}$$

Hence one has

$$\begin{aligned} H(z) &= \frac{\left(1-\widetilde{z}_1 z^{-1}\right)}{\left(1-\widetilde{p}_1 z^{-1}\right)\left(1-\widetilde{p}_2 z^{-1}\right)} \\ &= \frac{1-rz^{-1}}{1-2r\cos(\omega_0)z^{-1}+r^2z^{-2}} \end{aligned}$$

where $r = 0.9608$ and $\omega_0 = 0.4$.

For the impulse invariance method, one has

$$H(s) = \frac{1}{2}\left[\frac{1}{s+0.2-2j}+\frac{1}{s+0.2+2j}\right].$$

Thus

$$\begin{aligned} H(z) &= \frac{1}{2}\left[\frac{1}{1-e^{-0.2T}e^{j2T}z^{-1}}+\frac{1}{1-e^{-0.2T}e^{-j2T}z^{-1}}\right] \\ &= \frac{1-r\cos(\omega_0)z^{-1}}{1-2r\cos(\omega_0)z^{-1}+r^2z^{-2}}. \end{aligned}$$

It follows that the poles are the same but the zero is different.

(b) Explain the objectives and application of the bilinear transform

$$s = \frac{1-z^{-1}}{1+z^{-1}}.$$

What is the main drawback of the bilinear transform? [20%]

Answer. The objectives are

(cont.

•to map stable analogue filters to stable digital filters.

•map $s = j\Omega$ to $z = e^{j\omega}$ monotonically so as to preserve filter class (e.g. passband, lowpass etc.).

It is applied to design digital filters based on analogue filters. For example to design a digital lowpass filter one maps the (normalized) cutoff frequence $\omega_c \to \Omega_c = \tan\left(\frac{\omega_c}{2}\right)$, then design the appropriate analogue filter and finally replace s by $\left(1 - z^{-1}\right) / \left(1 + z^{-1}\right)$.

•The main problem with the bilinear transform is that it performs a nonlinear mapping of the phase leading to a distortion (or warping) of the digital frequency response. This effect is normally compensated for by prewarping the analogue filter before applying the bilinear transformation.

(c) It is required to design a lowpass digital filter by approximating the following analogue transfer function

$$H(s) = \frac{1}{s^2 + \sqrt{2}s + 1}.$$

Using the bilinear transform, obtain the transfer function $H(z)$ of the digital filter assuming a 3dB cutoff frequency of 150Hz and a sampling frequency of 1.28kHz. Sketch an implementation showing the required multiplier coefficients. [40%]

Answer. The normalized frequency is given by

$$\omega_c = \frac{150}{1280} 2\pi = 0.7363 \Rightarrow \Omega_c = 0.3857.$$

The prewarped analogue filter is given by

$$\begin{aligned} H'(s) &= H(s)|_{s=s/\Omega_c} = \frac{1}{(s/\Omega_c)^2 + \sqrt{2}(s/\Omega_c) + 1} \\ &= \frac{\Omega_c^2}{s^2 + \sqrt{2}\Omega_c s + \Omega_c^2} = \frac{0.1488}{s^2 + 0.5455s + 0.1488}. \end{aligned}$$

By applying the bilinear transform, one obtains

$$\begin{aligned} H(z) &= \frac{0.878z^2 + 0.1756z + 0.0878}{z^2 - 1.0048z + 0.3561} \\ &= \frac{0.878\left(1 - 2z^{-1} + z^{-2}\right)}{1 - 1.0048z^{-1} + 0.3561z^{-2}}. \end{aligned}$$

(TURN OVER

2 (a) Compute the frequency response of the FIR filter given by

$$h(0) = 0.25, \;\; h(1) = 0.5, \;\; h(2) = 0.25$$

exploiting symmetry to express its frequency response as the product of a pure delay term and a frequency-dependent gain. [15%]

Answer:

$$\begin{aligned} H(e^{j\theta}) &= \sum_{i=0}^{2} h(n)e^{-jn\theta} \\ &= h(1)e^{-j\theta} + h(0)(e^{0} + e^{-j2\theta}) \\ &= e^{-j\theta}(h(1) + 2h(0)cos(\theta)) \end{aligned}$$

Pure delay term is $e^{-j\theta}$, frequency-dependent gain is $h(1) + 2h(0)cos(\theta)$.

(b) An FIR filter has an impulse response $h\,(n)$ which is defined over the interval $0 \leq n \leq N-1$. Show that if N is odd and $h\,(n)$ satisfies the positive symmetry condition, that is $h\,(n) = h\,(N-1-n)$, the filter has a linear phase response, and give an expression for the frequency response of the filter. [35%]

Answer.

Define the 'central' coefficient: $h(M)$ where $M = (N-1)/2$.

Then,

$$\begin{aligned} H\left(e^{j\omega}\right) &= \sum_{n=0}^{N-1} h\,(n)\, e^{-j\omega n} \\ &= h(M)e^{-j\omega M} + \sum_{n=0}^{M-1} h(n)(e^{-j\omega n} + e^{-j\omega(N-n-1)}) \;\; \text{[Since } h(n) = h(N-n-1)\text{]} \\ &= e^{-j\omega M}\left(h(M) + \sum_{n=0}^{M-1} h(n)(e^{-j\omega(n-M)} + e^{-j\omega((N-n-1)-M)})\right) \end{aligned}$$

Looking at exponents in summation: $n - M = n - N/2 + 1/2$, and $N - n - 1 - M = N - n - 1 - N/2 + 1/2 = N/2 - n - 1/2$. Hence the two complex exponenials in summation are conjugates of one another and:

$$H\left(e^{j\omega}\right) = e^{-j\omega M}\left(h(M) + 2\sum_{n=0}^{M-1} h(n)\cos(\omega(n-M))\right)$$

(cont.

We can then clearly see that the frequency response is a linear phase term and a real, frequency-dependent gain - hence filter is linear phase.

(c) A stationary random process $\{x_n\}$ has autocorrelation function:

$$r_{XX}[k] = 0.8^{|k|}$$

The process is measured in additive, independent white noise v_n with unit variance, so that the observations y_n are:

$$y_n = x_n + v_n \, .$$

It is desired to estimate the underlying signal values $\{x_n\}$ based on the observations alone. One sample of time delay is allowable in the estimate. A 3-tap FIR filter is to be designed for this task such that the estimated value of x_{n-1} is given by

$$\hat{x}_{n-1} = \sum_{i=0}^{2} h(n) y_{n-i} \, .$$

Obtain a simplified expression for mean-squared error of this estimator when the filter is constrained to have *linear phase response*. [15%]

Answer:

The linear phase filter has $h_0 = h_2$. Hence we can write:

$$\hat{x}_{n-1} = h(0)(y_n + y_{n-2}) + h(1)y_{n-1} = \tilde{\mathbf{h}}^T \tilde{\mathbf{y}}_n$$

with

$$\tilde{\mathbf{h}} = [h_0 \; h_1]^T \; \tilde{\mathbf{y}}_n = [(y_n + y_{n-2}) \; y_{n-1}]^T$$

Then, with $\epsilon_n = x_{n-1} - \hat{x}_{n-1}$

$$E = E[|\epsilon_n|^2] = E[(x_{n-1} - \tilde{\mathbf{h}}^T \tilde{\mathbf{y}}_n)^2]$$
$$= r_{XX}[0] + \tilde{\mathbf{h}}^T \mathbf{R}_{\tilde{Y}\tilde{Y}} \tilde{\mathbf{h}} - 2\tilde{\mathbf{h}}^T E[x_{n-1}\tilde{\mathbf{y}}_n]$$

Obtain the optimum mean-squared error filter coefficients, under the constraint of linear phase response. Determine the mean-squared error for this optimum filter. [35%]

(CONTINUED OVER.

We can then obtain:

$$\mathbf{R}_{\tilde{Y}\tilde{Y}} = \begin{bmatrix} E[(y_n + y_{n-2})^2] & E[y_{n-1}(y_n + y_{n-2})] \\ E[y_{n-1}(y_n + y_{n-2})] & E[y_{n-1}^2] \end{bmatrix}$$

$$= \begin{bmatrix} 2(r_{XX}[0] + r_{VV}[0]) + 2r_{XX}[2] & 2r_{XX}[1] \\ 2r_{XX}[1] & r_{XX}[0] + r_{VV}[0] \end{bmatrix} = \begin{bmatrix} 4 + 2 * 0.64 & 1.6 \\ 1.6 & 1 + 1 \end{bmatrix}$$

and

$$\mathbf{r}_{X\tilde{Y}} = E[x_{n-1}\tilde{\mathbf{y}}_n] = [2r_{XX}[1] \;\; r_{XX}[0]]^T$$

The optimal solution is then (Yule Walker equation):

$$\tilde{\mathbf{h}}_{opt} = \mathbf{R}_{\tilde{Y}\tilde{Y}}^{-1}\mathbf{r}_{X\tilde{Y}} = [0.2 \; 0.34]^T$$

i.e.

$$h(0) = 0.2, \quad h(1) = 0.34, \quad h(2) = 0.2$$

Finally the error at the optimal solution is

$$r_{XX}[0] - \tilde{\mathbf{h}}_{opt}^T\mathbf{r}_{X\tilde{Y}} = 0.34$$

3 A random process is defined as:

$$x_n = u_n - \alpha u_{n-1}$$

where u_n is white noise with variance equal to 1 and α is a constant.

(a) Determine the autocorrelation function for the process. What type of process is this? [30%]

Answer:

$$r_{XX}[0] = E[x_n^2] = E[u_n]^2 + \alpha^2 E[u_{n-1}^2] = 1 + \alpha^2$$

$$r_{XX}[1] = E[x_n x_{n+1}] = -\alpha E[u_n^2] = -\alpha$$

For lags k greater than 1 $r_{XX}[k] = 0$.

Negative lags obtained by $r_{XX}[-k] = r_{XX}[+k]$.

This is a moving average (MA) process.

(b) Is the process mean ergodic?

Answer:

The mean of the process is zero. Hence $c_{XX}[k] = r_{XX}[k]$ thus

$$\text{Lim}_{k\to\infty} c_{XX}[k] = 0$$

and it is therefore mean ergodic.

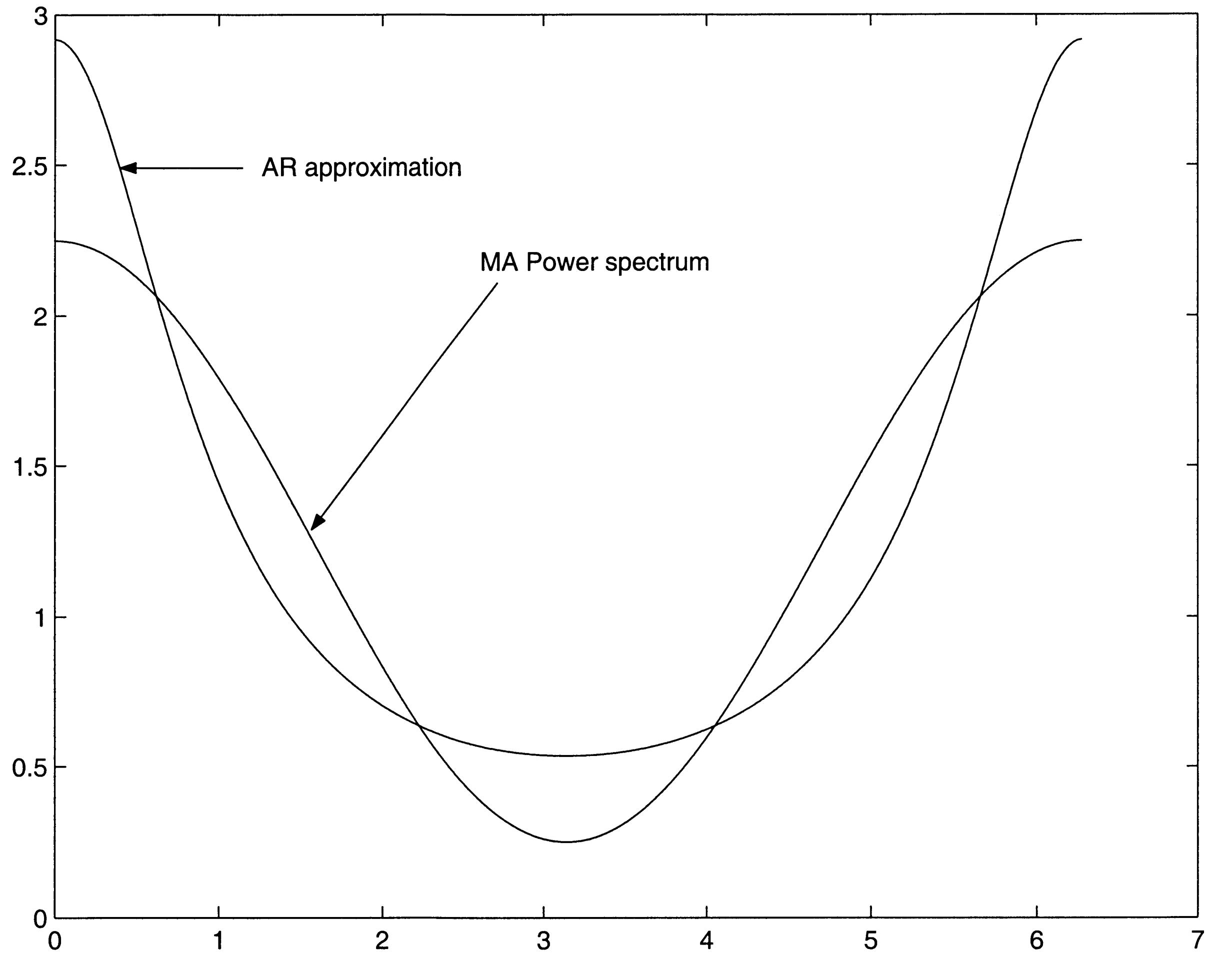
[10%]

(c) Determine and sketch the power spectrum of the process when $\alpha = 0.5$ [20%]

$$S_X(\exp(i\theta)) = |1 - \alpha \exp(-i\theta)|^2$$

At $\theta = 0$, takes value $(1-\alpha)^2$. At $\theta = \pi$, takes value $(1+\alpha)^2$. Hence power spectrum looks like:

(CONTINUED OVER.



(d) The process is now to be approximated as an autoregressive process. Use the autocorrelation function determined above to obtain a first order ($P = 1$) autoregressive model for this process. Sketch the power spectrum of this process for $\alpha = 0.5$ and comment on the quality of the approximation. Comment on whether this model is stable or not for all values of α . [30%]

Answer:

AR model is given by:

$$\mathbf{a} = -\mathbf{R}_{XX}^{-1}\mathbf{r}_{XX} = -r_{XX}[1]/r_{XX}[0] = \alpha/(1+\alpha^2)$$

$$\sigma_W^2 = r_{XX}[0] + r_{XX}[-1]a_1 = 1 + \alpha^2 - \alpha^2/(1+\alpha^2)$$

(cont.

The power spectrum is:

$$S_X(\exp(i\theta)) = \frac{\sigma_W^2}{|1 + a\exp(-i\theta)|^2}$$

With $\alpha = 0.5$ this is plotted above. Quite a good match considering how crude the approximation.

This AR model has a pole at $z = -a = -\alpha/(1+\alpha^2)$. Since $|a < 1|$ for all values of α , the process is unconditionally stable.

(TURN OVER

4 A two class classification problem is to be solved by building a linear decision boundary between the two classes.

(a) Contrast the training and classification performance on the training data of the linear decision boundaries that are generated using the *perceptron* algorithm and using *least mean squares estimation.* When might it be useful to use one of these approaches rather than Bayes' decision rule? [25%]

Answer: Answer should include

- Perceptron algorithm iterative, LMS is a single shot (but involves an inverse).
- For separable case perceptron algorithm with yield 100% correct performance. Otherwise will not converge. No guarantees for LMS, but can be used for non-separable data.
- Perceptron requires testing the training data, not required for LMS.

Bayes' decision rule is optimal when:

- the form of class-conditional distributions and prior are known;
- infinite training data is available;
- the global maximum is found.

Since these conditions are rarely satisfied it is usually reasonable to use these alternative approaches.

(b) The following set of feature vectors are to be used to build a linear classifier for a 2-dimensional problem,

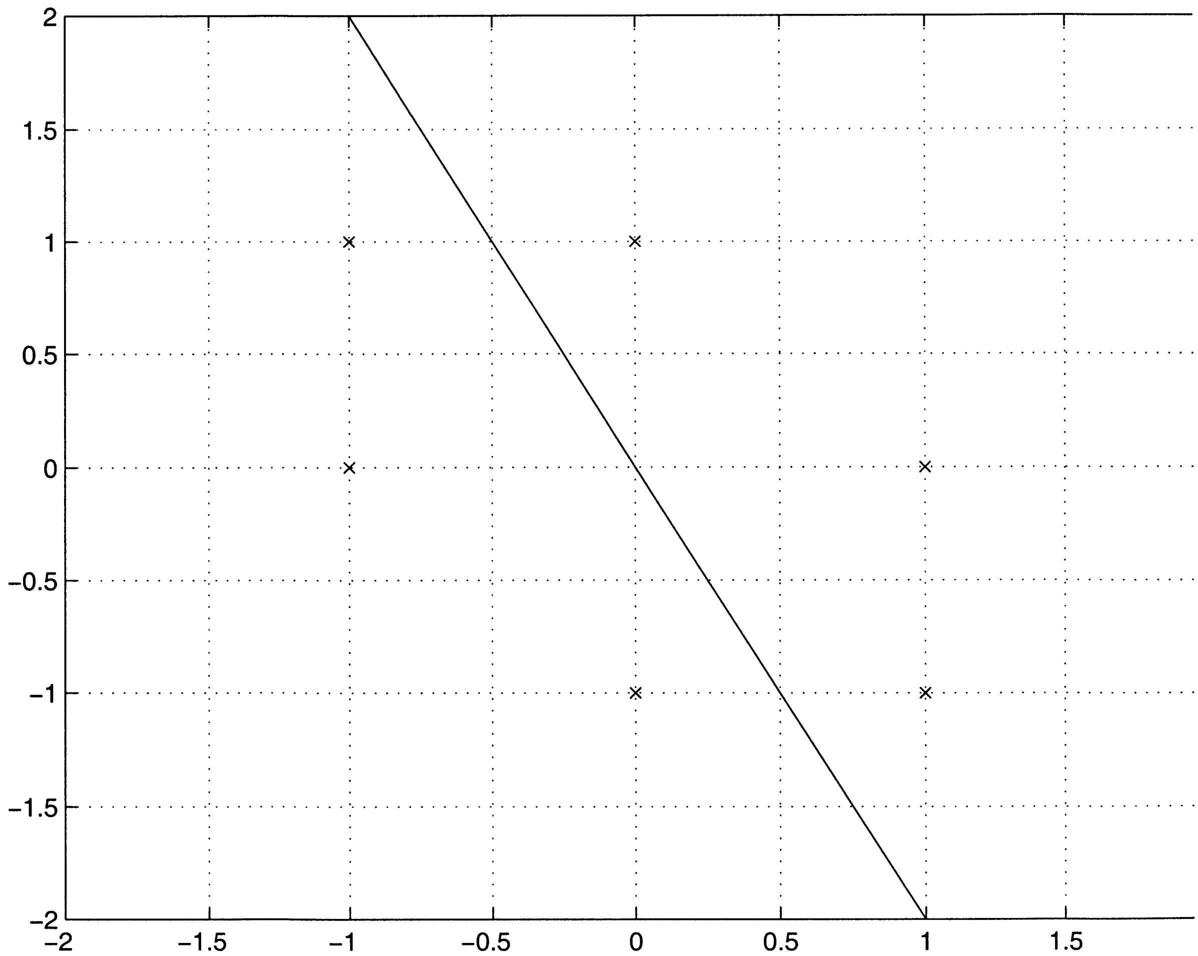
$$\begin{aligned}
\mathbf{x_1} &= [\,1\,,\ 0\,]', \\
\mathbf{x_2} &= [\,0\,,\ 1\,]', \\
\mathbf{x_3} &= [\,1\,,-1]', \\
\mathbf{x_4} &= [-1,\ 0\,]', \\
\mathbf{x_5} &= [-1,\ 1\,]', \\
\mathbf{x_6} &= [\,0\,,-1]'
\end{aligned}$$

(cont.

Feature vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are labelled class ω_1 and $\mathbf{x}_4$, $\mathbf{x}_5$ and $\mathbf{x}_6$ are labelled class ω_2. The decision boundary is required to go through the origin.

(i) Sketch the position of the points on a diagram and mark the range of possible decision boundaries that correctly classify the training data. [15%]

Answer: Points shown below.



(ii) What is the *cost function* for least mean squares estimation of a linear decision boundary? [10%]

(CONTINUED OVER.

Answer: The cost function is

$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{n}\left(\mathbf{w}'\mathbf{x}_k - t(\mathbf{x}_k)\right)^2$$
$$= \frac{1}{2}(\mathbf{X}\mathbf{w} - \mathbf{t})'(\mathbf{X}\mathbf{w} - \mathbf{t})$$

(iii) Using least mean squares estimation and the *pseudo-inverse* method, compute the decision boundary. The target value for class ω_1 is 1 , and for class ω_2 is -1 . Show this decision boundary on the sketch of part (b)(i). What is the training data classification performance? [30%]

Answer: Decision boundary requires solving

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{t}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

Now

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

Inverting this and solving gives

$$\mathbf{w} = \frac{1}{3}\begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

The decision boundary is shown on the figure. It classifies the training data perfectly.

(iv) Rather than using the pseudo-inverse to estimate the decision boundary for least mean squares estimation, gradient descent optimisation is to be used. Derive an appropriate update rule to find the decision boundary. Contrast this form of parameter estimation with the pseudo-inverse approach for solving least mean squares classification problems. [20%]

Answer: The gradient descent update rule is

$$\begin{aligned}\mathbf{w}^{(\tau+1)} &= \mathbf{w}^{(\tau)} - \eta \sum_{k=1}^{n} \left(\mathbf{w}^{(\tau)\prime}\mathbf{x}_k - t(\mathbf{x}_k)\right) \mathbf{x}_k \\ &= \mathbf{w}^{(\tau)} - \eta \mathbf{X}'(\mathbf{X}\mathbf{w} - \mathbf{t})\end{aligned}$$

The pseudo-inverse approach requires that $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ exists, which is not necessarily true. If the dimension is large it also avoids having to keep large matrices around.

END OF PAPER