Thursday May 2013 9.00 to 10.30

Module 3F3

SIGNAL AND PATTERN PROCESSING

WORKED SOLUTIONS.

STATIONERY REQUIREMENTS SPECIAL REQUIREMENTS

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 1

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

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

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%]

Solution:

This is a very detailed solution - more detailed than requires for the 50% marks.

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

where

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

To see how this simplifies, look at the original DFT in (*) above, but evaluated at

frequencies $p+N/2$:

$$X_{p+N/2} = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{(N/2)}n(p+N/2)} + e^{-j\frac{2\pi}{N}(p+N/2)} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{(N/2)}n(p+\frac{N}{2})}$$

Now, simplify terms as follows:

$$e^{-j\frac{2\pi}{(N/2)}n(p+N/2)} = e^{-j\frac{2\pi}{(N/2)}np}, \quad e^{-j\frac{2\pi}{N}(p+N/2)} = -e^{-j\frac{2\pi}{N}p}$$

Hence,

$$X_{p+N/2} = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{(N/2)}np} - e^{-j\frac{2\pi}{N}p} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{(N/2)}np}$$

$$= A_p - W^p B_p$$

with A_p W^p and B_p defined as before.

Look at the two required terms (W^p assumed precomputed and stored):

$$A_p = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{(N/2)}np}, \quad B_p = \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{(N/2)}np},$$

- •The terms A_p and B_p need only be computed for $p = 0,2,...,N/2-1$, since $X_{p+N/2}$ has been expressed in terms of A_p and B_p - hence we have uncovered redundancy in the DFT computation.
- •Thus calculate the A_p and B_p for $p = 0,1,...,N/2-1$ and use them for calculation of both X_p and $X_{p+N/2}$
- •The number of complex multiplies and additions is:
  - –A_p requires $N/2$ complex multiplies and additions; so does B_p. The total for all $p = 0,1,...,N/2-1$ is then $2(N/2)^2$ multiplies and additions for the calculation of all the A_p and B_p terms.
  - –$N/2$ multiplies for the calculation of $W^p B_p$ for all $p = 0,1,2,...,N/2-1$
  - –$N = N/2+N/2$ additions for calculation of $A_p + W^p B_p$ and $A_p - W^p B_p$
- •Thus total number of complex multiplies and additions is approximately $N^2/2$ for large N

 (TURN OVER for continuation of Question 1

•The computation is approximately halved compared to the direct DFT evaluation

(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%]

Solution: From definition:

$$X_{N-p} = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}n(N-p)} = e^{-j\frac{2\pi nN}{N}} \sum_{n=0}^{N-1} x_n e^{+j\frac{2\pi}{N}np} = 1.X^*_p$$

Hence shown.

(c) Hence or otherwise show how to efficiently compute the DFTs of two real data sequences $\{x^1_n\}$ and $\{x^2_n\}$ by computing the DFT of a single complex data sequence $\{x_n = x^1_n + jx^2_n\}$. What is the computational complexity of such a procedure compared witrh direct evaluation of the two DFTs separately? [30%]

Solution:

We want X^1_p and X^2_p separately, but they are mixed up as $X_p = X^1_p + jX^2_p$ (by linearity of DFT).

Now, from part b) result,

$$X_{N-p} = X^{1*}_p + jX^{2*}_p$$

So

$$X_p = X^{1R} - X^{2I} + j(X^{1I} + X^{2R})$$

and

$$X_{N-p} = X^{1R} + X^{2I} + j(-X^{1I} + X^{2R})$$

Hence,

$$X^1_p = \frac{X_p + X^*_{N-p}}{2}, \quad X^2_p = \frac{X_p - X^*_{N-p}}{2j}$$

as required.

2 (a) The rectangle 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:

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

Solution:

DTFT is defined as:

$$W(\exp(j\Omega)) = \sum_{n=-\infty}^{+\infty} w_n \exp(-jn\Omega)$$

so here we have

$$W(\exp(j\Omega)) = \sum_{n=-N/2}^{+N/2} 1.\exp(-jn\Omega)$$

$$= \frac{\exp(+jN/2\Omega)(1-\exp(-j\Omega)^{N+1})}{1-\exp(-j\Omega)}$$

[GP with constant ratio $\exp(-j\Omega)$]

$$= \frac{\exp(+jN/2\Omega - j\Omega(N+1)/2)(\exp(+j\Omega(N+1)/2) - \exp(-j\Omega(N+1)/2)}{(\exp(-j\Omega/2))(\exp(+j\Omega/2) - \exp(-j\Omega/2))}$$

[Factorising exponential terms to get complex pairs top and bottom]

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

as required.

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%]

Solution:

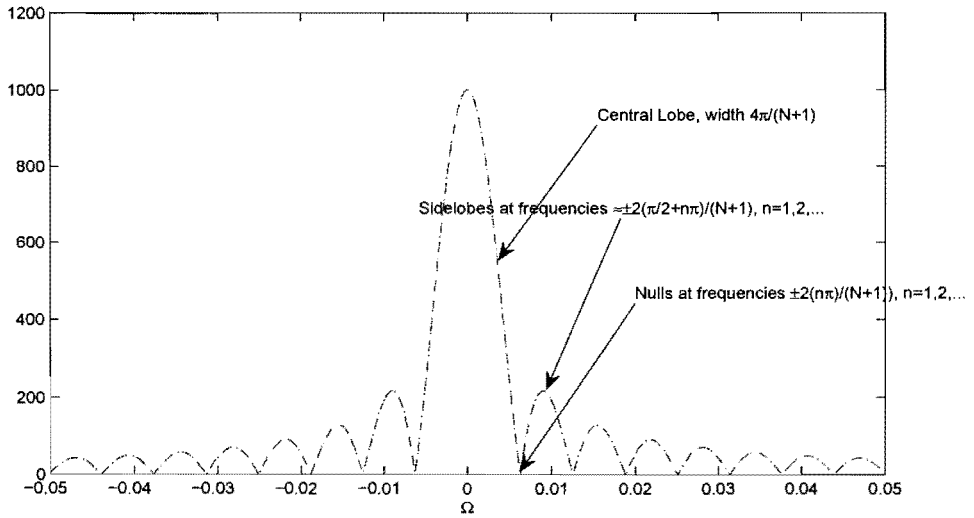
For N large, we will see 'fast' oscillations in the spectrum due to the $\sin(\Omega(N+1)/2)$ terms. These will lead to central lobe and sidelobes. Studying the first few of these, for low frequency Ω, the term $\sin(\Omega/2)$ behaves as $\Omega/2$ ('small x' approximation of sin(x)). Hence we have approximately:

$$\frac{\sin(\Omega(N+1)/2)}{\Omega/2)} = (N+1)\text{sinc}(\Omega(N+1)/2)$$

(TURN OVER for continuation of Question 2

Therefore we have value $N+1$ at frequency zero and the central lobe extends to the first null at $\Omega = \pm 2\pi/(N+1)$, and sidelobe maxima around $\Omega = \pm 2(\pi/2 + n\pi)/(N+1)$, $n = 1,2,3,....$

Sketching, we have:



(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

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

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

Solution:

$$\begin{aligned} Y(e^{j\Omega}) &= \sum_{n=-\infty}^{\infty} \{w_n x_n\} e^{-jn\Omega} \\ &= \sum_{n=-\infty}^{\infty} w_n \left\{ \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\theta}) e^{jn\theta} d\theta \right\} e^{-jn\Omega} \\ &= \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\theta}) \sum_{n=-\infty}^{\infty} w_n e^{-jn(\Omega-\theta)}\, d\theta \\ Y(e^{j\Omega}) &= \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\theta}) W(e^{j(\Omega-\theta)})\, d\theta \end{aligned}$$

where $V() = W()$ is the DTFT of the window function.

 (cont.

(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%]

Solution:

Using part b) result:

$$D_w(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D(\exp(j\lambda))\, W\left(\exp(j(\Omega-\lambda))\right) d\lambda$$

and substituting for D:

$$\begin{aligned} D_w(\Omega) &= \frac{1}{2\pi} \int_{-\Omega_c}^{\Omega_c} 1.W\left(\exp(j(\Omega-\lambda))\right) d\lambda \\ &= \frac{1}{2\pi} \int_{\Omega-\Omega_c}^{\Omega+\Omega_c} W\left(\exp(j(\lambda'))\right) d\lambda' \\ &\quad \text{[substituting } \lambda' = \Omega - \lambda \text{ and swapping limits]} \\ &= \frac{1}{2\pi} \int_{\Omega-\Omega_c}^{\Omega+\Omega_c} \frac{\sin(\lambda'(N+1)/2)}{\sin(\lambda'/2)} d\lambda' \end{aligned}$$

101

as required.

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%]

Solution:

Now we can see that the resulting frequency response is just the area under the 'sinc' function over different intervals of width $2\Omega_c$ and centered upon frequency Ω, see figure for commentary.

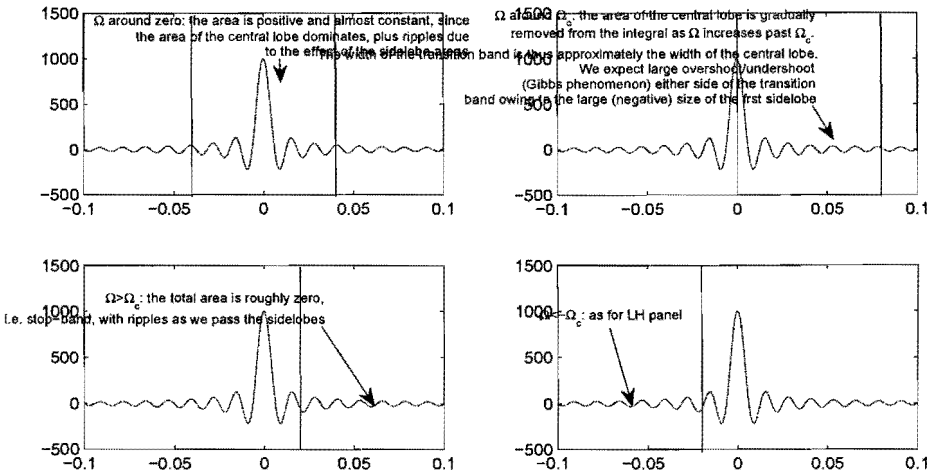


Fig. 1

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:

$$R_{EX}[k] = \sum_{l=-\infty}^{+\infty} h_l R_{EE}[k-l].$$

[20%]

Solution:

$$\begin{aligned} R_{EX}[k] &= E[e_n \sum_{m=-\infty}^{+\infty} h_m e_{n+k-m}] \\ &= \sum_{m=-\infty}^{+\infty} h_m E[e_n e_{n+k-m}] \\ &= \sum_{m=-\infty}^{+\infty} h_m r_{EE}[k-m] \end{aligned}$$

as required.

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

Solution: If the system is causal then $h_m = 0$ for $m < 0$. Thus,

$$R_{EX}[k] = \sum_{m=0}^{+\infty} h_m r_{EE}[k-m]$$

(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:

$$R_{XX}[k] = \alpha R_{XX}[k-1] + R_{EX}[-k].$$

[20%]

Solution:

$$\begin{aligned} R_{XX}[k] &= E[x_n x_{n+k}] \\ &= E[x_n(\alpha x_{n+k-1} + e_{n+k})] \\ &= \alpha r_{XX}[k-1] + r_{XE}[k] \\ &= \alpha r_{XX}[k-1] + r_{EX}[-k] \end{aligned}$$

since $r_{XE}[k] = r_{XE}[-k]$.

Hence show that the autocorrelation function for the AR process is

$$R_{XX}[k] = \alpha^{|k|}\sigma_e^2$$

where $\sigma_e^2 = E[e_n^2]$. [20%]

Solution: From the above, we need an expression for $r_{EX}[-k]$. Now,

$$R_{EX}[k] \sum_{m=0}^{+\infty} h_m r_{EE}[k-m] = h_k \sigma_e^2$$

for this case, since $r_{EE}[k] = \delta_k \sigma_e^2$ (zero mean white noise).

But, the AR system is causal, so $h_k = 0$ for $k < 0$. Also, $h_0 = 1$ from the AR difference equation with input $e_n = \delta_n$ and zero initial conditions on x_{-1}....

Thus,

$$r_{XX}[0] = \alpha r_{XX}[-1] + \sigma_e^2 = \alpha r_{XX}[1] + \sigma_e^2$$

 (TURN OVER for continuation of Question 3

(since $r_{XX}[1] = r_{XX}[-1]$) and

$$R_{XX}[k] = \alpha R_{XX}[k-1],\ k > 0$$

Solving for $r_{XX}[0]$ and $r_{XX}[0]$:

$$r_{XX}[0] = \alpha \times \alpha r_{XX}[0] + \sigma_e^2$$

or,

$$r_{XX}[0] = \frac{\sigma_e^2}{1-\alpha^2}$$

But the second equation implies that

$$r_{XX}[k] = r_{XX}[0](\alpha)^k$$

So, finally,

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

where the $|k|$ term arises because we know that $r_{XX}[k] = r_{XX}[-k]$.

(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:

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

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

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

Solution:

Version 1 (cont.

Take derivative of the Wiener error criterion:

$$\frac{\partial E[(x_n - \hat{x}_n)^2]}{\partial b_i} = 2E[(x_n - \hat{x}_n)\frac{\partial (x_n - \hat{x}_n)}{\partial b_i}] = 0,\ i = 0, 1$$

and:

$$\frac{\partial x_n - \hat{x}_n}{\partial b_0} = -y_n$$

$$\frac{\partial x_n - \hat{x}_n}{\partial b_1} = -y_{n-1}$$

Hence:

$$E[(x_n - \hat{x}_n)y_n] = 0,\ E[(x_n - \hat{x}_n)y_{n-1}] = 0$$

and expanding out:

$$E[(x_n - (b_0 y_n + b_1 y_{n-1}))y_n] = 0,\ E[(x_n - (b_0 y_n + b_1 y_{n-1}))y_{n-1}] = 0$$

$$E[(x_n - (b_0 y_n + b_1 y_{n-1}))y_n] = 0,\ E[(x_n - (b_0 y_n + b_1 y_{n-1}))y_{n-1}] = 0$$

$$r_{XY}[0] - b_0 r_{YY}[0] - b_1 r_{YY}[1],\ r_{XY}[-1] - b_0 r_{YY}[-1] - b_1 r_{YY}[0]$$

which, upon rearranging with $r_{YY}[-1] = r_{YY}[1]$, yields the required result.

...

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, $\sum_k s_{nk} = 1$ 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%]

SOLUTION

(a) Given the means, the problem decomposes into separate minimisations over each data point n. For data point n, the solution is to set $s_{nk} = 1$ for the value k which has the smallest distance $\|x_n - m_k\|^2$, and to set $s_{nk'} = 0$ for all $k' \neq k$. In terms of the k-means algorithm, the interpretation is that we assign each data point to the cluster with the nearest centre, as measured by Euclidean distance.

(b) Minimising C as a function of K is *not* a good idea. There are (at least) two ways to see this. One is that the optimal k-means cost for K could always be decreased by adding a new $K+1$st centre, since the optimal solution for K is generally a suboptimal case for $K+1$. A second way is to consider the extreme where we have as many clusters as data points, $K = N$. Then clearly we can obtain a cost $C = 0$ simply by placing each mean on a distinct data point (e.g. $m_n = x_n$). Since $C \geq 0$, this is the lowest possible cost solution no matter how the data is distributed, so it gives no insight into the actual number of clusters in the data.

(c) Running PCA dimensionality reduction on the data, and then k-means, will not in general result in the same solution as running k-means directly on the high dimensional data. To see this, consider in general that $y_n = Wx_n$ is the lower dimensional PCA projection of x_n. Running k-means on $\{y_n\}$ means that we are mininising $\tilde{C} = \sum_{n,k} s_{nk} \|Wx_n - \tilde{m}_k\|^2$ rather than $C = \sum_{n,k} s_{nk} \|x_n - m_k\|^2$. Assume we initialise all $\tilde{m}_k = Wm_k$ (the low dimensional projection of the means). Then one step of k-means for $\tilde{C}$ assigns $s_{nk} = 1$ if $\|Wx_n - Wm_k\|^2 = (x_n - m_k)^\top W^\top W(x_n - m_k)$ is minimised, rather than $\|x_n - m_k\|^2$. Therefore the PCA k-means corresponds to using a non-Euclidean norm to find the nearest mean, and generally only coincides with the original k-means when $W^\top W = I$.

END OF PAPER