ENGINEERING TRIPOS PART IIA

Wednesday 1st May 2013 2 to 3:30

Module 3F3

SIGNAL AND PATTERN PROCESSING

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

All questions carry the same number of marks.

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

There are no attachments.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
Engineering Data Book
CUED approved calculator allowed

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

Version FINAL

1 The Discrete Fourier Transform (DFT) for a data sequence $\{x_n\}$ of length N, where N is here assumed to be a power of 2, is defined as

$$X_p = \sum_{n=0}^{N-1} x_n \, e^{-j\frac{2\pi}{N}np}, \quad p = 0, 1, ..., N-1$$

(a) Show that the DFT values X_p and $X_{p+N/2}$ may be expressed as

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

where A_p is a series involving only the even numbered data points $(x_0, x_2, ...)$ and B_p is a series involving only the odd numbered data points $(x_1, x_3, ...)$ and W is a constant which should be carefully defined. [30%]

Find the computational complexity for evaluating X_p and $X_{p+N/2}$ for $p = 0, 1, ..., N/2 - 1$ and compare this with a full evaluation of the DFT (assume that complex exponentials are pre-computed and stored). [20%]

50

(b) Show that the N-point DFT of a real-valued data sequence has conjugate symmetry, i.e. that

$$X_p = X^*_{N-p}$$ [20%]

70

(c) Hence or otherwise show how to efficiently compute the DFTs of two real data sequences $\{x_n\}$ and $\{y_n\}$ by computing the DFT of a single complex data sequence $\{z_n = x_n + jy_n\}$. What is the computational complexity of such a procedure compared with direct evaluation of the two DFTs separately? [30%]

$8N^2$ $\qquad$ $10 N/2 \log_2 N$

2 (a) The rectangular window centred on $n = 0$ is defined as:

$$w_n = \begin{cases} 1, & |n| \leq N/2 \\ 0, & \text{otherwise}. \end{cases}$$

Show that the discrete time Fourier transform (DTFT) of this function is given by:

$$W(e^{j\Omega}) = \frac{\sin(\Omega(N+1)/2)}{\sin(\Omega/2)}.$$

Sketch the magnitude of this spectrum, paying particular attention to the main lobe and first few sidelobes. You may assume that N is large. [40%]

(b) A signal x_n is multiplied by a general window function w_n to give $y_n = x_n w_n$. Show from first principles that the DTFT of y_n is given by

$$Y(e^{j\Omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\lambda})V(e^{j(\Omega-\lambda)})d\lambda$$

where $V()$ should be defined. [25%]

(c) An FIR filter is to be designed using the window method. The ideal frequency response within the range $\Omega = -\pi$ to π is specified as

$$D(\Omega) = \begin{cases} 1, & |\Omega| < \Omega_c \\ 0, & \text{otherwise} \end{cases}$$

where $0 < \Omega_c < \pi$.

The ideal filter coefficients are to be truncated to zero for $|n| > N/2$. Show that the frequency response of the resulting filter can be expressed as:

$$D_w(\Omega) = \frac{1}{2\pi}\int_{\Omega-\Omega_c}^{\Omega+\Omega_c} \frac{\sin(\lambda(N+1)/2)}{\sin(\lambda/2)} d\lambda.$$

[20%]

Hence explain, with the aid of sketches, the shape of the resulting frequency response, including the width of the transition band and any ripples in the passband or stopband. [15%]

3 (a) A stationary random process $\{e_n\}$ with autocorrelation function r_{EE} is the input to a stable linear system with impulse response $\{h_n\}$, giving output $\{x_n\}$:

$$x_n = \sum_{m=-\infty}^{+\infty} h_m e_{n-m}$$

Show that the cross-correlation function between input and output is given by:

$$r_{EX}[k] = \sum_{m=-\infty}^{+\infty} h_m r_{EE}[k-m].$$ [20%]

How is this result modified when $\{e_n\}$ is white noise and the linear system is *causal*? [10%]

(b) A 1st order ($P = 1$) autoregressive (AR) process obeys the following equation, with parameter $|\alpha| < 1$:

$$x_n = \alpha x_{n-1} + e_n$$

where e_n is zero mean white noise. Show that the autocorrelation function for this process obeys the following recursion:

$$r_{XX}[k] = \alpha r_{XX}[k-1] + r_{EX}[-k].$$ [10%]

Hence show that the autocorrelation function for the AR process is 35

$$r_{XX}[k] = \frac{\sigma_e^2}{1-\alpha^2}\alpha^{|k|}$$

where $\sigma_e^2 = E[e_n^2]$. [30%] 50

(c) The AR signal x_n is observed in a noisy and reverberant environment

$$y_n = x_n - 0.8x_{n-1} + v_n,$$

where v_n is zero mean white noise having variance $\sigma_v^2 = 1$.

Now, take $\alpha = 0.9$ and $r_{XX}[k] = \alpha^{|k|}$. It is desired to estimate x_n from measurements only of y_n by filtering with an FIR filter having impulse response b_n, $n = 0, 1$:

$$\hat{x}_n = b_0 y_n + b_1 y_{n-1}.$$

Show that the optimal Wiener filter coefficients must satisfy the equations: 0

$$b_0 r_{YY}[0] + b_1 r_{YY}[1] = r_{YX}[0]$$

$$b_0 r_{YY}[1] + b_1 r_{YY}[0] = r_{YX}[1]$$

and hence determine the coefficients of the filter. [30%]

Version FINAL

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

(b) You would like to automatically learn the number of clusters K from data. One possibility is to minimise the cost C as a function of K. Explain whether this is a good idea or not, and what the solution to this minimisation is. [30%]

(c) Consider an algorithm for clustering high-dimensional data which first performs a principal components analysis (PCA) dimensionality reduction on the data, and then runs k-means on the low dimensional projection of the data. Will this result in the same clustering of the data as running k-means on the original high-dimensional data? Explain your answer. [40%]

END OF PAPER