Version FINAL

EGT3
ENGINEERING TRIPOS PART IIA

Thursday 25th April 2015 2 to 3.30

Module 3F3

SIGNAL AND PATTERN PROCESSING

WORKED SOLUTIONS

STATIONERY REQUIREMENTS

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
Engineering Data Book
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 ASSESSOR's comment:

A popular question, well-answered by most candidates. Parts a) and b) were well done by most. c) was OK, but some candidates did not make detailed enough comments to justify 20

(a) The Discrete-time Fourier Transform (DTFT) for a dataset $\{x_n\}$ is denoted $X(e^{j\Omega})$, where Ω is the normalised frequency such that $\Omega = 2\pi$ corresponds to the sampling frequency of the data.

A length-N Discrete Fourier Transform (DFT) is now calculated for the data,

$$X_p = \sum_{n=0}^{N-1} x_n e^{-2jn\pi/N}, \; p = 0, 1, \ldots, N-1.$$

Show that the DFT coefficients are related to the DTFT through the expression:

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

where $W_r(e^{j\Theta})$ is a function which should be carefully calculated and defined. [40%]

SOLUTION:

This is as bookwork as follows: Consider DTFT:

$$X(e^{j\omega T}) = \sum_{n=-\infty}^{\infty} x_n e^{-jn\omega T}$$

Truncate the summation, as for the DFT:

$$X_w(e^{j\omega T}) = \sum_{n=0}^{N-1} x_n e^{-jn\omega T}$$

Now, note that this is equivalent to an infinite summation, but with x_n pre-multiplied by a rectangle window function:

$$X_w(e^{j\omega T}) = \sum_{n=-\infty}^{\infty} w_n x_n e^{-jn\omega T}$$

with

$$w_n = \begin{cases} 1, & n = 0, 1, 2, \ldots, N-1 \\ 0, & \text{otherwise} \end{cases}$$

Now, take the DTFT of the windowed signal $x_w = w_n x_n$ directly:

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

We see that the spectrum of the windowed signal is the convolution of the infinite duration signal spectrum and the window spectrum. The result is obtained by substituting $\omega T = p2\pi/N$, the DFT bin frequency.

What is the DTFT of the window w_n?

[Subst. $\theta = \omega T$ - makes no difference to form of results]:

$$\begin{aligned}
W(e^{j\theta}) &= \sum_{n=0}^{N-1} 1.e^{-jn\theta} \\
&= e^{-j(N-1)\theta/2} \frac{\sin(N\theta/2)}{\sin(\theta/2)}
\end{aligned}$$

(b) Sketch $|W_r(e^{j\Theta})|$ for a small range of frequencies, showing the central lobe and several sidelobes. Calculate and clearly mark on the sketch the height of the central lobe and the positions of first few sidelobes and nulls (i.e. frequencies where $|W_r(e^{j\Theta})| = 0$). [20%]

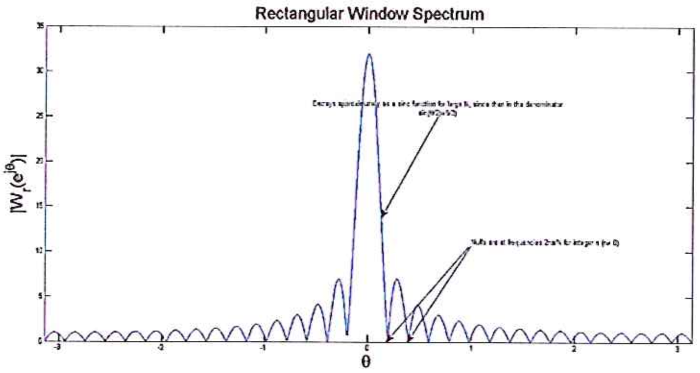
SOLUTION:

Height of central lobe is, using small angle approximations for sine:

$$\lim_{\theta \to 0} \frac{\sin(N\theta/2)}{\sin(\theta/2)} = \frac{\sin(N\theta/2)}{\theta/2} = N$$

Sketch is then:

(c) Suppose that $\{x_n\}$ is composed purely of complex frequency components of the form $a_i \exp(jn\Omega_i)$, where Ω_i is the fixed frequency of the ith component with magnitude a_i. Explain the effect of using the DFT to estimate the spectrum compared with the full DTFT of such a signal. Your explanation should include the effect of N, spectral smearing, leakage and the use of window functions. [20%]



SOLUTION:

This is bookwork - smearing of components due to the convolution of delta function frequency components with the window spectrum. Both effects inversely proportional to N since spectrum width of W_r is inversely proportional to N as in (b). Mention generalised Hamming window and how it cancels sidelobes at the expense of wider central lobe (i.e. greater spectral smearing, less leakage), ...

(d) Consider a signal which contains frequencies which lie exactly on one of the DFT bins, i.e. $\Omega_i = k2\pi/N$ where k is an integer. How, if at all, would your comments in the previous section be modified? If it is known in advance that the frequencies satisfy this constraint, what would be the minimum gap between two frequency components in order to detect their presence individually (i.e. 'resolve' them). Consider the unwindowed case

[20%]

SOLUTION: In this case the nulls of the rectangle window coincide exactly with the frequency bins adjacent to the frequency component, hence there is no leakage into adjacent components. Thus can 'resolve' up to 1 frequency bin, or $\delta_\Omega = 2\pi/N$rad/sample. [In case of Hamming window the resolution will be poorer since now the central lobe is smeared further into the adjacent frequency bins either side].

2 **ASSESSOR's comment:**

Generally well answered. In a), some candidates only gained partial marks by simply transforming the s-domain poles via the bilinear transform, without showing that these are the poles of the resulting digital filter. b) also well answered by many, with some ingenious methods used to prove that the poles lie inside the unit circle. c) This was very rough and ready. Most could sketch a cascade form, but were very unclear about what scale factors apply and how to assign the biquadratic coefficients to the designed filter from a). d) was quite well done; many candidates were able to quote the necessary result for the output power spectrum, and a few being able to accurately sketch it.

A digital filter equaliser is to be designed based on an all-pole analogue prototype filter of the form:

$$H(s) = \prod_{i=1}^{P} \frac{1}{s^2 - 2r_i \cos(\theta_i)s + r_i^2}$$

where $\pi/2 < \theta_i \leq \pi$, $0 < r_i < \infty$, all θ_i are assumed distinct from one another and all r_i are assumed distinct from one another.

The change of variables $s = \frac{1-z^{-1}}{1+z^{-1}}$ is used to design a digital filter $\hat{H}(z)$ based on this prototype.

(a) Show, by keeping $\hat{H}(z)$ factorised in a manner corresponding to $H(s)$, that the new digital filter has poles at

$$\hat{p}_i = \frac{1+p_i}{1-p_i} \text{ and } \hat{p}_i^* = \frac{1+p_i^*}{1-p_i^*}$$

where $p_i = r_i \exp(j\theta_i)$, and show that the digital filter also has $2P$ zeros, all co-located at $z = -1$. [30%]

SOLUTION:

Note that the analogue filter has poles at:

$$p_i = r_i \exp(\pm j\theta_i)$$

and that these are stable poles by the constraints on r and θ.

Consider ith section of filter only:

$$H_i(s) = \frac{1}{(s-p_i)(s-p_i^*)}$$

Making the substitution (the bilinear transform) gives:

$$\hat{H}_i(z) = \frac{1}{(\frac{1-z^{-1}}{1+z^{-1}} - p_i)(\frac{1-z^{-1}}{1+z^{-1}} - p_i^*)}$$
$$= \frac{(1+z^{-1})^2}{(1-z^{-1}-(1+z^{-1})p_i)(1-z^{-1}-(1+z^{-1})p_i^*)}$$
$$= \frac{(1+z^{-1})^2}{(1-p_i-(1+p_i)z^{-1})(1-p_i^*-(1-p_i^*)z^{-1})}$$

Hence it is clear that, for each section, there are two zeros at $z = -1$ and poles at:

$$\hat{p}_i = \frac{1+p_i}{1-p_i} \quad \hat{p}_i^* = \frac{1+p_i^*}{1-p_i^*}$$

ASSESSOR's comment:

Most candidates handled this well and worked it through completely.Those who simply converted the s-plane poles using the bilinear transform did not get full marks unless they showed the result that the bilinear transform also transforms the poles via the bilinear transform.

(b) By considering the poles and/or zeros of the new digital filter as necessary, show whether $\hat{H}(z)$ is guaranteed to be stable or not. [20%]

SOLUTION:

Consider just the modulus of the poles for stability:

$$|\hat{p}_i| = \frac{|1+p_i|}{|1-p_i|}$$
$$= \frac{\sqrt{(1+r_i\cos\theta_i)^2 + r_i^2\sin(\theta_i)^2}}{\sqrt{(1-r_i\cos\theta_i)^2 + r_i^2\sin(\theta_i)^2}}$$
$$= \frac{\sqrt{1+r_i^2+2r_i\cos(\theta_i)}}{\sqrt{1+r_i^2-2r_i\cos(\theta_i)}}$$

Now $\cos(\theta_i) < 0$ for the range of possible θ_i, so the nemerator is always less than the denominator and the pole is therefore stable.

Hence filter is guaranteed stable.

ASSESSOR's comment:

Again, candidates who just stated that the bilinear transform gives stable poles would not get full marks. Many candidates spotted the above reasoning and some provided ingenious geometric solutions.

(c) Sketch a cascade realisation of Direct form II biquadratic sections, using as few multiplications as possible, which would be suitable for implementation of $\hat{H}(z)$. [20%]

SOLUTION:

Implement as biquadratic sections. Denominator of ith section is:

$$(1-p_i-(1+p_i)z^{-1})(1-p_i^*-(1-p_i^*)z^{-1}) = \; = \frac{|1+p_i|^2}{\hat{r}_i^2}(1-2\hat{r}_i\cos\hat{\theta}_i z^{-1}+\hat{r}_i^2 z^{-2})$$

where $\hat{p}_i = \hat{r}_i\exp(j\hat{\theta}_i)$. Here we recall that the feedback part of an (2nd order) all-pole IIR filter needs to have a unity first coefficient a_0 so it can be implemented as:

$$y_n = x_n - a_1 y_{n-1} - a_2 y_{n-2}$$

Hence the coefficients of the feedback part are: $a_1 = -2\hat{r}_i\cos(\hat{\theta}_i)$ and $a_2 = \hat{r}_i^2$, noting that each section has a gain of $|1+p_i|^2/\hat{r}_i^2$, which can be accumulated and applied at the last biquadratic section. The feedforward part has coefficients $b_0 = 1$, $b_1 = 2$, $b_2 = 1$ since trhe numerator is $(1+z^{-1})^2 = (1+2z^{-1}+z^{-2})$. Note that this avoids any multiplications for b_0 and b_2, so we would not include the gain at each stage of the biquadratic cascade.

ASSESSOR's comment:

This part was handled rather poorly by many candidates. Most were broadly familiar with the Direct Form II structure and cascade implementations, but were not careful to calculate the coefficients of the feedback/ feedforward parts correctly, or to allow for the overall gain factor on the cascade. Although not strictly necessary, it helps to know the formula for the biquadratic section in terms of $(1-2\hat{r}_i\cos\hat{\theta}_i z^{-1}+\hat{r}_i^2 z^{-2})$, but most did not use this.

(d) A zero-mean white noise sequence with variance $\sigma_w^2 = 1$ is input to a filter $\hat{H}(z)$ as above with $P = 1$, $r_1 = 0.2$ and $\theta_1 = 5\pi/8$. Determine the power spectrum of the output from the filter, sketching the result between normalised frequency $\Omega = 0$ and π, paying particular attention to the DC gain, gain at $\Omega = \pi$ and any maxima/minima of the gain. [30%]

SOLUTION:

With $r_1 = 0.2$ and $\theta_1 = 5\pi/8$ we have that $\hat{r}_1 = 0.862$ and $\hat{\theta}_1 = 0.368$ (by $\hat{p}_1 = \frac{1+p_1}{1-p_1}$.

Power spectrum is:

$$\mathscr{S}_x(\exp(j\Omega) = \sigma_w^2|\hat{H}(\exp(j\Omega))|^2$$

with $\sigma_w^2 = 1$.

For $|\hat{H}(\exp(j\Omega))|^2$ we expect a maximum around $\Omega = 0.368$, the pole frequency and a minumum (gain=0) at $\Omega = \pi$. By symmetry also a (non-zero) minimum at $\Omega = 0$. Gain at $\Omega = 0$ is the same as the gain of the analogue filter at $\omega = 0$, i.e $1/r_1^2 = 25$.

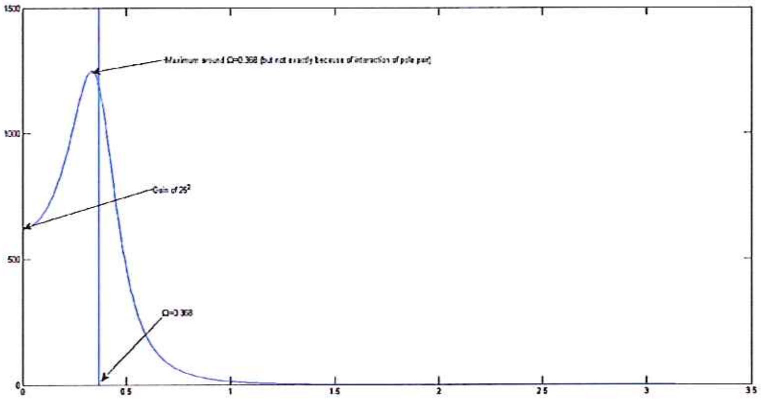


Fig. 1

See plot for how the sketch should come out.

ASSESSOR's comment: This was mostly well handled by candidates.

3 **ASSESSOR's comment:**

a) Was well done by most. Candidates were clearly familiar with autocorrelation function calculation and stationarity definitions. b) Likewise, many good attempts at these correlation functions. c) Most had a reasonable attempt at this, but very few actually got the right coefficients. d) Many got the right idea here, but few got the right answer or discussed the result.

(a) A discrete random process is generated as a sequence of independent time points, assigned a value of +1 or -1 with probabilities p_1 and $p_{-1} = 1 - p_1$, respectively. A possible sequence of bits generated from the process is, for example:

$$\{x_n\} = \{..., +1, +1, -1, +1, -1, +1, ...\}$$

Determine the mean and autocorrelation function for the process. Hence determine whether the process is wide-sense stationary. [30%]

SOLUTION:

$$E[X_n] = p_1 \times 1 + p_{-1} \times -1 = 2p_1 - 1$$

$$\begin{aligned} r_{XX}[n, n+l] &= E[X_n X_{n+l} \\ &= \begin{cases} 1^2 \times p_1 + (-1)^2 \times p_{-1} = 1, & l = 0 \\ 1^2 \times p_1^2 + 1 \times (-1) p_1 p_{-1} + (-1) \times 1 p_{-1} p_1 + (-1)^2 p_{-1}^2 = (2p_1 - 1)^2 & l \neq 0 \end{cases} \\ &= (1 - (2p_1 - 1)^2)\delta[l] + (2p_1 - 1)^2 \end{aligned}$$

Both are finite and clearly do not depend on n, so the process is WSS

ASSESSOR's comment:

This was mostly well handled by candidates.

(b) A noisy communications channel is applied to the random bit stream $\{x_n\}$ such that the received sampled data can be modelled as

$$y_n = 0.9x_n + 0.5x_{n-1} + v_n$$

where $\{v_n\}$ is zero mean white noise having variance $\sigma_v^2 = 0.1$ and which is independent of $\{x_n\}$ and it is known that $p_1 = p_{-1} = 0.5$.

Determine the autocorrelation function of the received process $\{y_n\}$ and also the cross-correlation function between $\{x_n\}$ and $\{y_n\}$.

SOLUTION:

With $p_1 = p_{-1}$ we have that $r_{XX}[l] = \delta[l]$ and $E[X_n] = 0$.

Now,

$$\begin{aligned} r_{YY}[l] &= E[Y_n Y_{n+l}] \\ &= E[(0.9x_n + 0.5x_{n-1} + v_n)(0.9x_{n+l} + 0.5x_{n-1+l} + v_{n+l})] \\ &= (0.81 + 0.25)r_{XX}[l] + r_{VV}[l] + 0.45(r_{XX}[l-1] + r_{XX}[l+1]) + \\ &\quad \text{cross-correlations between } X \text{ and } V \text{, all equal to zero} \\ &= 1.16\delta[l] + 0.45(\delta[l+1] + \delta[l-1]) \end{aligned}$$

since $r_{VV}[l] = 0.1\delta[l]$ and $r_{XX}[l] = \delta[l]$.

$$\begin{aligned} r_{XY}[l] &= E[x_n y_{n+l}] \\ &= E[x_n(0.9x_{n+l} + 0.5x_{n-1+l} + v_{n+l})] \\ &= 0.9r_{XX}[l] + 0.5r_{XX}[l-1] \\ &= 0.9\delta[l] + 0.5\delta[l-1] \end{aligned}$$

ASSESSOR's comment:

Again, good attempts from most candidates.

[30%]

(c) Hence determine the coefficients of an optimal filter for estimation of $\{x_n\}$ from $\{y_n\}$ according to the formula

$$\hat{x}_n = h_0 y_n + h_1 y_{n-1},$$

where h_0 and h_1 are to be determined according to the Wiener criterion.

SOLUTION:

Wiener Filter minimizes:

$$\begin{aligned} E[(x_n - \hat{x}_n)^2] &= E[(x_n - h_0 y_n - h_1 y_{n-1})^2] \\ &= E[x_n^2] - 2h_0 E[x_n y_n] - 2h_1 E[x_n y_{n-1}] + h_0^2 E[y_n^2] + h_1^2 E[y_{n-1}^2] + 2h_0 h_1 E[y_n y_{n-1}] \\ &= 1 - 2h_0 \times 0.9 - 2h_1 \times 0 + h_0^2 \times 1.16 + h_1^2 \times 1.16 + 2h_0 h_1 \times 0.45 \end{aligned}$$

where we have substitued correlation/ cross-correlations from part (b), being careful to observe that $E[x_n y_{n-1} = 0$.

Now, differentiate wrt h_0 and h_1:

$$\frac{\partial}{\partial h_0} = -1.8 + 2 \times 1.16 h_0 + 2h_1 \times 0.45$$
$$\frac{\partial}{\partial h_1} = 2 \times 1.16 h_1 + 2h_0 \times 0.45$$

Now, solving for the partial derivatives equal to zero we get:

$$\begin{bmatrix} 2.32 & 0.9 \\ 0.9 & 2.32 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} 0.913 \\ -0.354 \end{bmatrix}$$

ASSESSOR's comment:

Most people got the right general approach here, but few arrived at a plausible final solution.

[30%]

(d) Determine the mean-squared error corresponding to this Wiener filter and comment on the result.

SOLUTION:

Plugging in the estimated coefficients, the error is

$$1 - 2h_0 \times 0.9 - 2h_1 \times 0 + h_0^2 \times 1.16 + h_1^2 \times 1.16 + 2h_0 h_1 \times 0.45 = 0.1786$$

[could also calculate from the expression given in the notes forthe minimum error - either is fine and equivalent]

Comment: this is a pleasingly small error that should enable decoding of the bit sequence accurately.

ASSESSOR's comment:

Not many people made a serious attempt at this part...

[10%]

4 ASSESSOR's comment:

This year question 4 was popular, which is a pleasing shift compared to previous exams. a) Was handled well by most. b) Caused some confusion, in that many tried set the joint likelihood to be > 0.5, which is a harder problem, whereas the question requires the each of the two data points be classified with prob. > 0.5, a much easier problem. This error then made it difficult to find the parameter regions and the maximum likelihood value. d) Most had a vague idea of how Bayesian methods modifiy the situation, though few had enough knowledge to make the full 20

Consider a binary classification problem where the data $\mathscr{D}$ consists of N data points, $\mathscr{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, x_n is a real scalar and $y_n \in \{0, 1\}$, and the goal is to predict class labels y for new x.

Assume a very simple logistic classification model in which the class labels were produced independently and identically from the following model:

$$P(y_n = 1 | x_n, a, b) = \sigma(ax_n + b)$$

where σ is the logistic function, $\sigma(z) = \frac{1}{1+\exp(-z)}$, and a and b are the parameters of the classifier.

(a) Write down the likelihood of a and b for the data $\mathscr{D}$ and describe an algorithm to optimise this likelihood as a function of a and b. [40%]

(b) Consider a data set consisting of only two data points, $\mathscr{D} = \{(-2,0), (3,1)\}$. For this data set, describe the set of parameters which classify both data points correctly with probability greater than 0.5. Furthermore, what is the maximum achievable likelihood and describe the set of parameters achieving this maximum. [40%]

(c) Explain how Bayesian learning of the parameters might give more reasonable inferences about a and b from the data set in part (b) than maximum likelihood (ML) and how the Bayesian predictions about future labels differ from the ML predictions. [20%]

SOLUTION:

(a) The likelihood is

$$p(y_1,\ldots,y_N|x_1,\ldots,x_N,a,b) = \prod_{n=1}^{N} p(y_n|x_n,a,b) \quad (1)$$

$$= \prod_{n=1}^{N} \sigma(ax_n+b)^{y_n}(1-\sigma(ax_n+b))^{1-y_n} \quad (2)$$

An algorithm to optimise this is steepest gradient ascent in the likelihood which takes derivatives of the log likelihood and moves parameters in the direction of these derivatives (Pattern Processing Lecture 3, slide 8).

[Optional but not required for full marks: the steps for the batch version of this algorithm with step size *eta* are:

$$a^{[t+1]} = a^{[t]} + \eta \sum_n (y_n - \sigma(a^{[t]}x_n + b^{[t]}))x_n \quad (3)$$

$$b^{[t+1]} = b^{[t]} + \eta \sum_n (y_n - \sigma(a^{[t]}x_n + b^{[t]})) \quad (4)$$

]

(b) To classify both data points correctly (with prob. > 0.5) we require that

$$-2a+b < 0 \quad (5)$$

$$3a+b > 0 \quad (6)$$

Negating the first inequality and adding the two we find that

$$a > 0$$

and subsequently solving for b we get that

$$-3a < b < 2a.$$

The region of (a,b) parameter space defined by these two ineqalities gives us correct classification. The larger the value of a (for b satisfying the above constraint), the higher the likelihood, since the σ function increases monotonically to 1. In the limit of $a \to \infty$ the likelihood is 1. So the maximum achievable likelihood is 1, and this occurs when $a \to \infty$ as long as b stays in $-3a < b < 2a$.

(c) In part (b), because the data is linearly separable, the ML parameters can go to infinity, resulting in a very sharp and confident classification boundary. If you put a prior on (a,b) the posterior will put more mass in the same region defined as in part (b), but still reflect a reasonable amount of uncertainty about what a and b should be. By averaging over the posterior we get predictions that are not very confident; this makes sense, since we've only observed two data points.

END OF PAPER