PART IIA 2007
3F3 Signal and pattern processing
Prof S Godsill

May 2007

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 2 - 29 May 2007

1 (a) In a digital communication system random bits are transmitted as a Bernoulli random process, that is, each time point is independently assigned a value of +1 or -1 with equal probabilities of 0.5 . A possible sequence of bits generated from the process is, for example:

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

Explain intuitively the meaning of stationarity for a discrete time random process. Define wide-sense stationarity. Show that the above random process is wide sense stationary. [30%]

Solution:

Stationarity means that the statistical characteristics of the process do not change with time. Intuitively, if you observe a process starting at a particular time then there would be no difference in the characteristics of an observation started at an entirely different time.

Wide sense stationarity: A random process is *wide-sense stationary (WSS)* if:

(i) $\mu_n = E[X_n] = \mu$, (mean is constant)

(ii) $r_{XX}[n,m] = r_{XX}[m-n]$, (autocorrelation function depends only upon the difference between n and m).

(iii) The variance of the process is finite:

$$E[(X_n - \mu)^2] < \infty$$

Wide-sense stationarity for a random process

For Bernoulli process:

(i) $E[X_n] = 0.5 * -1 * 0.5 * -1 = 0$, (mean is constant)

(ii) $r_{XX}[n,m] = E[X_n X_m] = E[X_n]E[X_m] = 0$ for $m \neq n$, since process is independent. $r_{XX}[n,n] = 0.5 * (1)^2 + 0.5 * (-1)^2 = 1$

(iii) The variance of the process is finite:

$$E[(X_n - \mu)^2] = 1 < \infty$$

Hence wide-sense stationary.

(b) In the communications channel the bits are distorted according to a FIR filter,

$$x_n = \sum_{i=0}^{1} c_i b_{n-i}$$

where $c_0 = 1$ and $c_1 = 0.1$. Determine the cross-correlation function between $\{b_n\}$ and $\{x_n\}$, and also the autocorrelation function of $\{x_n\}$. [40%]

Solution:

$$
\begin{aligned}
E[b_n x_{n+m}] &= E[b_n \sum_{i=0}^{1} c_i b_{n+m-i}] \\
&= \sum_{i=0}^{1} c_i E[b_n b_{n+m-i}] \\
&= \sum_{i=0}^{1} c_i E[b_n b_{n+m-i}] \\
&= \begin{cases} c_0, & m = 0 \\ c_1, & m = 1 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

[i.e. the filter coefficients themselves]

$$
\begin{aligned}
E[x_n x_{n+m}] &= E[\sum_{i=0}^{1} c_i b_{n-i} \sum_{j=0}^{1} c_j b_{n+m-j}] \\
&= \sum_{i=0}^{1} \sum_{j=0}^{1} c_i c_j E[b_{n-i} b_{n+m-j}] \\
&= \sum_{i=0}^{1} c_i c_{m-i} \\
&= ...0, 0.1\, 1.01, 0.1, 0, ...
\end{aligned}
$$

[i.e. the convolution of the filter impulse responses]

(c) It is desired to optimally estimate the bit sequence $\{b_n\}$ from the channel data $\{x_n\}$. Design the second order FIR Wiener filter for this task, i.e. form an estimate of the type:

$$\hat{b}_n = \sum_{i=0}^{1} h_i x_{n-i}$$

where h_0 and h_1 are to be determined according to the Wiener criterion. [30%]

Solution:

(TURN OVER for continuation of Question 1

Require to minimise:

$$
\begin{aligned}
E[(b_n - \hat{b}_n)^2] &= E[(b_n - \sum_{i=0}^{1} h_i x_{n-i})^2] \\
&= E[(b_n - \mathbf{h}^T \mathbf{x}_n)^2] \\
&= E[b_n^2] + \mathbf{h}^T E[\mathbf{x}_n \mathbf{x}_n^T]\mathbf{h} - 2E[b_n \mathbf{h}^T \mathbf{x}_n]
\end{aligned}
$$

Differentiate wrt **h** and equate to zero:

$$2E[\mathbf{x}_n \mathbf{x}_n^T]\mathbf{h} = 2E[b_n \mathbf{x}_n]$$

hence

$$\mathbf{h} = \mathbf{R}_x^{-1} \mathbf{r}_{bx}$$

In this case:

$$\mathbf{R}_x = E[\mathbf{x}_n \mathbf{x}_n^T] = \begin{bmatrix} 1.01 & 0.1 \\ 0.1 & 1.01 \end{bmatrix}$$

$$\mathbf{r}_{bx} = E[b_n \mathbf{x}_n] = \begin{bmatrix} 1 \\ 0.1 \end{bmatrix}$$

and

$$\mathbf{h} = \mathbf{R}_x^{-1} \mathbf{r}_{bx} = \begin{bmatrix} 0.99 \\ 0.001 \end{bmatrix}$$

2 (a) Describe the principal means for reduction of errors in fixed precision digital filter implementations. Your description should include a discussion of overflow, limit cycles, saturation arithmetic and scaling. [30%]

Solution:

Bookwork - describe overflow and the use of saturation arithmetic to ensure that wrap around does not occur. Also, briefly describe the three main types of scaling (l1, l2, frequency-response), plus the occurrence of limit cycles.

(b) An IIR digital filter has the following transfer function:

$$H(z) = \frac{1 - 0.6z^{-1}}{1 - 0.9z^{-1}}$$

The filter is to be implemented in direct form using 16-bit fixed point arithmetic.

(i) Determine appropriate scalings to ensure overflow does not occur when the input signals are sine waves and sketch this implementation.

Solution:

For sine-wave inputs, require frequency response scaling.

Consider first stage (all-pole) section:

$$H(z) = \frac{1}{1 - 0.9z^{-1}}$$

which has a pole at $z = 0.9$.

Now, frequency response is:

$$H(exp(j\Omega)) = \frac{1}{1 - 0.9\exp(-j\Omega)}$$

which clearly has maximum magnitude when $\Omega = 0$ and a value of $1/(1 - 0.9) = 10$.

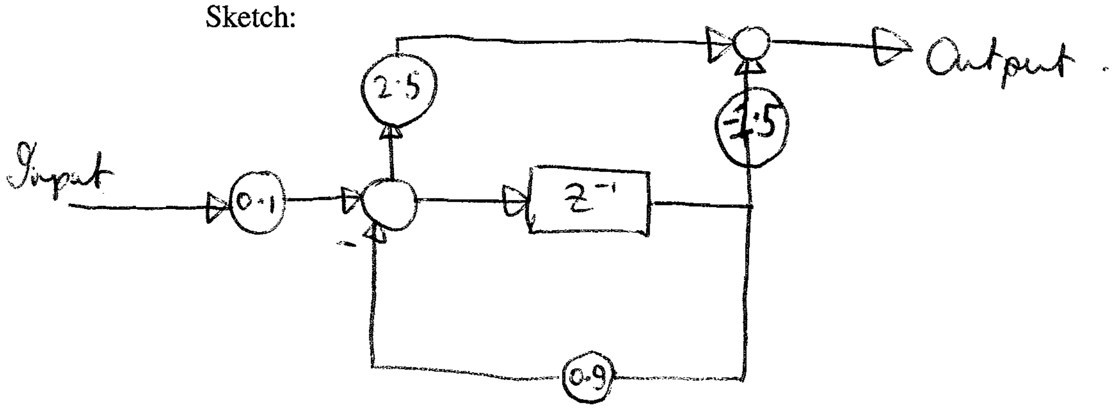
Hence the input should be scaled by a factor of $1/10$ (or 1/16 if binary shifting used). This ensures the internal signal does not overflow.

Now, consider the entire filter, including the scaling just computed (use 1/10 - 1/16 also acceptable):

$$H(z) = 0.1\frac{1 - 0.6z^{-1}}{1 - 0.9z^{-1}}$$

This has in addition a zero at 0.6. Thus overall the response is still at a maximum when $\Omega = 0$, when the gain is 0.1*0.4/0.1=0.4. Hence can scale the two FIR coefficients by 1/0.4=2.5 without overflow, resulting in $[2.5 - 1.5]$.

Sketch:



The input signal is assumed bounded between -1 and +1 and and overflow occurs if the magnitude of the output or the internal signal exceeds 1. [30%]

(c) If the input to the above filter, with a scaling calculated as in part (b)(ii) above, is Gaussian white noise with variance equal to 10, determine:

(i) the mean value at the output of the filter

Solution:

$E[X_n] = E[\sum_n h_m X_{n-m}] = 0$

(ii) the power spectrum at the output of the filter,

Solution:

Power spectrum is:

$$S_X(\exp(j\Omega)) = 10 * 0.25 \frac{(|1-0.6\exp(-j\Omega)|^2)}{|1-0.9\exp(-j\Omega)|^2}$$

(iii) the mean-squared signal value at the output of the filter, and hence

Solution:

The impulse response (including the scaling of 0.25) is:

$$h_n = 0.25(0.9^n - H(n-1) * 0.6 * (0.9)^{n-1})$$

and the sum of squares is:

$$\sum_n h_n^2 = 0.25^2(1 + 0.3^2(1/(1-0.81))) = 0.0921$$

Average Power (mean-sqaured value) at output is:

$$E[x_n^2] = \sum_n h_n^2 \sigma_v^2$$
$$= 0.0921 * 10 = 0.921$$

(iv) the probability that overflow occurs at any given sample time (assuming that the effects of any previous overflows have died away).

Solution: Output is a zero mean Gaussian random variable with variance 0.921. Hence probability of overflow is:

$$2 * \int_1^\infty \frac{1}{\sqrt{2\pi 0.921}} \exp(-1/2 * 0.921x^2) dx = 2 * (1 - 0.8612) = 0.28$$

(from tables).

[40%]

3 (a) Describe the steps involved in the window method of digital filter design. Explain its advantages and drawbacks when compared with the bilinear transform method of design. [30%]

Solution:

There are 5 steps in the window design method for FIR filters:

Select a suitable window function. Specify an 'ideal' response $D(\Omega)$. Compute the coefficients of the ŞidealŤ filter by inverse DTFT. Multiply and truncate the ideal coefficients by the window function to give the filter coefficients. Evaluate the frequency response of the resulting filter, and iterate if necessary to get the desired response.

Advantages: stable FIR filter, simple to design, doesn't warp frequency axis, generates linear phase.

Disadvantages: not optimal design, iterative, can't guarantee particular performance in various bands (stop band atten., pass band atten, etc.). ă

(b) It is desired to design a lowpass digital filter having frequency response $D(\Omega)$. The ideal impulse response of the filter, d_n is determined as the inverse DTFT of $D(\Omega)$:

$$d_n = \frac{1}{2\pi}\int_{-\pi}^{+\pi} D(\Omega)\exp(+j\Omega n)d\Omega$$

If d_n is truncated by multiplication with a finite duration window function w_n, show that the resulting filter's frequency response is

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

where $W(\exp(j\Omega))$ is the DTFT of the window function. Use this formula to explain qualitatively the effect on the filter's frequency response of truncating the ideal impulse response in this way. [30%]

Solution:

Take the DTFT of the windowed response $d_w = w_n d_n$ directly:

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

(TURN OVER for continuation of Question 3

We see that the spectrum of the windowed signal is the convolution of the infinite duration signal spectrum and the window spectrum.

(c) Instead of calculating d_n exactly using the inverse DTFT as above, the coefficients are estimated using the inverse DFT of a sampled version of $D(\Omega)$, as follows:

$$\hat{d}_n = \frac{1}{N}\sum_{p=0}^{N-1} D\left(\frac{2\pi p}{N}\right)\exp\left(\frac{j2\pi np}{N}\right)$$

Show that the resulting coefficients are related to the ideal coefficients by the following result:

$$\hat{d}_n = \sum_{m=-\infty}^{+\infty} d_{n-mN}$$

Explain how to reduce the effects of this approximation in a practical implementation of the window method of filter design. [40%]

Solution:

$$\begin{aligned}
\hat{d}_n &= \frac{1}{N}\sum_{p=0}^{N-1} D\left(\frac{2\pi p}{N}\right)\exp\left(\frac{j2\pi np}{N}\right) \\
&= \frac{1}{N}\sum_{p=0}^{N-1}\sum_{m=-\infty}^{+\infty} d_m \exp\left(-jm\left(\frac{2\pi p}{N}\right)\right)\exp\left(\frac{j2\pi np}{N}\right) \\
&= \frac{1}{N}\sum_{m=-\infty}^{+\infty} d_m \sum_{p=0}^{N-1}\exp\left(\frac{j2\pi p(n-m)}{N}\right) \\
&= \frac{1}{N}\sum_{m=-\infty}^{+\infty} d_m N\delta[n-m-kN] \\
&= \sum_{m:n-m=kN} d_m \\
&= \sum_{k=-\infty}^{+\infty} d_{n-kN}
\end{aligned}$$

as required (using result below for summation).

This error may be reduced by increasing the number of points in the DFT using zero padding, hence reducing overlap between the repetitions of d_{n-kN}.

[You may use the following result, which applies for integers m and k:

$$\sum_{p=0}^{N-1}\exp(j2\pi mp/N) = \begin{cases} N, & \text{for } m = kN \\ 0 & \text{otherwise} \end{cases}$$

]

4 Assume you want to build an automatic berry classification machine which, based on the measured weight of the berry, x, classifies it into one of three classes, "Strawberry", "Raspberry" and "Cranberry", denoted $Y = s, Y = r$ and $Y = c$ respectively. Your goal is to compute $P(Y|x)$. Assume that $P(Y = s) = 0.5$, $P(Y = r) = 0.3$ and $P(Y = c) = 0.2$, which are obtained by measuring the observed frequencies of these berries.

(a) We assume that $P(x|Y = s)$ is Gaussian with mean 4 and variance 1; $P(x|Y = r)$ is Gaussian with mean 2 and variance 1; and $P(x|Y = c)$ is Gaussian with mean 1 and variance 1. Compute the region of x for which the Raspberry class is more probable than the other two classes given x. Show all working, and sketch $P(Y = r|x)$. [30%]

Solution:

$$\begin{aligned} P(Y = i|x) &\propto P(x)p(x|Y = i) \\ &= N(x|\mu_i, \sigma_i)P_i \end{aligned}$$

Need to find place where:

$$0.2exp(-1/2(x_l - 1)^2) = 0.3exp(-1/2(x_l - 2)^2)$$

i.e., solving,

$$x_l = 1.0945$$

Similarly for right hand edge:

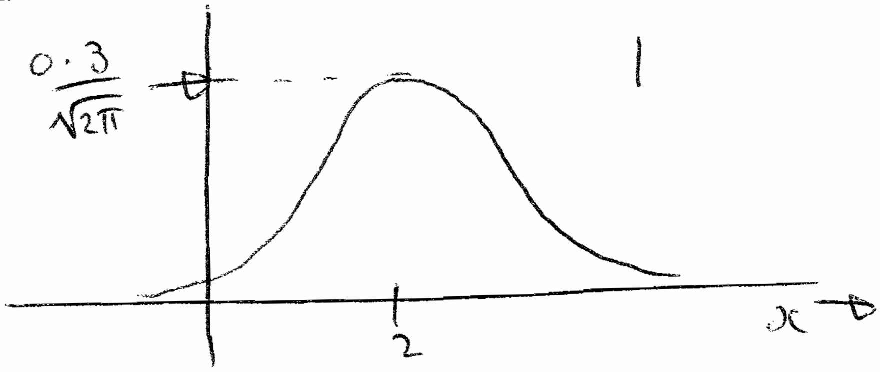
$$0.3exp(-1/2(x_r - 2)^2) = 0.5exp(-1/2(x_r - 4)^2)$$

giving,

$$x_r = 2.74$$

(TURN OVER for continuation of Question 4

Sketch:



(b) Given the following data set of 6 points and their class labels:

$$\mathscr{D} = \{(1.0,c),(2.1,c),(1.0,r),(3.0,r),(2.0,s),(4.0,s)\},$$

find the maximum likelihood parameter estimates of the means and variances of Gaussians fitting each class. [30%]

Solution:

Need $p(x|Y=i) = N(\mu_i, \sigma_i^2)$.

Take e.g. class c:

$$\sum_i log(p(x_i|Y=c)) = -log(2\pi\sigma_c^2) - 0.5(1-\mu_c)^2/\sigma_c^2 - 0.5(2-\mu_c)^2/\sigma_c^2$$

Take derivative wrt μ_c:

$$\frac{\partial log(p(x_i|Y=c))}{\partial \mu_c} = (1-\mu_c)/\sigma_c^2 + (2-\mu_c)/\sigma_c^2$$

Equate to zero:

$$3 - 2\mu_c = 0, \ \mu_c = 3/2$$

Similarly for σ_c^2:

$$\frac{\partial log(p(x_i|Y=c))}{\partial \sigma_c^2} = -1/(\sigma_c^2) + (0.5(1-\mu_c)^2 + 0.5(2-\mu_c)^2))/(\sigma_c^2)^2$$

Equate to zero and solve:

$$\sigma_c^2 = 0.5(1-\mu_c)^2 + 0.5(2-\mu_c)^2) = 0.25$$

Similarly for the others:

$$\mu_r = 2, \sigma_r^2 = 1, \mu_s = 3, \sigma_s^2 = 1$$

(c) Consider the problem of finding the means of the above 3 Gaussians which maximise the probability of the observed classes for the data set $\mathscr{D}$. The parameters of this model are $\theta = (\mu_s, \mu_r, \mu_c)$. Assume the variances are all set to one, and the observed class frequencies are set to $\pi_s = 0.5$, $\pi_r = 0.3$, $\pi_c = 0.2$ as above. We can write

$$P(Y|x,\theta) = \frac{\pi_Y e^{-(x-\mu_Y)^2/2}}{\pi_s e^{-(x-\mu_s)^2/2} + \pi_r e^{-(x-\mu_r)^2/2} + \pi_c e^{-(x-\mu_c)^2/2}}$$

and our goal is to maximize the probability of the observed classes for the above data set $\mathscr{D}$. Describe a method or algorithm for doing this. Would this algorithm give the same means as in part (b) of this question? Explain. [40%]

Solution:

Compute $L(\theta) = \prod_{n=1}^{6} P(y_n|x_n,\theta)$ and find the derivative $\frac{\partial L}{\partial \theta}$, then use gradient ascent:

$$\theta_{t+1} = \theta_t + \mu \frac{\partial L}{\partial \theta}$$

Would not give same result in general as (b) since we are maximising a different quantity - the probability of getting a correct label rather than the probability of X for each class.

END OF PAPER