Version SJG/1

EGT2
ENGINEERING TRIPOS PART IIA

Friday 22 April 2016 9.30 to 11

Module 3F3

SIGNAL AND PATTERN PROCESSING - SOLUTIONS

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed

10 minutes reading time is allowed for this paper.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 Examiner's comments:

This question was found reasonably straightforward by most, though hardly anyone could sketch the phase response in part (a)(iii). Similarly in (b)(i) a surprising number of students could not find the 8th roots of unity.

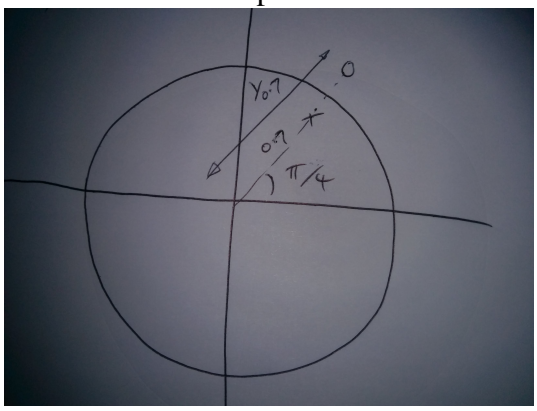
(a) A complex-coefficient digital filter has a transfer function of the following form:

$$H(z) = \frac{z^{-1} - r\exp(-j\phi)}{1 - z^{-1}r\exp(j\phi)}$$

(i) Sketch the pole-zero diagram for such a filter when $r = 0.7$ and $\phi = \pi/4$. [10%]

Solution:

1 zero at $r^{-1}e^{j\phi}$. 1 pole at $re^{j\phi}$:



(ii) Determine the frequency response of such a filter for any (r, ϕ) and show by geometrical arguments from the pole-zero diagram, or otherwise, that all frequencies are passed with equal gain, i.e. the filter is *all-pass*. [25%]

For frequency response substitute $z = e^{j\theta}$:

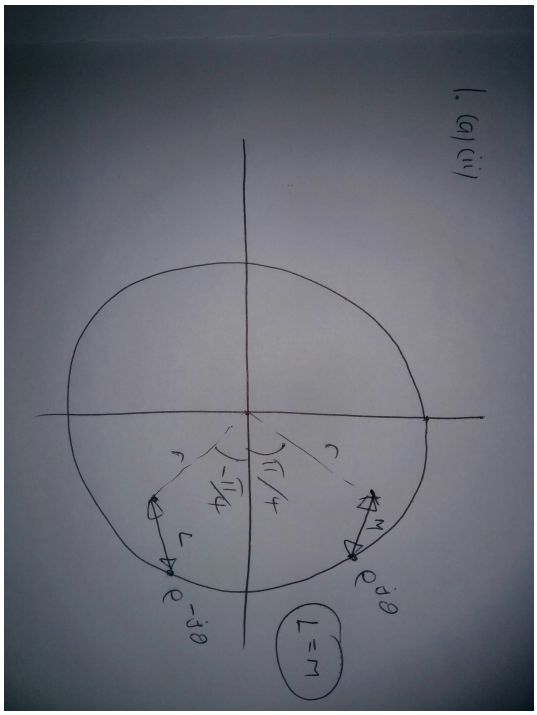
$$H(e^{j\theta}) = \frac{e^{-j\theta} - r\exp(-j\phi)}{1 - e^{-j\theta}r\exp(j\phi)}$$

Gain of filter is:

$$\begin{aligned} H(e^{j\theta}) &= \frac{|e^{-j\theta} - r\exp(-j\phi)|}{|1 - e^{-j\theta}r\exp(j\phi)|} \\ &= \frac{|e^{-j\theta} - r\exp(-j\phi)|}{|e^{j\theta} - r\exp(j\phi)|} \\ &= L/M = 1, \end{aligned}$$

see diagram:

(cont.



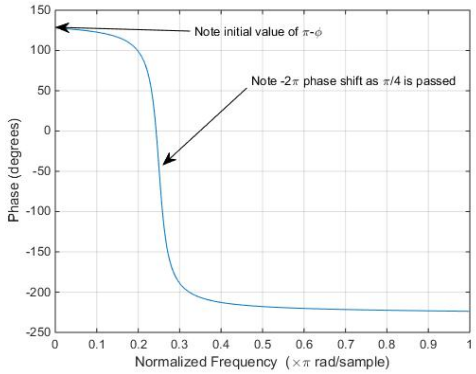
(iii) Sketch the phase response of the above filter when $r = 0.95$ and $\phi = \pi/4$, for a range of normalised frequencies from 0 to 2π. Your justification should be geometric and based on the pole-zero diagram; you may utilise the fact that pole/zero pairs are close to each other (and the unit circle) when $r = 0.95$. [20%]

Solution:

Use the standard result that phase response is the angle of $e^{j\theta} - c$ minus the angle of $e^{j\theta} - d$, where c is the zero of the filter and d is the pole. But note that there is an extra $\pi - \phi$ term since:

$$H(z) = \frac{-r\exp(-j\phi)(z - 1/r\exp(j\phi))}{z - r\exp(j\phi)} = \frac{r\exp(j(\pi-\phi))(z - 1/r\exp(j\phi))}{z - r\exp(j\phi)}$$

The first observation is that the pole and zero are close together, so that the phase from the pole/zero term is nearly zero when we are far from the pole/zero on the unit circle. Moreover, as we approach $\theta = \phi$, the phase sharply decreases by decreases by 2π:



(iv) Explain how a real-coefficient filter can be generated from a combination of two such structures, retaining the all-pass property but with modified phase response. [10%]

Solution:

We cascade it with a second all-pass complex filter having pole at $r, -\theta$, hence providing pole-zero conjugate pairs for the first filter. Thus the coefficients are real but it is still all-pass.

(b) (i) A second digital filter has the following transfer function:

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

where $0 < r < 1$ and $P > 1$ is an integer.

A. Sketch the pole-zero diagram for this filter when $P = 8$ and $r = 0.8$. [10%]

Solution:

Poles are at:

$$rz^{-P} = 1,$$

i.e.

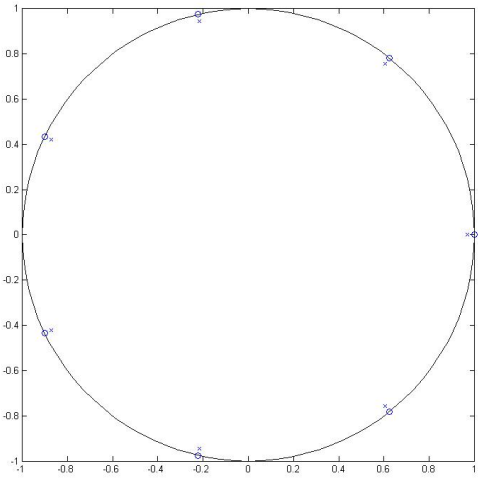
$$z = r^{1/P} e^{i2\pi p/P},\ p = 0, 1, 2, \ldots$$

Zeros similarly are at

$$z = e^{i2\pi p/P},\ p = 0, 1, 2, \ldots$$

Pole-zero diagram:

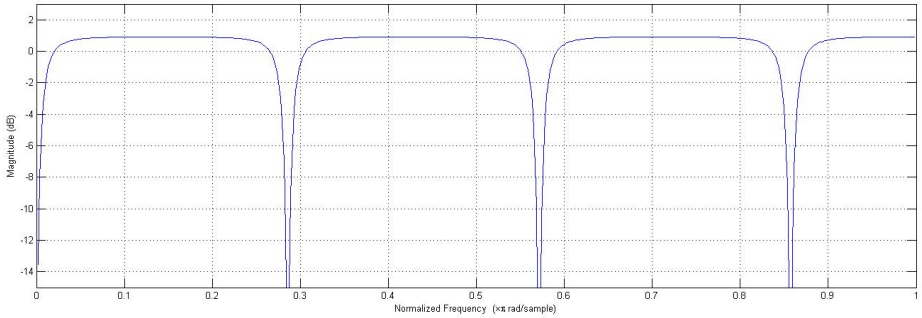
(cont.



B. Sketch also the frequency magnitude response of this filter when r is close to 1, and hence suggest an application for such a filter. What trade-off would be made by choice of the value of r? [25%]

Solution:

Pole and zero pairs are close to each other, so gain is approximately unity when distant from a zero/pole (ratio of distances is approximately 1). When moving close to a pole on the frequency axis (unit circle) the gain drops sharply to zero as it passes through the zero. Hence plot looks like:



This is a filter with multiple nulls at integer multiples of $2\pi/P$. Hence it can filter out periodic disturbances, e.g. mains noise by appropriate choice of the fundamental frequency $2\pi/P$. With r approaching 1 we get excellent frequency magnitude response, at the expense of poor time domain transient response (since poles cause components in the impulse response with envelope r^k).

2 Examiner's comment:

Attempted by nearly all candidates with very good results in general.

The discrete Fourier transform (DFT) of a sequence x_n, $n = 0, 1, ..., N-1$ is given by:

$$X_p = \sum_{n=0}^{N-1} x_n e^{-\frac{jnp2\pi}{N}}$$

(a) Show that the DFT spectrum values X_p are related to the true DTFT spectrum $X(e^{j\theta})$ of x_n, $n = -\infty, ..., -1, 0, 1, 2..., \infty$, by the following convolution formula:

$$X_p = \frac{1}{2\pi} \int_0^{2\pi} W(e^{j\theta}) X(e^{j(p(2\pi/N)-\theta)}) \, d\theta$$

where $W(e^{j\theta})$ is the DTFT of the appropriate rectangular window function. What frequency, in rads/s would correspond to $\theta = 0.2$, if the digital sampling frequency is 44.1kHz? [30%]

(b) Explain with the aid of diagrams how the DFT modifies the DTFT spectrum of a complex exponential signal $\exp(j\omega_0 t)$ where ω_0 is a fixed frequency. Describe the effects of spectral smearing and spectral leakage, how the use of window functions might aid the analysis, and why this is important for analysis of multiple frequency components. [20%]

(c) Show that when N is even, the DFT of a data sequence x_n may be expressed in the form:

$$X_p = A_p + W^p B_p \text{ and } X_{p+N/2} = A_p - W^p B_p$$

where A_p and B_p are DFTs of length $N/2$ sub-sequences of x_n, and W is a suitable complex exponential (you should derive the formulae for A_p, B_p and W). [30%]

(d) Explain briefly how the representation in (c) above allows a very efficient implementation of the DFT when N is a power of 2. Show that the computational load, neglecting additions, is approximately $(N/2)\log_2(N)$ complex multiplications and additions. How does this compare with direct evaluation of the DFT? [20%]

SOLUTION:

(a) Consider applying a recangular window $w_n = 1, n = 0, 1, ..., N-1$ to the signal and x_n and take the DTFT:

$$\begin{aligned} X_w(e^{j\omega T}) &= \sum_{n=-\infty}^{\infty} \{x_n w_n\} e^{-jn\omega T} \\ &= \sum_{n=0}^{N-1} \{x_n\} e^{-jn\omega T} \end{aligned}$$

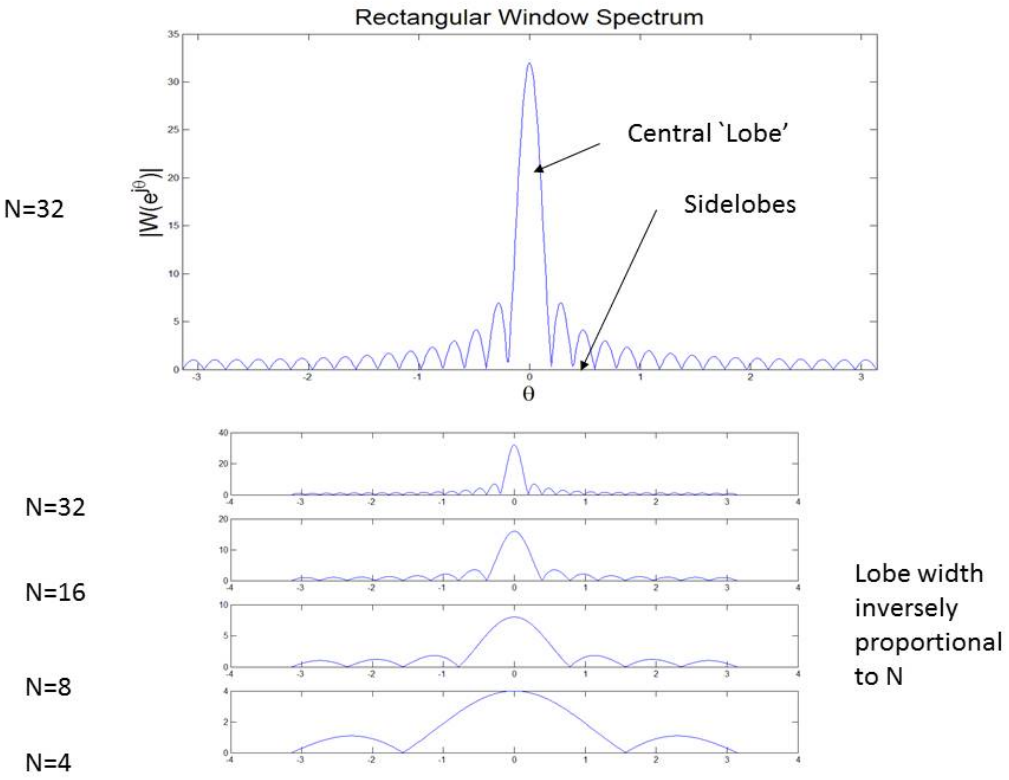
 (cont.

and clearly this is equal to the DFT when we evaluate at $\omega_p = 2\pi p/N$.
Now, proceeding from the first line:

$$\begin{aligned} X_w(e^{j\omega T}) &= \sum_{n=-\infty}^{\infty} x_p \left\{ \frac{1}{2\pi} \int_0^{2\pi} W(e^{j\theta}) e^{jn\theta} d\theta \right\} e^{-jn\omega T} \\ &= \frac{1}{2\pi} \int_0^{2\pi} W(e^{j\theta}) \sum_{n=-\infty}^{\infty} x_n e^{-jn(\omega T-\theta)} d\theta \\ X_w(e^{j\omega T}) &= \frac{1}{2\pi} \int_0^{2\pi} W(e^{j\theta}) X(e^{j(\omega T-\theta)}) d\theta \end{aligned}$$

and evaluating this at discrete frequencies $\omega_p = p2\pi/N$ gives the required result. [Knowledgable students could quote this result directly using the discrete time convolution theorem].
Now, the frequency at $\theta = 0.2$ is $0.2 \times 44100/2\pi = 1404Hz$.

(b) This is part is standard bookwork. Based on result of part (a) we see that the delta-funtions of the pure complex exponentials are convolved with the window spectrum, leading to spectral smearing and spectral leakage. Window functions from e.g. generalised Hamming families will increase specral smearing but can dramatically improve spectral leakage through much lower sidelobe levels. Suitable diagrams from lecture notes are:



Now, imagine what happens when the sum of two frequency components is DFT-ed:

$$x_n = \exp(j\omega_1 nT) + \exp(j\omega_2 nT)$$

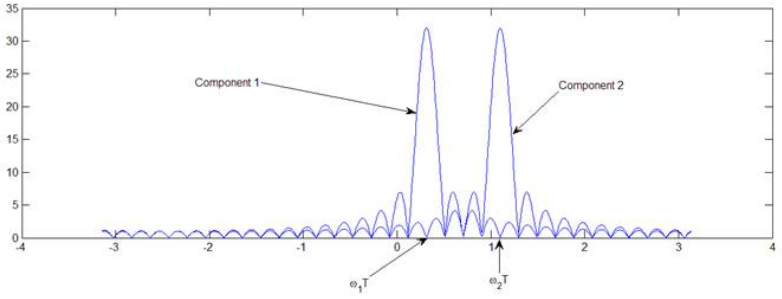
The DTFT is given by a train of delta functions:

$$X(e^{j\omega T}) = 2\pi \sum_{n=-\infty}^{+\infty} \delta(\omega T + 2n\pi - \omega_1 T) + \delta(\omega T + 2n\pi - \omega_2 T)$$

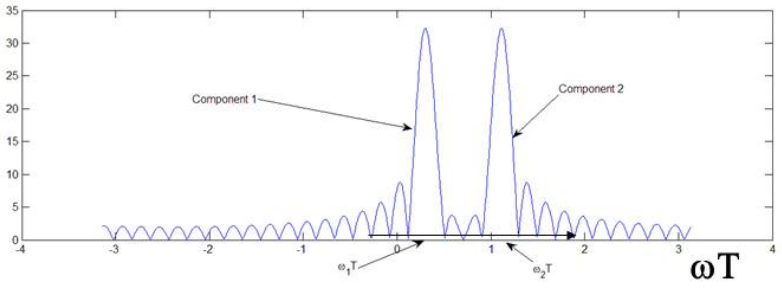
Hence the windowed spectrum is just the convolution of the window spectrum with the delta functions:

Now consider the DFT for the data:

Both components separately



Both components Together



(cont.

(c) First rewrite the DFT equation in terms of the even-indexed and odd-indexed data x_n:

$$\begin{aligned} X_p &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n}\, e^{-j\frac{2\pi}{N}(2n)p} + \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{N}(2n+1)p} \\ &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{(N/2)}np} + e^{-j\frac{2\pi}{N}p} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{(N/2)}np} \\ &= A_p + W^p B_p \end{aligned} \tag{1}$$

where

$$\begin{aligned} A_p &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{(N/2)}np} \\ B_p &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{(N/2)}np} \\ W &= e^{-j\frac{2\pi}{N}} \end{aligned}$$

Look at the DFT values $X_{p+N/2}$:

$$\begin{aligned} & X_{p+N/2} \\ &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n}\, e^{-j\frac{2\pi}{(N/2)}n(p+N/2)} \\ &\qquad + e^{-j\frac{2\pi}{N}(p+N/2)} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1}\, e^{-j\frac{2\pi}{(N/2)}n(p+\frac{N}{2})} \\ &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n}\, e^{-j\frac{2\pi}{(N/2)}np} \\ &\qquad - e^{-j\frac{2\pi}{N}p} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1}\, e^{-j\frac{2\pi}{(N/2)}np} \\ &= A_p - W^p B_p \end{aligned} \tag{2}$$

[Here we have used the results: $\exp(i(\theta+2\pi)) = \exp(i\theta)$ and $\exp(i(\theta+\pi)) = -\exp(i\theta)$.]

(d) Repeated application of the above procedure to each of the subsequence DFTs successively reduces the computation of the N-DFT to N single point DFTs, with no computational burden. The remaining computations are calculated as follows:

- Each stage of the FFT reduces computation by half but introduces an extra $\frac{N}{2}$ multiplications $W^p B_p$.
- For $N = 2^M$, the process can be repeated M times to reduce the computation to that of evaluating N single point DFTs which requires no computation.
- However, at each of the M stages of reduction an extra $\frac{N}{2}$ multiplications are introduced so that the total number of complex multiplies required to evaluate an N-point DFT is:

$$\frac{N}{2} \times (\text{Number of levels}) = \frac{N}{2} \log_2(N)$$

Compare with the DFT,which has roughly N^2 complex multiplications. We see that for large N there are very significant savings from use of the FFT.

3 Examiner's comment:

Popular question, with high marks in general. Lots of people did not know how to find average energy within a frequency band in part (a). Part (b)(ii) âĂŞ lots of candidates stated the result for the case when y is stationary, which lost some credit and made part (iii) awkward to demonstrate convincingly. In (b)(iii) many people were careless in checking the condition for finite variance and many forgot to check constant mean.

(a) For a discrete-time random process, explain the concept of *stationarity* and *ergodicity*, and define the terms autocorrelation function, wide-sense stationarity and power spectrum. Show how to obtain the average power of a real-valued process between two normalised frequencies ω_1 and ω_2, with $0 \leq \omega_1 < \omega_2 < \pi$. [30%]

SOLUTION:

Stationarity: a process is stationary if the the statistical characteristics centred at one time cannot be distinguished from those at any other time.

Ergodicity: an ergodic process 'forgets' its initial conditions with time and always converges on a stationary distribution of values.

Autocorrelation function (ACF):

$$R_{XX}[n,m] = E[X_n X_m]$$

Wide-sense stationary (WSS) : ACF depends only on time difference $m-n$, mean is constant over time, and variance is finite.

Power spectrum S_X is the DTFT of the ACF for a WSS process.

Calculate power as:

$$2\int_{\omega_1}^{\omega_2} S_X(e^{j\omega T})d\omega$$

(b) A random process $\{y_n\}$ is passed through a causal linear system having impulse response $h_m = \alpha^m$:

$$z_n = \sum_{m=0}^{+\infty} h_m y_{n-m}.$$

(i) Show how the linear system may be implemented as a first-order infinite impulse response (IIR) digital filter. [10%]

The z-transform of the impulse response is:

$$\frac{1}{1-\alpha z^{-1}}$$

so this can be implemented as

$$z_n = \alpha z_{n-1} + y_n$$

(ii) If $\{y_n\}$ has autocorrelation function $E[y_k y_l] = R_{YY}[k,l]$, find an expression for the cross-correlation function between $\{y_n\}$ and $\{z_n\}$, and also the autocorrelation function of the output process $\{z_n\}$. [20%]

SOLUTION:

Note that we haven't yet stated that $\{y_n\}$ is stationary, so:

$$\begin{aligned} R_{YZ}[k,l] &= E[y_k z_l] \\ &= E[y_k \sum_{m=0}^{\infty} h_m y_{l-m}] \\ &= E[\sum_{m=0}^{\infty} h_m y_k y_{l-m}] \\ &= \sum_{m=0}^{\infty} h_m R_{YY}[k,l-m] = \sum_{m=0}^{\infty} \alpha^m R_{YY}[k,l-m] \end{aligned}$$

$$\begin{aligned} R_{ZZ}[k,l] &= E[z_k z_l] \\ &= E[\sum_{n=0}^{\infty} h_n y_{k-n} \sum_{m=0}^{\infty} h_m y_{l-m}] \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} h_n h_m R_{YY}[k-n,l-m] \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha^{n+m} R_{YY}[k-n,l-m] \end{aligned}$$

(iii) If $\{y_n\}$ is wide-sense stationary, show that $\{z_n\}$ is also wide-sense stationary, provided the condition $|\alpha| < 1$ applies. [Hint: the largest value of the autocorrelation function is always at lag zero.] [20%]

Check mean of $\{z_n\}$:

$$E[z_n] = E[\sum_{m=0}^{\infty} h_m y_{k-m}] = \sum_{m=0}^{\infty} h_m \mu_Y$$

which is constant since it does not depend on n.

Autocorrelation function:

$$R_{ZZ}[k,l] = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha^{n+m} R_{YY}[k-n,l-m] = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha^{n+m} R_{YY}[l-m+n-k]$$

which depends only on the lag difference $l-k$, since the n and m variables are summed out.

Check variance of z:

$$R_{ZZ}[k,k] = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha^{n+m} R_{YY}[n-m]$$

Now, $|R_{YY}[n-m]| \leq R_{YY}[0]$, since the maximum autocorrelation value is at lag 0. Hence:

$$R_{ZZ}[k,k] \leq \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} |\alpha^{n+m}| R_{YY}[0] = R_{YY}[0] \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} |\alpha^{n+m}| = R_{YY}[0] \sum_{n=0}^{\infty} |\alpha^n| \sum_{m=0}^{\infty} |\alpha^m|$$

and both sums are finite only for $|\alpha| < 1$ - can check by summing GPs.
Hence $\sigma^2 = R_{ZZ}[k,k] - \mu_Z^2$ is also finite under the same condition and the process is WSS.

(iv) If $\{y_n\}$ is zero-mean white noise with variance 1, and $|\alpha| < 1$, determine the power spectrum of $\{y_n\}$, the autocorrelation function for $\{z_n\}$ and the power spectrum of $\{z_n\}$. Sketch the power spectrum $S_Z(\exp(j\theta))$ for $\alpha = 0.8$ over the range of normalised frequencies $\theta = 0$ to 2π. [20%]

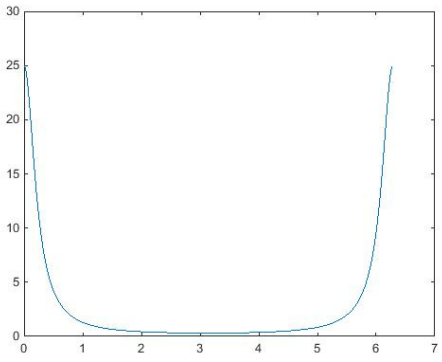
SOLUTION:
Power spectrum of $\{y_n\}$ is just flat:

$$S_Y(e^{j\theta}) = 1$$

Easiest way to the power spectrum for $\{z\}$ is through the frequency domain, and using the filtering result from part (b):

$$S_Z = |H(e^{j\theta})|^2 S_Y(e^{j\theta}) = \frac{1}{|1-\alpha e^{-j\theta}|^2}$$

though it could also be got as the DTFT of the autocorrelation function.
The sketch will look like this, with a resonance at zero and 2π corresponding to the single pole at $z = \alpha$, i.e. at zero frequency:



The autocorrelation function is:

$$R_{ZZ}[\delta] = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha^{n+m} R_{YY}[\delta - m + n] = \alpha^{\delta} \sum_{m=0}^{\infty} \alpha^{2n} = \alpha^{\delta}/(1-\alpha^2)$$

4 Examiner's comment:

The least popular question, but well handled by most.

Consider the k-means clustering algorithm which seeks to minimise the cost function

$$C = \sum_{n=1}^{N} \sum_{k=1}^{K} s_{nk} \|x_n - m_k\|^2$$

where m_k is the mean (centre) of cluster k, x_n is data point n, $s_{nk} = 1$ signifies that data point n is assigned to cluster k, and there are N data points and K clusters.

(a) Given all the cluster assignments s_{nk} (with the constraint that each data point must be assigned to one cluster, that is, $\sum_k s_{nk} = 1$ for all n, and $s_{nk} \in \{0,1\}$ for all n and k), derive the value of the means $\{m_k\}$ which minimise the cost C and give an interpretation in terms of the k-means algorithm. [30%]

(b) Give an interpretation of the k-means algorithm in terms of a probabilistic model. Describe up to three generalisations based on this probabilistic model. [40%]

(c) You are applying the k-means algorithm to a large collection of images, where most of the images are not labelled, but you have labels for a few of the images (e.g. "cat", "dog", "person", "car"). You would like to modify your k-means algorithm so that images with the same label are always in the same cluster, and images with different labels are never in the same cluster. Describe a modified version of the algorithm that would do this. [30%]

SOLUTION

(a) Given the cluster assignments, the problem decomposes into separate minimisations over each mean k, that is, $C = \sum_k C_k$. Since the cluster assignments are binary, for each mean we have a cost function

$$C_k = \sum_{n=1}^{N} s_{nk} \|x_n - m_k\|^2 = \sum_{n:s_{nk}=1} \|x_n - m_k\|^2$$

Minimising over m_k results in

$$m_k = \frac{\sum_{n:s_{nk}=1} x_n}{\sum_n s_{nk}}$$

which is simply the Euclidean mean of the data points assigned to cluster k.

(b) The k-means algorithm is closely related to the Gaussian mixture model, a probabilistic model for density esimation. In fact, the k-means cost is equal up to a constant to the (negative) log likelihood of a Gaussian mixture model under the

following assumptions: (1) the Gaussians have means m_k and covariances that are a multiple of the identity matrix, (2) the Gaussians all have equal mixing proportions, (3) the assignment variables which are actually hidden are treated as parameters and optimised rather than summed out. Upto these three constraints, k-means is almost identical to the EM algorithm. Once this relationship is established several generalisations become possible: (1) using different covariance matrices for each cluster to allow for elongated clusters at different orientations, (2) allowing different mixing proportions so that some clusters can be bigger than others, (3) handling partial membership of data points in clusters by accounting for the uncertainty in the assignment variables s_{nk}, (4) use of models other than the Gaussian to capture each cluster (e.g. mixtures of any other distribution), and (5) Bayesian generalisations whereby the number of clusters can be learned from data, and the uncertaintly in clustering is represented in the inference.

(c) Assume that the number of clusters K is equal or greater than the number of labels (otherwise the constraints can't be satisfied). Initalise assignments so that the labelled images belong to separate clusters (e.g. all "cats" in cluster 1, all "dogs" in cluster 2, etc). Run the k-means algorithm as before on all the unlabelled data, but ensure that the assignments for the labelled data remain unchanged. Since the constraints are imposed at initialisation and kept at each iteration, this will converge to a solution which is a (local) minimum of the cost C subject to the imposed constraints.

END OF PAPER

THIS PAGE IS BLANK