RANDOM PROCESSES. THE SOLUTION TO FINAL TEST.

April, 2000

Problem 1.

(a) The optimal linear estimate $\widetilde{\xi}_0 = \sum_{k\neq 0} a_k \xi_k$ satisfies the orthogonality principle:

$$\mathbf{E}\big(\xi_0 - \sum_{k\neq 0} a_k \xi_k\big)\xi_\ell = 0, \quad \ell \neq 0$$

Note that if we set $a_0 \equiv 0$ and choose some constant γ (which does not necessarily equals 0), the orthogonality eq. becomes:

$$\mathbf{E}\big(\xi_0 - \sum_{k=-\infty}^{\infty} a_k \xi_k\big)\xi_\ell = \gamma\delta_\ell, \quad \text{for all } \ell$$

Let $R(m) = \mathbf{E}\xi_n \xi_{n+m}$, then

$$R(\ell) - \sum_{k=-\infty}^{\infty} a_k R(k+\ell) = \gamma\delta_\ell, \quad \text{for all } \ell$$

Now calculate the Fourier transform of both sides:

$$f(\lambda) - A^*(\lambda) f(\lambda) = \gamma$$

Clearly $A(\lambda)$ is real and since $f(\lambda) > 0$:

$$A(\lambda) = 1 - \frac{\gamma}{f(\lambda)}$$

The constant γ is determined by the constrain $a_0 \equiv 0$:

$$a_0 = \frac{1}{2\pi}\int_{[-\pi,\pi]} A(\lambda)d\lambda = 1 - \gamma\frac{1}{2\pi}\int_{[-\pi,\pi]} 1/f(\lambda)d\lambda \equiv 0$$

which implies:

$$\gamma = \frac{2\pi}{\int_{[-\pi,\pi]} d\lambda/f(\lambda)}$$

Now the filter is completely specified.

(b)

$$\begin{aligned}\widetilde{P} &= \mathbf{E}(\xi_0 - \sum_{k=-\infty}^{\infty} a_k\xi_k)^2 = R(0) - 2\sum_k a_k R(k) + \sum_k\sum_m a_k a_m R(k-m) \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} \big(f(\lambda) - 2A(\lambda)f(\lambda) + A(\lambda)^2 f(\lambda)\big)\,d\lambda = \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} f(\lambda)\,(1-A(\lambda))^2\,d\lambda = \frac{1}{2\pi}\int_{[-\pi,\pi]} \frac{\gamma^2}{f(\lambda)}d\lambda = \gamma\end{aligned}$$

(c) For white noise (i.e. $f(\lambda) \equiv \sigma^2$), we expect that $\widetilde{\xi_0} \equiv 0$. Indeed, in this case $A(\lambda) \equiv 0$.

(d) Of course, the solution can be obtained as a special case of (a).

Alternatively, if one notes that $\{\xi_k, k \neq 0\}$ and $\{\xi_1, \xi_{-1}, \varepsilon_k, k \neq 0, 1\}$ are related by a one-to-one linear transformation, the solution can be simplified, since then [1] with prob. one

$$\mathbf{E}(\xi_0|\xi_k, k\neq 0) = \mathbf{E}(\xi_0|\xi_1, \xi_{-1}, \varepsilon_k, k \neq 0, 1\} = \mathbf{E}(\xi_0|\xi_1, \xi_{-1})$$

where the last equality follows from independence of $\{\varepsilon_k, k \neq 0, 1\}$ and $\{\xi_{-1}, \xi_0, \xi_1\}$. Now the problem is reduced to estimating a component of a Gaussian vector:

$$\begin{aligned}\xi_1 &= a\xi_0 + b\varepsilon_1 \\ \xi_{-1} &= \xi_0/a - b/a\varepsilon_0\end{aligned}$$

Since the process is stationary

$$\mathbf{E}\xi_n = 0, \quad \mathbf{E}\xi_n^2 = \frac{b^2}{1-a^2}$$

and

$$\begin{aligned}\mathbf{E}\xi_0\xi_1 &= \mathbf{E}\xi_0(a\xi_0 + b\varepsilon_1) = ab^2/(1-a^2) \\ \mathbf{E}\xi_0\xi_{-1} &= \mathbf{E}\xi_{-1}(a\xi_{-1} + b\varepsilon_0 = ab^2/(1-a^2) \\ \mathbf{E}\xi_{-1}\xi_1 &= \mathbf{E}(a\xi_0 + b\varepsilon_1)(\xi_0/a - b/a\varepsilon_0) = \mathbf{E}\xi_0^2 - b\mathbf{E}\xi_0\varepsilon_0 = \\ &= b^2/(1-a^2) - b^2 = b^2a^2/(1-a^2)\end{aligned}$$

So that:

$$\begin{aligned}\mathbf{E}(\xi_0|\xi_1,\xi_{-1}) &= \begin{pmatrix}\mathbf{E}\xi_0\xi_1 & \mathbf{E}\xi_0\xi_{-1}\end{pmatrix}\begin{pmatrix}\mathbf{E}\xi_1\xi_1 & \mathbf{E}\xi_0\xi_1 \\ \mathbf{E}\xi_0\xi_1 & \mathbf{E}\xi_1\xi_1\end{pmatrix}^{-1}\begin{pmatrix}\xi_1 \\ \xi_{-1}\end{pmatrix} = \\ &= \frac{ab^2}{1-a^2}\begin{pmatrix}1 & 1\end{pmatrix}\left(\frac{b^2}{1-a^2}\right)^{-1}\begin{pmatrix}1 & a^2 \\ a^2 & 1\end{pmatrix}^{-1}\begin{pmatrix}\xi_1 \\ \xi_{-1}\end{pmatrix} = \\ &= \frac{a}{1-a^4}\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}1 & -a^2 \\ -a^2 & 1\end{pmatrix}^{-1}\begin{pmatrix}\xi_1 \\ \xi_{-1}\end{pmatrix} = \\ &= \frac{a}{1+a^2}[\xi_1 + \xi_{-1}]\end{aligned}$$

[1] since ξ_n is Gaussian, the orthogonal projection is replaced by conditional expectation

To calculate the corresponding error note that:

$$\frac{a}{1+a^2}[\xi_1+\xi_{-1}] = \frac{a}{1+a^2}\big((a+1/a)\xi_0 + b\varepsilon_1 - b/a\varepsilon_0\big) = \\ = \xi_0 + \big(b\varepsilon_1 - b/a\varepsilon_0\big)$$

from which it follows that:

$$P = \mathbf{E}(\xi - \widehat{\xi}_0)^2 = \frac{b^2}{1+a^2}$$

(e) Note that vectors $\{\xi_1, \xi_2, ..., \xi_n\}$ and $\{\xi_1, \varepsilon_2, ..., \varepsilon_2\}$ are related by one-to-one *linear* transformation. Then with probability one

$$\widehat{\mathbf{E}}(\xi_0|\xi_1^n) = \widehat{\mathbf{E}}(\xi_0|\xi_1, \varepsilon_2^n) = \widehat{\mathbf{E}}(\xi_0|\xi_1)$$

where the last inequality follows from independence of ξ_1 and ε_k, $k > 1$.

For $n \ge 1$:

$$\widehat{\xi}_0(n) = \frac{\mathbf{E}\xi_0\xi_1}{\mathbf{E}\xi_1^2}\xi_1 = \frac{ab^2/(1-a^2)}{a^2b^2/(1-a^2)+b^2}\xi_1 = a\xi_1$$

and the error is:

$$P = \mathbf{E}(\xi_0 - a\xi_1)^2 = \mathbf{E}\big(\xi_0(1-a^2) - ab\varepsilon_1\big)^2 = b^2$$

(f) Identical to (e)

Problem 2

(a) Introduce:

$$X_n = \begin{bmatrix} I(\theta_n = a_1) \\ I(\theta_n = a_2) \\ \vdots \\ I(\theta_n = a_d) \end{bmatrix}, \quad J = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}$$

Clearly $\theta_n = J^\top X_n$ and $Y_n = J^\top X_n + \gamma J^\top X_{n-1} + \xi_n$. Let (X'_n, Y'_n) be generated by a recursion:

$$\begin{aligned} X'_n &= \Lambda^\top X'_{n-1} + \varepsilon_n \\ Y'_n &= J^\top X'_n + \gamma J^\top X'_{n-1} + \xi_n \end{aligned}$$

where ε_n is a sequence of independent Gaussian vector r.v. in $\mathbb{R}^d$, such that:

$$\mathbf{E}\varepsilon_n = 0, \quad \mathbf{E}\varepsilon_n\varepsilon_n^\top = \mathbf{diag}\,(p_n) - \Lambda^\top \mathbf{diag}\,(p_{n-1})\,\Lambda := D_n$$

and

$$p_n = \Lambda^\top p_{n-1}, \quad \text{subject to } p_0$$

Note that X_n and X'_n have the same correlation structure (see lecture note No. 9), so that the optimal *linear* estimate of X_n from Y_1^n is obtained by applying the Kalman filter for the pair (X'_n, Y'_n) to the observations Y_n:

$$\begin{aligned}\widehat{X}_n &= \Lambda^\top \widehat{X}_{n-1} + P^{xy}_{n-1}[P^y_{n-1}]^{-1}(Y_n - J^\top \Lambda^\top \widehat{X}_{n-1} - \gamma \Lambda^\top \widehat{X}_{n-1}) \\ P_n &= P^x_{n-1} - P^{xy}_{n-1}[P^y_{n-1}]^{-1}[P^{xy}_{n-1}]^\top\end{aligned}$$

where

$$\begin{aligned}P^x_{n-1} &= \Lambda^\top P_{n-1}\Lambda + D_n \\ P^{xy}_{n-1} &= \Lambda^\top P_{n-1}(\Lambda J + \gamma J) + D_n J \\ P^y_{n-1} &= (J^\top \Lambda^\top + \gamma J^\top) P_{n-1}(\Lambda J + \gamma J) + J^\top D_n J + \mathbf{E}\xi_n^2\end{aligned}$$

(b) Let:

$$\pi_n(i) = \mathbf{P}\{\theta_n = a_i | Y_1^n\} = \mathbf{E}[I(\theta_n = a_i)|Y_1^n] := G(Y_n, Y_1^{n-1})$$

and

$$\pi_{n|n-1}(i) = \mathbf{P}\{\theta_n = a_i | Y_1^{n-1}\} = \mathbf{E}[I(\theta_n = a_i)|Y_1^{n-1}]$$

Then for any bounded $h(x)$ and $H(x_1, ..., x_{n-1})$

$$\mathbf{E}h(Y_n)H(Y_1, ..., Y_{n-1})\left[I(\theta_n = a_i) - G(Y_n, Y_1^{n-1})\right] = 0$$

or equivalently:

$$\mathbf{E}\big(h(Y_n)\left[I(\theta_n = a_i) - G(Y_n, Y_1^{n-1})\right]|Y_1^{n-1}\big) = 0$$

Calculate each term separately:

$$\begin{aligned}&\mathbf{E}\big[I(\theta_n = a_i)h(Y_n)|Y_1^{n-1}\big] = \mathbf{E}\left\{\mathbf{E}\big[I(\theta_n = a_i)h(Y_n)|Y_1^{n-1}, \theta_{n-1}\big]\Big|Y_1^{n-1}\right\} \\ &= \mathbf{E}\left\{\mathbf{E}\big[I(\theta_n = a_i)h(a_i + \gamma\theta_{n-1} + \xi_n)|Y_1^{n-1}, \theta_{n-1}\big]\Big|Y_1^{n-1}\right\} = \\ &= \sum_j \pi_{n-1}(j)\int_{\mathbb{R}} \lambda_{ji} h(a_i + \gamma a_j + x) f(x)dx = \\ &= \int_{\mathbb{R}} \sum_j \pi_{n-1}(j)\lambda_{ji} h(x) f(x - a_i - \gamma a_j)dx \qquad (1)\end{aligned}$$

and similarly:

$$\begin{aligned}&\mathbf{E}\big(h(Y_n)G(Y_n, Y_1^{n-1})|Y_1^{n-1}\big) = \mathbf{E}\mathbf{E}\left\{\big(h(\theta_n + \gamma\theta_{n-1} + \xi_n)\cdot\right. \\ &\left.\cdot G(\theta_n + \gamma\theta_{n-1} + \xi_n, Y_1^{n-1})|\theta_{n-1}, Y_1^{n-1}\big)\right\} = \qquad (2) \\ &\sum_j \pi_{n-1}(j)\sum_i \int_{\mathbb{R}} \lambda_{ji} h(a_i + \gamma a_j + x) G(a_i + \gamma a_j + x, Y_1^{n-1}) f(x)dx = \\ &= \int_{\mathbb{R}} \sum_j \pi_{n-1}(j)\sum_i \lambda_{ji} h(x) G(x, Y_1^{n-1}) f(x - a_i - \gamma a_j)dx\end{aligned}$$

Since (1) and (2) should be equal for any $h(x)$, we deduce:

$$\sum_j \pi_{n-1}(j)\lambda_{ji} f(x - a_i - \gamma a_j) =$$

$$= \sum_j \pi_{n-1}(j) \sum_i \lambda_{ji} G(x, Y_1^{n-1}) f(x - a_i - \gamma a_j)$$

or:

$$G(x, Y_1^{n-1}) = \frac{\sum_j \pi_{n-1}(j)\lambda_{ji} f(x - a_i - \gamma a_j)}{\sum_i \sum_j \pi_{n-1}(j)\lambda_{ji} f(x - a_i - \gamma a_j)} \tag{3}$$

and the recursion is obtained by $\pi_n(j) = G(Y_n, Y_1^{n-1})$.

(c) If $\gamma = 0$, a conventional Wonham filter is obtained.

(d) Note that Y_n is a Gaussian r.v. given θ_n and θ_{n-1} with mean:

$$\mathbf{E}(Y_n|\theta_n, \theta_{n-1}) = \theta_n + \gamma\theta_{n-1}$$

and variance:

$$\mathbf{E}\Big(\big[Y_n - \mathbf{E}(Y_n|\theta_n, \theta_{n-1})\big]^2 \theta_n, \theta_{n-1}\Big) = \theta_{n-1}^2\sigma_\gamma^2 + \sigma_\xi^2 := \sigma^2(\theta_{n-1})$$

So (1) reads:

$$\mathbf{E}\big[I(\theta_n = a_i)h(Y_n)|Y_1^{n-1}\big] = \cdots =$$

$$= \sum_j \pi_{n-1}(j) \int_{\mathbb{R}} \lambda_{ji} h(x)\varphi\left(x, a_i + \gamma a_j, \sigma(a_j)\right) dx$$

where

$$\varphi(x, a, b) = \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{(x-a)^2}{2b^2}\right\}$$

Similarly modifying (2), we conclude that the optimal filter is given by (3), with $f(Y_n - a_i - \gamma a_j)$ replaced by $\varphi\big(Y_n, a_i + \gamma a_j, \sqrt{a_j^2\sigma_\gamma^2 + \sigma_\xi^2}\big)$.

Problem 3

Let for brevity[2] $g(x) = |x|/(|x| + 1)$.

[2] By the way, $d(X, Y) = \mathbf{E}g(X - Y)$ is indeed a metric. All the properties are obvious, except maybe for the triangle inequality. This is proved as follows: we should verify that for any z:

$$\frac{|x-y|}{|x-y|+1} \le \frac{|x-z|}{|x-z|+1} + \frac{|z-y|}{|z-y|+1}$$

To prove this, not that for fixed x and y the right hand side expression obeys a global minimum, which equals to the left hand side and attained at $z = x$ and $z = y$. E.g. let $z > y > x$, then:

$$\frac{|x-z|}{|x-z|+1} + \frac{|z-y|}{|z-y|+1} = \frac{z-x}{z-x+1} + \frac{z-y}{z-y+1} \ge \frac{z-x}{z-x+1} \ge \frac{y-x}{y-x+1}$$

etc.

(a) For any $\varepsilon > 0$

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} = \mathbf{P}\left\{g(\xi_n - \xi) > g(\varepsilon)\right\} \le \frac{\mathbf{E}g(\xi_n - \xi)}{g(\varepsilon)} \to 0, \quad n \to \infty$$

where the equality holds since $g(x)$ is one to one and Chebyshev inequality holds (non trivially) since $g(x)$ is bounded ($\mathbf{E}g(\xi_n - \xi) < \infty$).

(b) By the way, note that since $g(x)$ is a continuous function (see exam 1999)

$$\xi_n \xrightarrow{\mathbf{P}} \xi \implies \xi_n - \xi \xrightarrow{\mathbf{P}} 0 \implies g(\xi_n - \xi) \xrightarrow{\mathbf{P}} g(0) = 0$$

So that the sequence $\zeta_n := g(\xi_n - \xi)$ converges to 0 in probability. Since $0 \le \zeta_n < 1$, we conclude (why?) that $\mathbf{E}\zeta_n \to 0$, which completes the proof.

A straight forward approach is also possible: note that $g(x) < 1$, so for any $\varepsilon > 0$

$$\begin{aligned} &d(\xi_n, \xi) = \mathbf{E}g(\xi_n - \xi) = \\ &= \mathbf{E}g(\xi_n - \xi)I(|\xi_n - \xi| > \varepsilon) + \mathbf{E}g(\xi_n - \xi)I(|\xi_n - \xi| \le \varepsilon) \le \\ &\le 1 \cdot \mathbf{P}\{|\xi_n - \xi| > \varepsilon\} + g(\varepsilon) \to g(\varepsilon), \quad n \to \infty \end{aligned}$$

Since $g(\varepsilon)$ is a strictly decreasing function of ε and ε can be chosen arbitrary small we conclude:

$$d(\xi_n, \xi) \to 0, \quad n \to \infty$$

The proof of (a) and (b) can be also easily deduced from

Lemma 1.1. *For any fixed $\varepsilon > 0$:*

$$\mathbf{E}\frac{|X|}{1+|X|} - \frac{\varepsilon}{1+\varepsilon} \le \mathbf{P}\big(|X| \ge \varepsilon\big) \le \frac{1+\varepsilon}{\varepsilon}\mathbf{E}\frac{|X|}{1+|X|} \tag{4}$$

Proof.

$$\begin{aligned} &\mathbf{E}\frac{|X|}{1+|X|} = \mathbf{E}\frac{|X|}{1+|X|}I(|X| \ge \varepsilon) + \mathbf{E}\frac{|X|}{1+|X|}I(|X| < \varepsilon) \ge \\ &\ge \mathbf{E}\frac{\varepsilon}{1+\varepsilon}I(|X| \ge \varepsilon) = \frac{\varepsilon}{1+\varepsilon}\mathbf{P}\big(|X| \ge \varepsilon\big) \end{aligned}$$

which implies the upper bound. The lower bound is derived similarly

$$\begin{aligned} &\mathbf{E}\frac{|X|}{1+|X|} = \mathbf{E}\frac{|X|}{1+|X|}I(|X| \ge \varepsilon) + \mathbf{E}\frac{|X|}{1+|X|}I(|X| < \varepsilon) \le \\ &\le \mathbf{E}I(|X| \ge \varepsilon) + \frac{\varepsilon}{1+\varepsilon} \end{aligned}$$

□

(c)

(I) For example $d'(\xi_n, \xi) = \mathbf{E}|\xi_n - \xi|$, i.e. convergence in prob. does not imply convergence in the mean (take e.g. $\xi_n = \xi/n$ with ξ a r.v. with $\mathbf{E}\xi = \infty$)

(II) For another example, set $d''(\xi_n,\xi) = \mathbf{E}I(\xi_n \neq \xi) = \mathbf{P}\{\xi_n \neq \xi\}$. It is indeed a metric (with prob. 1): for any two r.v. η and ξ

(i) $\xi \equiv \eta \implies d''(\eta,\xi) = 0$ and

$$d''(\xi,\eta) = 0 \implies \mathbf{P}\{\xi \neq \eta\} = 0 \implies \xi = \eta \text{ with prob. } 1$$

(ii) $d''(\xi,\eta) > 0$

(iii) For any numbers a, b, c

$$I(a \neq b) \leq I(a \neq c) + I(b \neq c)$$

(which is verified by trying all the combinations $a = b \neq c$, $a \neq b \neq c$, etc.) Using this inequality with r.v. and taking expectation from both sides leads to the triangle inequality.

Now take some ξ_n, so that $\xi_n \xrightarrow{\mathbf{P}} 0$ and $\mathbf{P}\{\xi_n \neq 0\} = 1$, clearly $d''(\xi_n,\xi) \not\to 0$.

(d) The idea is to define a metric, convergence in which will be equivalent to convergence in distribution. Once such metric is chosen, one can pick a sequence which converges in distribution and does not converge in probability. Construction of such metric is possible[3], but non trivial.

(e) Since $\xi_n \xrightarrow{d} C$, by definition for any bounded and continuous function $f(x)$:

$$\mathbf{E}f(\xi_n) \to \mathbf{E}f(C)$$

Take special function $f'(x) = |x - C|/(|x - C| + 1)$, then:

$$\mathbf{E}f'(\xi_n) \to \mathbf{E}f'(C) \equiv 0$$

which is nothing but

$$d(\xi_n, C) \to 0 \implies \xi_n \xrightarrow{\mathbf{P}} C$$

[3]refer 'Probability', Second edition, A.N. Shiryaev - look for weak convergence and Prokhorov-Levy metric