SOLUTION TO EXAM 2002

Problem 1

(a) false: e.g. take $Y \sim \mathcal{N}(0,1)$ and $X = \xi Y$, where ξ is an independent symmetric random sign. Then X is Gaussian (check its characteristic function) and $\mathbf{E}XY = \mathbf{E}\xi Y^2 = \mathbf{E}\xi = 0$. However the vector (X,Y) is not Gaussian, since e.g.

$$P(X+Y=0) = 1/2$$

i.e. linear combination of the two has an atom.

(b) true: if $\widehat{\mathbf{E}}(X|Y) = \alpha + \beta Y$ and its variance is nonzero, then $\beta \neq 0$ and hence $Y = (\widehat{\mathbf{E}}(X|Y) - \alpha)/\beta$. So Y is a linear function of a Gaussian random variable and hence itself is Gaussian.

(c) false: example similar to **(a)**, let $Y \sim \mathcal{N}(0,1)$ and $X = (\xi+1)Y$. Then

$$\mathbf{E}(X|Y) = Y\mathbf{E}(\xi+1) = Y$$

and so also (why?) $\widehat{\mathbf{E}}(X|Y) = Y$. Moreover Y is Gaussian (with positive variance) and hence also $\mathbf{E}(X|Y)$, however (X,Y) is still not, e.g.

$$P(X=0) = 1/2$$

(d) true by direct calculation (the case $\mathbf{E}Y^2 = 0$ is trivial)

$$\mathbf{E}X\widehat{\mathbf{E}}\big(X|Y\big) = \frac{\mathbf{E}XY}{\mathbf{E}Y^2}\mathbf{E}XY = 0 \quad \Longrightarrow \quad \mathbf{E}XY = 0$$

(e) false, in the example of **(a)** we have $\mathbf{E}(X|Y) = 0$ so that $\mathbf{E}(X|Y)$ and X are independent, however X and Y clearly depend, e.g.

$$\mathbf{E}(X^2|Y) = Y^2$$

(f) true, by direct calculation

$$\widehat{\mathbf{E}}\Big(\widehat{\mathbf{E}}\big(Z|X\big)\big|Y\Big) = Y\frac{\mathbf{E}\widehat{\mathbf{E}}\big(Z|X\big)Y}{\mathbf{E}Y^2} = Y\frac{\mathbf{E}\frac{\mathbf{E}ZX}{\mathbf{E}X^2}XY}{\mathbf{E}Y^2} =$$
$$Y\frac{\frac{\mathbf{E}ZX}{\mathbf{E}X^2}\alpha\mathbf{E}X^2}{\mathbf{E}Y^2} = Y\frac{\mathbf{E}Z(\alpha X+V)}{\mathbf{E}Y^2} = Y\frac{\mathbf{E}ZY}{\mathbf{E}Y^2} = \widehat{\mathbf{E}}(Z|Y)$$

(g) true: for any bounded function ψ

$$\mathbf{E}\psi(Y)\mathbf{E}\Big(\mathbf{E}(Z|X)|Y\Big) = \mathbf{E}\psi(Y)\mathbf{E}(Z|X) = \mathbf{E}\psi(\alpha X+V)\mathbf{E}(Z|X) =$$
$$\mathbf{E}\int\psi(\alpha X+s)\mathbf{E}(Z|X)dF_v(s) = \mathbf{E}\mathbf{E}\Big(Z\int\psi(\alpha X+s)dF_v(s)\big|X\Big) =$$
$$\mathbf{E}\Big(Z\int\psi(\alpha X+s)dF_v(s)\Big) = \mathbf{E}\mathbf{E}\Big(Z\psi(\alpha X+V)\big|X,Z\Big) =$$
$$\mathbf{E}\mathbf{E}\Big(Z\psi(Y)\big|X,Z\Big) = \mathbf{E}\psi(Y)Z$$

where $F_v(s) = \mathbf{P}(V \le s)$. The claim follows from the definition of conditional expectation.

Problem 2

(a) As usual we look for $G(Y_2^{n-1}; Y_n)$, such that

$$\mathbf{E}\Big(I(X_n = 1)h(Y_n)|Y_2^{n-1}\Big) = \mathbf{E}\Big(G(Y_2^{n-1}; Y_n)h(Y_n)|Y_2^{n-1}\Big) \tag{1}$$

The right hand side becomes

$$\mathbf{E}\Big(I(X_n = 1)h(Y_n)|Y_2^{n-1}\Big) =$$
$$\mathbf{E}\Big(I(X_n = 1)h(1 - X_{n-1})|Y_2^{n-1}\Big) = 1/2\big\{h(0)\pi_{n-1} + h(2)(1 - \pi_{n-1})\big\}$$

whereas the left hand side is

$$\mathbf{E}\Big(G(Y_2^{n-1}; Y_n)h(Y_n)|Y_2^{n-1}\Big) =$$
$$\mathbf{E}\Big(G(Y_2^{n-1}; X_n - X_{n-1})h(X_n - X_{n-1})|Y_2^{n-1}\Big) =$$
$$\mathbf{E}\Big(I(X_n = 1)G(Y_2^{n-1}; 1 - X_{n-1})h(1 - X_{n-1})|Y_2^{n-1}\Big) +$$
$$\mathbf{E}\Big(I(X_n = -1)G(Y_2^{n-1}; -1 - X_{n-1})h(-1 - X_{n-1})|Y_2^{n-1}\Big) =$$
$$1/2\Big(G(Y_2^{n-1}; 0)h(0)\pi_{n-1} + G(Y_2^{n-1}; 2)h(2)(1 - \pi_{n-1})\Big) +$$
$$1/2\Big(G(Y_2^{n-1}; 0)h(0)(1 - \pi_{n-1}) + G(Y_2^{n-1}; -2)h(-2)\pi_{n-1}\Big)$$

Comparing $h(0)$, $h(-2)$ and $h(2)$ terms in the above expressions we find:

$$\pi_n = G(Y_2^{n-1}; Y_n) = \begin{cases} 1 & Y_n = 2 \\ \pi_{n-1} & Y_n = 0 \\ 0 & Y_n = -2 \end{cases} \tag{2}$$

The same answer can be obtained by a shortcut - note that $Y_n \in \{2, 0, -2\}$. If $\{Y_n = 2\}$ then $\{X_n = 1, X_{n-1} = -1\}$; if $\{Y_n = -2\}$ then $\{X_n = -1, X_{n-1} = 1\}$. $\{Y_n = 0\}$ means that $\{X_n = X_{n-1}\}$, so

$$\mathbf{P}\big(X_n = 1|Y_2^{n-1}, Y_n = 0\big) =$$
$$\mathbf{P}\big(X_n = 1|Y_2^{n-1}, X_n = X_{n-1}\big) =$$
$$\mathbf{P}\big(X_{n-1} = 1|Y_2^{n-1}, X_n = X_{n-1}\big) =$$
$$\mathbf{P}\big(X_{n-1} = 1|Y_2^{n-1}, X_n = 1\big) =$$
$$\mathbf{P}\big(X_{n-1} = 1|Y_2^{n-1}\big) = \pi_{n-1}$$

Summarizing the above we get (2) subject to $\pi_1 = 1/2$ or which is the same (why ?)

$$\pi_n = \frac{1 - 2\pi_{n-1}}{8} Y_n^2 + \frac{1}{4} Y_n + \pi_{n-1}$$

(b) The model suitable for Kalman filter application is

$$\begin{aligned}\theta_n &= X_n \\ Y_n &= \theta_n - X_{n-1}, n \ge 2\end{aligned}$$

Now $\widehat{\theta}_{n|n-1} = \widehat{\mathbf{E}}(\theta_n|Y_2^{n-1}) = \widehat{\mathbf{E}}(X_n|Y_2^{n-1}) = 0$, since Y_2^{n-1} is a linear combination of $\{X_1, ..., X_{n-1}\}$; $\widehat{Y}_{n|n-1} = \widehat{\mathbf{E}}(Y_n|Y_2^{n-1}) = \widehat{\mathbf{E}}(\theta_n - X_{n-1}|Y_2^{n-1}) = -\widehat{\theta}_{n-1}$. So $P^{\theta}_{n|n-1} = \mathbf{E}\theta_n^2 = 1$; $P^{Y}_{n|n-1} = \mathbf{E}\big(\theta_n - (\theta_{n-1} - \widehat{\theta}_{n-1})\big)^2 = 1 + P_{n-1}$, where $P_{n-1} = \mathbf{E}(\theta_{n-1} - \widehat{\theta}_{n-1})^2$. Finally $P^{\theta Y}_{n|n-1} = \mathbf{E}\theta_n\big(\theta_n - (\theta_{n-1} - \widehat{\theta}_{n-1})\big) = 1$. Hence

$$\begin{aligned}\widehat{\theta}_n &= \frac{1}{1+P_{n-1}}\big(Y_n + \widehat{\theta}_{n-1}\big) \\ P_n &= 1 - \frac{1}{1+P_{n-1}}\end{aligned}$$

subject to $\widehat{\theta}_1 = 0$ and $P_1 = 1$.

(c) Once again the conventional approach works (see (1)):

$$\begin{aligned}&\mathbf{E}\Big(I(X_n = 1)h(Z_n)|Z_2^{n-1}\Big) = \\ &\mathbf{E}\Big(I(X_n = 1)h(X_n/X_{n-1})|Z_2^{n-1}\Big) = \\ &\mathbf{E}\Big(I(X_n = 1)h(X_{n-1})|Z_2^{n-1}\Big) = \\ &1/2\Big\{h(1)\rho_{n-1} + h(-1)(1-\rho_{n-1})\Big\}\end{aligned}$$

and

$$\begin{aligned}&\mathbf{E}\Big(G(Z_2^{n-1}; Z_n)h(Z_n)|Z_2^{n-1}\Big) = \\ &\mathbf{E}\Big(G(Z_2^{n-1}; X_n/X_{n-1})h(X_n/X_{n-1})|Z_2^{n-1}\Big) = \\ &1/2\Big(G(Z_2^{n-1}; X_{n-1})h(X_{n-1})|Z_2^{n-1}\Big) + \\ &1/2\Big(G(Z_2^{n-1}; -X_{n-1})h(-X_{n-1})|Z_2^{n-1}\Big) = \\ &1/2\Big(G(Z_2^{n-1}; 1)h(1)\rho_{n-1} + G(Z_2^{n-1}; -1)h(-1)(1-\rho_{n-1})\Big) + \\ &1/2\Big(G(Z_2^{n-1}; -1)h(-1)\rho_{n-1} + G(Z_2^{n-1}; 1)h(1)(1-\rho_{n-1})\Big) = \\ &1/2G(Z_2^{n-1}; 1)h(1) + 1/2G(Z_2^{n-1}; -1)h(-1)\end{aligned}$$

which leads to the conclusion

$$\rho_n = \begin{cases}\rho_{n-1}, & Z_n = 1 \\ 1-\rho_{n-1}, & Z_n = -1\end{cases}$$

Now since $\rho_2 = \mathbf{P}(X_2|X_2/X_1) = 1/2$, we get $\rho_n \equiv 1/2$.

The answer can be obtained intuitively - we feel that Z_2^n contains[1] no information about X_n, since it "scrambles" the signal, i.e. $\rho_n \equiv \mathbf{P}(X_n = 1) = 1/2$. To prove

[1] it can be even shown that $\{Z_2, ..., Z_n, X_n\}$ is an i.i.d. vector.

fix $\psi : \mathbb{R}^{n-1} \to \mathbb{R}$, then

$$\begin{aligned}
&\mathbf{E}\big(I(X_n = 1)\psi(Z_2, ..., Z_n)\big) = \\
&\mathbf{E}\mathbf{E}\big(I(X_n = 1)\psi(Z_2, ..., Z_n)|X_1^{n-1}\big) = \\
&\mathbf{E}\mathbf{E}\big(I(X_n = 1)\psi(Z_2, ..., Z_{n-1}, X_n/X_{n-1})|X_1^{n-1}\big) = \\
&\mathbf{E}\big(I(X_n = 1)\psi(Z_2, ..., 1/X_{n-1})|X_1^{n-1}\big) = \\
&1/2\mathbf{E}\big(\psi(Z_2, ..., X_{n-1})\big)
\end{aligned} \tag{3}$$

On the other hand

$$\begin{aligned}
&\mathbf{E}\big(\psi(Z_2, ..., Z_n)\big) = \mathbf{E}\psi(Z_2, ..., X_n/X_{n-1}) = \\
&1/2\mathbf{E}\psi(Z_2, ..., X_{n-1}) + 1/2\mathbf{E}\psi(Z_2, ..., -X_{n-1}) = \\
&1/2\mathbf{E}\psi(Z_2, ..., X_{n-1}) + 1/2\mathbf{E}\psi(X_2/X_1, ..., X_{n-1}/X_{n-2}, -X_{n-1}) = \\
&1/2\mathbf{E}\psi(Z_2, ..., X_{n-1}) + 1/2\mathbf{E}\psi(-X_2/-X_1, ..., -X_{n-1}/-X_{n-2}, -X_{n-1}) \stackrel{\dagger}{=} \\
&1/2\mathbf{E}\psi(Z_2, ..., X_{n-1}) + 1/2\mathbf{E}\psi(X_2/X_1, ..., X_{n-1}/X_{n-2}, X_{n-1}) = \\
&\mathbf{E}\psi(Z_2, ..., X_{n-1})
\end{aligned} \tag{4}$$

where the equality † is due to symmetry of the distribution of $\{X_1, X_2, ..., X_{n-1}\}$. Eq. (3) and (4) imply $\rho_n \equiv 1/2$.

(d) It immediately follows from **(c)** that $\widehat{\mathbf{E}}(X_n|Z_2^n) = 0$ and $\mathbf{E}\big((X_n - \widehat{\mathbf{E}}(X_n|Z_2^n)\big)^2 = 1$, for any $n \geq 1$.

(e), (d) Note that $X_n = X_1 + \sum_{k=2}^n Y_k$, so that

$$\mathbf{E}(X_n|Y_2^n) = \mathbf{E}(X_1|Y_2^n) + \sum_{k=2}^n Y_k$$

and hence

$$\mathbf{E}\big(X_1 - \mathbf{E}(X_1|Y_2^n)\big)^2 = \mathbf{E}\big(X_n - \mathbf{E}(X_n|Y_2^n)\big)^2$$

Both from **(a)** and **(b)** it can be seen that $\lim_{n\to\infty} \mathbf{E}\big(X_n - \mathbf{E}(X_n|Y_2^n)\big)^2 = 0$, which means that $\mathbf{E}(X_1|Y_2^n)$ converges to X_1 in $\mathbb{L}^2$ (hence also in $\mathbb{L}^1$ and with probability and in law). P-a.s. convergence follows from Borel-Cantelli Lemma, since

$$\begin{aligned}
&\mathbf{P}\big(\mathbf{E}(X_1|Y_2^n) \neq X_1\big) = \mathbf{P}\big(Y_2 = 0, Y_3 = 0, ..., Y_n = 0\big) = \\
&1/2\mathbf{P}\big(Y_2 = 0, ..., Y_n = 0|X_1 = 1\big) + 1/2\mathbf{P}\big(Y_2 = 0, ..., Y_n = 0|X_1 = -1\big) = \\
&1/2\mathbf{P}\big(X_2 = 1, ..., X_n = 1|X_1 = 1\big) + 1/2\mathbf{P}\big(X_2 = -1, ..., X_n = -1|X_1 = -1\big) = \\
&= (1/2)^{n-1}
\end{aligned}$$

so that $\sum_k \mathbf{P}\big(\mathbf{E}(X_1|Y_2^k) \neq X_1\big) < \infty$.

By **(c)** we get $\mathbf{E}(X_1|Z_2^n) \equiv 0$ (i.e. trivially converges to zero in all senses).

Problem 3

(a) Since $\widehat{\theta}_t = Y_t/t = (t\theta + W_t)/t = \theta + W_t/t$, we have

$$\mathbf{E}(\theta - \widehat{\theta}) = 0, \quad \mathbf{E}(\theta - \widehat{\theta})^2 = \frac{\mathbf{E}W_t^2}{t^2} = \frac{1}{t}$$

(b) By Itô formula

$$d\widehat{\theta}_t = -\frac{Y_t}{t^2}dt + \frac{dY_t}{t} = -\frac{1}{t}\widehat{\theta}_t dt + \frac{1}{t}dY_t = \frac{1}{t}\big(dY_t - \widehat{\theta}_t dt\big)$$

(c) We have

$$\begin{aligned}
&\mathbf{E}\frac{1}{1+\exp\{t/2-Y_t\}} = \mathbf{E}\frac{1}{1+\exp\{(1/2-\theta)t - W_t\}} \stackrel{\dagger}{=} \\
&= \frac{1}{2}\mathbf{E}\frac{1}{1+\exp\{-1/2t-W_t\}} + \frac{1}{2}\mathbf{E}\frac{1}{1+\exp\{1/2t-W_t\}} = \\
&= \frac{1}{2}\mathbf{E}\frac{1}{1+\exp\{-1/2t-W_t\}} + \frac{1}{2}\mathbf{E}\frac{\exp\{-1/2t+W_t\}}{\exp\{-1/2t+W_t\}+1} \stackrel{\ddagger}{=} \\
&= \frac{1}{2}\mathbf{E}\frac{1}{1+\exp\{-1/2t-W_t\}} + \frac{1}{2}\mathbf{E}\frac{\exp\{-1/2t-W_t\}}{\exp\{-1/2t-W_t\}+1} = \\
&\frac{1}{2}\mathbf{E}\left(\frac{1}{1+\exp\{-1/2t-W_t\}} + \frac{\exp\{-1/2t-W_t\}}{\exp\{-1/2t-W_t\}+1}\right) \equiv 1/2
\end{aligned}$$

where $\dagger$ is due to independence of θ and W_t and $\ddagger$ is due to symmetry of the distribution of W_t. So $\mathbf{E}(\pi_t - \theta) = 0$, i.e. the estimate is unbiased.

(d) Let $\xi_t = \exp\{t/2 - Y_t\}$. Then by Ito formula

$$\begin{aligned}
d\xi_t &= \exp\{t/2-Y_t\}\,(1/2dt - dY_t) + 1/2\exp\{t/2-Y_t\}dt = \\
&= \exp\{t/2-Y_t\}dt - \exp\{t/2-Y_t\}dY_t = \xi_t dt - \xi_t dY_t
\end{aligned}$$

Now since $\pi_t = 1/(1+\xi_t)$ and $\xi_t = 1/\pi_t - 1$ we have

$$\begin{aligned}
d\pi_t &= -\frac{1}{(1+\xi_t)^2}d\xi_t + \frac{1}{(1+\xi_t)^3}\xi_t^2 dt = \\
&= -\frac{1}{(1+\xi_t)^2}\xi_t\big(dt - dY_t\big) + \frac{1}{(1+\xi_t)^3}\xi_t^2 dt \\
&= -\pi_t^2(1/\pi_t - 1)\big(dt - dY_t\big) + \pi_t^3(1/\pi_t-1)^2 dt = \\
&= -\pi_t(1-\pi_t)\big(dt - dY_t\big) + \pi_t(1-\pi_t)^2 dt = \\
&= \pi_t(1-\pi_t)\big(-dt + dY_t + (1-\pi_t)dt\big) = \\
&= \pi_t(1-\pi_t)\big(dY_t - \pi_t dt\big)
\end{aligned}$$

Appendix: what so special about π_t anyway ?

It can be shown that $\pi_t = \mathbf{E}(\theta|Y_0^t)$, and moreover it is the particular case of the Wonham filter for continuous time processes. This is of course beyond the scope of the course. But let's see that π_t is a at least better estimate than $\widehat{\theta}_t$.

$$\begin{aligned} Q_t &= \mathbf{E}\big(\theta-\pi_t\big)^2 = \mathbf{E}\Big(\theta - \frac{1}{1+\exp\{(1/2-\theta)t-W_t\}}\Big)^2 = \\ &= \frac{1}{2}\mathbf{E}\Big(\frac{1}{1+\exp\{1/2t-W_t\}}\Big)^2 + \frac{1}{2}\mathbf{E}\Big(1-\frac{1}{1+\exp\{-1/2t-W_t\}}\Big)^2 = \\ &= \frac{1}{2}\mathbf{E}\Big(\frac{1}{1+\exp\{1/2t-W_t\}}\Big)^2 + \frac{1}{2}\mathbf{E}\Big(\frac{1}{\exp\{1/2t+W_t\}+1}\Big)^2 = \\ &= \frac{1}{2}\mathbf{E}\Big(\frac{1}{1+\exp\{1/2t-W_t\}}\Big)^2 + \frac{1}{2}\mathbf{E}\Big(\frac{1}{\exp\{1/2t-W_t\}+1}\Big)^2 = \\ &= \mathbf{E}\Big(\frac{1}{1+\exp\{1/2t+W_t\}}\Big)^2 \end{aligned}$$

Let $\eta(t)$ be a Gaussian random variable with $\mathbf{E}\eta(t) = t/2$ and variance $\mathbf{E}(\eta(t) - t/2)^2 = t$, then

$$Q_t = \mathbf{E}\frac{1}{(1+\eta(t))^2} = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} \frac{1}{(1+e^x)^2} e^{-(x-t/2)^2/2t}dx \tag{5}$$

Note that the function $(1+e^x)^{-2}$ is less than 1 for any x and less than $e^{-2x} \le e^{-x}$ for $x \ge 0$. So the integral in (5) can be bounded as

$$Q_t \le \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{0} e^{-(x-t/2)^2/2t}dx + \frac{1}{\sqrt{2\pi t}}\int_{0}^{\infty} e^{-x}e^{-(x-t/2)^2/2t}dx := I_1 + I_2$$

Integrating I_1 w.r.t $y = (x-t/2)/\sqrt{t}$ we get, $t \ge 0$

$$I_1 = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{-\sqrt{t}/2} e^{-y^2/2}\sqrt{t}dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\sqrt{t}/2} e^{-y^2/2}dy = \mathbf{P}\big(\zeta \le -\sqrt{t}/2\big)$$

where ζ is a standard Gaussian r.v.

Similarly

$$\begin{aligned} I_2 &= \frac{1}{\sqrt{2\pi t}}\int_{0}^{\infty} e^{-x}e^{-(x-t/2)^2/2t}dx = \frac{1}{\sqrt{2\pi t}}\int_{0}^{\infty} e^{-(x+t/2)^2/2t}dx = \\ &= \frac{1}{\sqrt{2\pi}}\int_{\sqrt{t}/2}^{\infty} e^{-y^2/2}dx = \mathbf{P}(\zeta \ge \sqrt{t}/2) \end{aligned}$$

That is

$$Q_t \le 2P(\zeta \ge \sqrt{t}/2)$$

so using the well known bound

$$\mathbf{P}(\zeta \ge x) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} e^{-s^2/2}ds \le \frac{1}{\sqrt{2\pi}}\frac{e^{-2x^2}}{x+\sqrt{x^2+2/\pi}}$$

we get

$$Q_t \le \frac{1}{\sqrt{2\pi}}\frac{e^{-t^2}}{\sqrt{t}/2+\sqrt{t/2+2/\pi}} \tag{6}$$

which is much better than the rate in the linear case.