Version SJG/4

# Worked Solutions

EGT2
ENGINEERING TRIPOS PART IIA

---

Wednesday 7 May 2025    9.30 to 11.10

---

**Module 3F3**

**STATISTICAL SIGNAL PROCESSING**

*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
Engineering Data Book

**10 minutes reading time is allowed for this paper at the start of the exam.**

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

**You may not remove any stationery from the Examination Room.**

1 The vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots$ follow a first order vector autoregressive process defined by

$$\boldsymbol{X}_{t+1} = A\boldsymbol{X}_t + \boldsymbol{W}_{t+1}$$

where $A$ is a symmetric matrix, $\boldsymbol{W}_t$ are iid multivariate normal random variables with

$$E[\boldsymbol{W}_t] = \boldsymbol{0}$$

$$E\left[\boldsymbol{W}_t \boldsymbol{W}_t^T\right] = I$$

where $\boldsymbol{0} = (0, 0, \ldots, 0)^T$ is the vector of zeros and $I$ is the identity matrix.

(a) Assuming that $\boldsymbol{X}_0 = \boldsymbol{0}$, prove that for $t > 0$

$$\boldsymbol{X}_t = \sum_{s=0}^{t-1} A^s \boldsymbol{W}_{t-s}$$

and thus argue that $\boldsymbol{X}_t$ is multivariate normal. [25%]

---

**Solution:** We can show this by induction. The base case $X_1 = A^0 W_{1-0} = I W_1 = W_1$ holds. Assume the relation for $t$, then for $t+1$ we have

$$\begin{aligned} X_{t+1} &= A \sum_{s=0}^{t-1} A^s W_{t-s} + W_{t+1} \\ &= \sum_{s=1}^{t} A^s W_{t+1-s} + W_{t+1} \\ &= \sum_{s=0}^{t} A^s W_{t+1-s} \end{aligned}$$

as required.

The sum of independent multivariate normals is multivariate normal, hence $X_t$ is multivariate normal.

---

(b) Show that $E[\boldsymbol{X}_t] = \boldsymbol{0}$. [10%]

---

**Solution:** By linearity of expectation

$$E[X_t] = E[\sum_{s=0}^{t-1} A^s W_{t-s}] = \sum_{s=0}^{t-1} A^s E[W_{t-s}]$$

and, because $E[W_s] = 0$ for all $s$, we have $E[X_t] = \sum_{s=0}^{t-1} A^s \boldsymbol{0} = \boldsymbol{0}$.

---

(c) Let $\boldsymbol{Y}_t = A^t \boldsymbol{W}_t$. Show that $E[\boldsymbol{Y}_t \boldsymbol{Y}_t^T] = A^{2t}$. [20%]

**Solution:**

$$\begin{aligned} E[Y_t Y_t^T] &= E[A^t W_t W_t^T (A^t)^T] \\ &= E[A^t W_t W_t^T A^t] && \text{(by symmetry of } A) \\ &= A^t E[W_t W_t^T] A^t && \text{(by linearity of } E) \\ &= A^t I A^t \\ &= A^{2t} \end{aligned}$$

(d) Show that $E\left[\boldsymbol{X}_t \boldsymbol{X}_t^T\right] = \sum_{s=0}^{t-1} A^{2s}$. [20%]

**Solution:** From the expression of $X_t$ in part (a), and using the linearity of expectation and the symmetry of $A$, we have

$$E[X_t X_t^T] = \sum_{s=0}^{t-1} \sum_{r=0}^{t-1} A^s E[W_{t-s} W_{t-r}^T] A^r$$

Additionally, $E[W_{t-s} W_{t-r}^T] = \delta_{r,s} I$ since $W_t$ are independent. This leads to

$$\begin{aligned} E[X_t X_t^T] &= \sum_{s=0}^{t-1} \sum_{r=0}^{t-1} \delta_{r,s} A^s I A^r \\ &= \sum_{s=0}^{t-1} A^{2s} \end{aligned}$$

(e) Consider the distribution for $\boldsymbol{X} = \lim_{t \to \infty} \boldsymbol{X}_t$.

(i) Show that the covariance of $\boldsymbol{X}$ can be written $\left(I - A^2\right)^{-1}$ (hint: recall the expression for a Geometric Progression).

(ii) What happens if $A$ has any eigenvalues with magnitude larger than 1?

[25%]

**Solution:** We have

$$E[XX^T] = \lim_{t \to \infty} \sum_{s=0}^{t-1} A^{2s}$$

Assuming that this sum converges, we have

$$E[XX^T](I - A^2) = \sum_{s=0}^{\infty} A^{2s} - \sum_{s=1}^{\infty} A^{2s} = A^0 = I$$

hence, $E[XX^T]$ is the inverse of $(I - A^2)$, i.e., $E[XX^T] = (I - A^2)^{-1}$.

If $A$ has eigenvalues larger than 1 in magnitude, the sum will not converge. The consequence of this is that the variance of the process will diverge and there will not be a well defined stationary distribution in the limit of infinite time.

---

**Examiner's comment:** A very popular question, answered by almost every candidate. This question was a straightforward generalisation of material from the lectures and there were a lot of good and very good answers. Part (a): several candidates simply wrote out the expression for t=1 and t=2, followed by "..." and then the expression given in the question for general t. In part (c) and (d), several candidates assumed (without any comment) that matrices commute or made no note of potential cross terms. In part (e)ii, some answers didn't relate the diverging sum to the process in any way, or incorrectly stated the covariance would be negative.

2 Let $X_t$ be a discrete random walk. Initially $X_0 = 0$. Then, at each subsequent step, we move to the left or to the right with equal probability, i.e.,

$$X_{t+1} = \begin{cases} X_t + 1 & \text{with probability } 1/2 \\ X_t - 1 & \text{with probability } 1/2 \end{cases}$$

(a) After $2t$ steps, the random walk will be at the origin if the same number of steps to the right have been made as to the left. Hence, show that the probability that $X_{2t} = 0$ is

$$P(X_{2t} = 0) = \frac{(2t)!}{2^{2t}(t!)^2}$$

[20%]

---

**Solution:** Out of $2t$ steps, the probability of taking $t$ steps to the left and $t$ steps to the right is the binomial probability, $\binom{n}{k}p^n(1-p)^{n-k}$, with $p = 1/2$, $n = 2t$ and $k = t$. This gives

$$\binom{2t}{t}\left(\frac{1}{2}\right)^t\left(1-\frac{1}{2}\right)^{2t-t} = \frac{(2t!)}{t!(2t-t)!}\frac{1}{2^{2t}} = \frac{(2t!)}{2^{2t}(t!)^2}$$

---

(b) Stirling's approximation for the factorial is

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Use this to show that the probability that $X_{2t} = 0$ is approximately proportional to a power of $t$, i.e.,

$$P(X_{2t} = 0) \propto t^{-\alpha},$$

and give the value of $\alpha$. [20%]

---

**Solution:**

$$P(X_{2t} = 0) = \frac{\sqrt{4\pi t}\left(\frac{2t}{e}\right)^{2t}}{2^{2t}2\pi t\left(\frac{t}{e}\right)^{2t}} = \frac{2\sqrt{\pi t}}{2\pi t} = \frac{1}{\sqrt{\pi}}t^{-1/2}$$

so $\alpha = 1/2$.

---

(c) Show that the expected number of returns to the origin is infinite. (It will be useful to recall that the series $\sum_{k=1}^{\infty} k^{-\alpha}$ is finite for $\alpha > 1$ and diverges for $\alpha \leq 1$). [20%]

---

**Solution:** The walk has returned at time $2t$ if $X_{2t} = 0$. The total number of returns for a particular walk is

$$\sum_{t=1}^{\infty} \delta_{X_{2t},0}$$

the expected value of the number of returns to the origin is thus

$$E[\sum_{t=1}^{\infty} \delta_{X_{2t},0}] = \sum_{t=1}^{\infty} E[\delta_{X_{2t},0}] = \sum_{t=1}^{\infty} P(X_{2t} = 0) \approx \frac{1}{\sqrt{\pi}} \sum_{t=1}^{\infty} t^{-1/2} = \infty$$

because the sum diverges (as per the hint).

---

(d) A random walk in 2 dimensions can be modelled as two independent random walks, $X_t$ and $Y_t$ for the $x$ and $y$ coordinates respectively. The random walk has returned to the origin at time $2t$ only if $X_{2t} = Y_{2t} = 0$. Show that in 2 dimensions, the expected number of returns to the origin is also infinite. [20%]

---

**Solution:** The two walks are independent, so the expectation of the event that both of them are at the origin is $E[X_{2t} = Y_{2t} = 0] = E[X_{2t} = 0]E[Y_{2t} = 0] = P(X_{2t} = 0)P(Y_{2t} = 0) \approx \frac{1}{\pi} t^{-1}$ hence the expected number of returns is (approximately)

$$\frac{1}{\pi} \sum_{t=1}^{\infty} t^{-1} = \infty$$

which still diverges, and hence the total number of returns to the origin is infinite.

---

(e) Repeat the calculation for 3 dimensions and comment on the result. [20%]

---

**Solution:** In three dimensions, we have three independent random walks and so $P(X_{2t} = 0, Y_{2t} = 0, Z_{2t} = 0) \approx \left(\frac{1}{\sqrt{\pi}} t^{-1/2}\right)^3$. Taking the sum over all $t$ we have that the expected number of returns is

$$\frac{1}{\pi^{3/2}} \sum_{t=1}^{\infty} t^{-3/2} < \infty$$

and so a random walk in three dimensions is only expected to return to the origin a finite number of times.

Equivalently, there is a finite probability of a random walk in 3 dimensions never returning to the origin, even if we wait for an infinite length of time. This result also clearly generalizes to any dimension greater than 3.

The upshot is that the Markov process is transient in 3 or more dimensions, even though it was recurrent in 1 and 2 dimensions.

---

**Examiner's comment:** The second most popular question, and a high average score. Nevertheless, there was a very large standard deviation. Parts (a) and (b) were fairly straightforward, simply needing the Binomial distribution and algebra. Parts (c)-(e) were well answered by a majority of candidates, hence the high mean mark. However, a substantial minority of answers were confused about how to set up the relevant expectation, often including extra factors of t or multiple sums. Some other poor answers provided absolutely no justification or comment for the result, giving the impression that they simply guessed an answer based on provided hints.

3 It is desired to estimate a random process $X_n$ formed as an AR process:

$$X_n = \alpha X_{n-1} + 0.2W_n \, .$$

Observations are made as

$$Y_n = X_n + 0.3V_n$$

where $\{W_n\}$ and $\{V_n\}$ are zero mean, mutually uncorrelated white noise random processes with unity variance.

(a) Determine the autocorrelation function of $Y_n$ and find the range of values of $\alpha$ for which $Y_n$ is wide-sense stationary. [25%]

---

Solution:

Mean (needs to be const. for WSS):

$$E[X_n] = \alpha E[X_{n-1}] + E[0.2W_n] = \alpha E[X_{[}n-1]]$$

Clearly mean can only be constant for $\alpha = 1$ (this case later excluded), or $E[X_n] = 0$.

For $k > 0$,

$$R_{XX}[n, n+k] = E[X_n(\alpha X_{n-1+k} + 0.2W_{n+k})] = \alpha R_{XX}[n, n+k-1]$$

But

$$R_{XX}[n,n] = E[X_n^2] = E[(\alpha X_{n-1}+0.2W_n)(\alpha X_{n-1}+0.2W_n)] = \alpha^2 R_{XX}[0]+0.2^2 E[W_n^2] = \alpha^2 R_{XX}[n,n]+0.$$

So

$$R_{XX}[n,n] = 0.04/(1-\alpha^2)$$

which is fixed over $n$.

Finally then, the ACF is stationary and symmetric, and hence given by:

$$R_{XX}[k] = \alpha^{|k|} R_{XX}[0]$$

As $|\alpha| \to 1$ variance goes to $\infty$, and $|\alpha| > 1$ leads to negative variance (undefined), so for WSS we require $|\alpha| < 1$.

We have then, since $Y_n = X_n + 0.3V_n$:

$$R_{YY} = R_{XX} + 0.3^2 R_{VV} = 0.04\alpha^{|k|}/(1-\alpha^2) + 0.09\delta_k$$

Same WSS condition as for $X$.

---

(b) Under the condition for wide-sense stationarity, show that the power spectrum of $\{X_n\}$ is given by

$$\frac{0.04}{|1-\alpha\exp(-j\Omega)|^2}$$

[15%]

---

Solution:

AR model $X_n$ is white noise $0.2V_n$ driving a filter $H(\exp(j\Omega)) = \frac{1}{(1-\alpha\exp(-j\Omega)}$.

Linear system result states that

$$\mathcal{S}_X(\Omega) = 0.2^2\mathcal{S}_V|H(\exp(j\Omega))|^2 = \frac{0.04}{|(1-\alpha\exp(-j\Omega)|^2}$$

---

(c) For $\alpha$ within the stationary range, show that the optimal Wiener Filter for estimating $X_n$ from the observations $, ...., Y_0, ...., Y_n, ...$ has frequency response:

$$H(\exp(j\Omega)) = \frac{A}{(1-B\cos\Omega)}$$

where $A$ and $B$ are constants that should be determined in terms of $\alpha$. Sketch the filter frequency response over the range $0 \leq \Omega < 2\pi$ for $\alpha = -0.9$. [25%]

---

Solution:

Optimal Wiener filter is:

$$H(\exp(j\Omega)) = \frac{\mathcal{S}_X}{\mathcal{S}_Y}$$

$$\mathcal{S}_X = DTFT\{\alpha^{|k|}R_{XX}[0]\} = \sum_{k=-\infty}^{+\infty}\alpha^{|k|}R_{XX}[0]\exp(-jn\Omega)$$

But easier to use Linear System result:

$$\mathcal{S}_X() = \frac{0.04}{|(1-\alpha\exp(-j\Omega)|^2}$$

and

$$\mathcal{S}_Y() = \frac{0.04}{|(1-\alpha\exp(-j\Omega)|^2} + 0.09$$

So

$$H(\exp(j\Omega)) = \frac{0.04}{0.04 + 0.09|(1 - \alpha \exp(-j\Omega)|^2)} = \frac{0.04}{0.13 - 0.18\alpha \cos\Omega + 0.09\alpha^2} = \frac{0.04/(0.13 + 0.0}{1 - 0.18\alpha \cos(\Omega)/(0.1}$$

So

$$A = 0.04/(0.13 + 0.09\alpha^2), \quad B = 0.18\alpha/(0.13 + 0.09\alpha^2)$$

---

(d) Show that the impulse response of this filter has the form

$$h_n = C\beta^{|n|},\ n = -\infty, ...., +\infty$$

giving equations that express $A$ and $B$ in terms of $C$ and $\beta$. [20%]

---

Solution:

Probably easiest to work from the proposed $h_n$:

$$DTFT\{\beta^{|n|}\} = \frac{(1-\beta^2)}{|1 - \beta \exp(-j\Omega)|^2} = \frac{(1-\beta^2)}{(1+\beta^2) - 2\beta\cos(\Omega)} = \frac{(1-\beta^2)/(1+\beta^2)}{1 - 2\beta/(1+\beta^2)\cos(\Omega)}$$

(do this from 1st principles by geometric progressions from $0 \rightarrow \infty$ and $-1 \rightarrow -\infty$).

So we can identify:

$$B = 2\beta/(1+\beta^2)$$

$$A = C(1+\beta^2)/(1-\beta^2)$$

---

(e) Can this optimal filter be implemented on-line in a computer programme? If so, then explain how. If not, suggest a suitable alternative strategy for optimal filter design that could be implemented. In either case, there is no need to give a code implementation or the details of the alternative filter design. [15%]

---

Solution:

Solution:

No - the filter is non-causal, and also IIR.

Could be implemented using a causal FIR Wiener filter.

---

---

**Examiner's comment:**

A reasonably popular question, rather poorly answered by many. Part a) quite well answered by most, but a lot of confusion over the reasoning for WSS in this model. Part b) some spotted the simple and quick solution based on power spectrum at output of a linear system. Others derived directly with DTFT which took much more working. c) Most answered this OK – some algebra errors, generally well sketched. d) Poorly answered – required a careful DTFT calculation. e) an easy part very well answered by most students. A common error was to say that the filter could not be implemented because it is IIR, whereas non-causality is the issue here.

4 (a) It is desired to perform Bayesian estimation $\hat{\theta}$ of a random parameter $\theta$ as a function $\hat{\theta}(x)$ of the measured data $x$, by minimising the expected value of a cost function:

$$C(\hat{\theta}, \theta) \geq 0,$$

which satisfies $C(\theta, \theta) = 0$ for any value of $\theta$.

The likelihood function is $p(x|\theta)$ and the prior distribution is $p(\theta)$. Show that the optimal choice of estimator $\hat{\theta}$ satisfies the following condition, assuming that $C(\hat{\theta}, \theta)$ is differentiable:

$$\int_\theta \frac{\partial C(\hat{\theta}, \theta)}{\partial \hat{\theta}} p(x|\theta) p(\theta) d\theta = 0$$

where integration is over the full range of possible $\theta$ values. [30%]

---

Solution:

Expected cost is:

$$\int C(\hat{\theta}, \theta) p(\theta|x) d\theta$$

Stationary point at

$$\frac{\partial}{\partial \hat{\theta}} \int C(\hat{\theta}, \theta) p(\theta|x) d\theta = \int \frac{\partial}{\partial \hat{\theta}} C(\hat{\theta}, \theta) p(\theta|x) d\theta \propto \int \frac{\partial}{\partial \hat{\theta}} C(\hat{\theta}, \theta) p(x|\theta) p(\theta) d\theta = 0$$

Will be a minimum if, at the stationary point,

$$\frac{\partial^2}{\partial \hat{\theta}^2} \int C(\hat{\theta}, \theta) p(\theta|x) dx \propto \int \frac{\partial^2}{\partial \hat{\theta}^2} C(\hat{\theta}, \theta) p(x|\theta) p(\theta) d\theta > 0$$

---

(b) If minimum mean squared error (MMSE) is required of the estimator, show that the required solution is in the form

$$\hat{\theta} = \frac{1}{Z(x)} \int_\theta \theta p(x|\theta) p(\theta) d\theta$$

and specify $Z(x)$, which does not depend on $\theta$. [20%]

---

Solution:

SE cost is:

$$C(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

whose first partial derivative is $\hat{\theta} - \theta$.

Thus the optimal solution is at

$$\int (\hat{\theta} - \theta)p(x|\theta)p(\theta)d\theta = 0$$

Rearranging and multiplying through by $1/Z(x) = 1/p(x)$:

$$\int \hat{\theta}p(x|\theta)p(\theta)/p(x)d\theta = \hat{\theta}\int p(x|\theta)p(\theta)/p(x)d\theta = \hat{\theta} = \int \theta p(x|\theta)p(\theta)/p(x)d\theta$$

i.e. the standard posterior mean MMSE estimator.

---

(c) A new cost function is proposed with the form $C(\hat{\theta}, \theta) = 1 - U(\hat{\theta}, \theta)$ where

$$U(\hat{\theta}, \theta) = \begin{cases} 1, & |\hat{\theta} - \theta| < \epsilon \\ 0, & \text{Otherwise} \end{cases}$$

Show that in this case, for the posterior probability $p(\theta|x)$, the optimal estimator $\hat{\theta}$ must satisfy

$$p(\theta = \hat{\theta} + \epsilon|x) = p(\theta = \hat{\theta} - \epsilon|x)$$

and hence explain why, as $\epsilon$ becomes small, $\hat{\theta}$ approaches the Maximum *a posteriori* (MAP) estimator $\theta^{MAP}$. [25%]

---

Solution:

$$\frac{\partial}{\partial\hat{\theta}}\int C(\hat{\theta}, \theta)p(\theta|x)d\theta = \frac{\partial}{\partial\hat{\theta}}\int p(\theta|x)d\theta - \frac{\partial}{\partial\hat{\theta}}\int_{\hat{\theta}-\epsilon}^{\hat{\theta}+\epsilon} p(\theta|x)d\theta = 0 - \frac{\partial}{\partial\hat{\theta}}\int_{\theta-\epsilon}^{\theta+\epsilon} p(\theta|x)d\theta$$

Following the hint:

$$\frac{\partial}{\partial\hat{\theta}}\int_{\theta-\epsilon}^{\theta+\epsilon} p(\theta|x)d\theta = \frac{\partial}{\partial\hat{\theta}}\int_{0}^{\theta+\epsilon} p(\theta|x)d\theta - \frac{\partial}{\partial\hat{\theta}}\int_{0}^{\theta-\epsilon} p(\theta|x)d\theta = p(\hat{\theta}+\epsilon|x) - p(\hat{\theta}-\epsilon|x) = 0$$

and so the staationary point is

$$p(\hat{\theta}+\epsilon|x) = p(\hat{\theta}-\epsilon|x)$$

as required.

Clearly as $\epsilon \longrightarrow 0$ this will home in on the maximum of $p(\theta|x)$, i.e. the MAP estimator.

Some students in the exam differentiated $C$ directly to give $\delta$-functions and this method gave a neat (and correct) alternative solution.

---

[Hint: recall that the derivative of an integral is given by $\frac{d\int_0^{\phi} f(\theta)d\theta}{d\phi} = f(\phi)$ and assume that the probability density functions are smooth and continuous.]

(d) The joint probability density function for two random variables is given by

$$f_{\theta,x}(\theta, x) = \begin{cases} 6\theta, & 0 \le \theta \le 1;\ 0 \le x \le 1-\theta \\ 0, & \text{Otherwise} \end{cases}$$

Find the conditional (posterior) density $p(\theta|x)$ and hence determine the MAP estimator and the MMSE estimator for $\theta$ conditional upon data $x$. Your solution should include a sketch of the non-zero probability region for $(\theta, x)$. [25%]

---

Solution:

Sketch:

...

$p(\theta|x) = p(\theta, x)/p(x)$

Now,

$$p(x) = \int_0^{1-x} p(\theta, x)d\theta = \int_0^{1-x} 6\theta d\theta = 3[\theta^2]_0^{1-x} = 3(1-x)^2,\ x \in [0, 1]$$

So

$$p(\theta|x) = p(\theta, x)/p(x) = \begin{cases} 6\theta/(3(1-x)^2), & 0 \le \theta \le 1;\ 0 \le x \le 1-\theta \\ 0, & \text{Otherwise} \end{cases}$$

The largest possible $\theta$ value for a given $x$ is $1-x$. So, MAP estimate is $\hat{\theta} = (1-x)$.

The MMSE estimate is the posterior mean:

$$\int_0^{1-x} \theta p(\theta|x)d\theta = \int_0^{1-x} 2\theta^2/(1-x)^2 = 2[\theta^3]_0^{1-x}/(3(1-x)^2) = 2(1-x)^3/(3(1-x)^2)$$

---

**Examiner's comment:** A less popular question, but quite well answered. Parts a) – c) well handled in general. Part d) which combines basic probability calculations with Bayesian inference was poorly done by most – lots of slips in algebra/concepts and evidence of running out of time.

**END OF PAPER**