STOCHASTIC PROCESSES. SOLUTIONS TO HOME ASSIGNMENTS

2. Stationary Random Processes

Problem 2.1

It is well known, that characteristic function of a random vector defines its distribution. Introduce the vector

$$\xi_n = \begin{bmatrix} \xi(t_1) \\ \xi(t_2) \\ \cdots \\ \xi(t_n) \end{bmatrix}$$

In this case

$$\Phi(\lambda) \stackrel{\triangle}{=} \mathbb{E}e^{i\lambda^T \xi_n} = \mathbb{E}\mathbb{E}(e^{i\lambda^T \xi_n}|\alpha, \beta)$$

since γ is independent of α, β we get

$$\Phi(\lambda) = \mathbb{E}\frac{1}{2\pi}\int_0^{2\pi} \exp\left(\sum_{k=1}^n i\lambda_k \alpha \sin(\beta t_k + \gamma)\right) d\gamma$$

Denote by $\Phi_h(\lambda)$ the characteristic function of the time shifted vector, namely

$$\begin{aligned}
\Phi_h(\lambda) &= \mathbb{E}\mathbb{E}\left[\exp\left\{i\sum_{k=1}^n \lambda_k \xi(t_k + h)\right\}\Big|\alpha, \beta\right] = \\
&= \mathbb{E}\frac{1}{2\pi}\int_0^{2\pi} \exp\left\{i\sum_{k=1}^n \lambda_k \alpha \sin(\beta t_k + \beta h + \gamma)\right\} d\gamma = \\
&= \mathbb{E}\frac{1}{2\pi}\int_{\beta h}^{2\pi+\beta h} \exp\left\{i\sum_{k=1}^n \lambda_k \alpha \sin(\beta t_k + \gamma')\right\} d\gamma' = \\
&= \mathbb{E}\frac{1}{2\pi}\int_{\beta h}^{2\pi} \exp\left\{i\sum_{k=1}^n \lambda_k \alpha \sin(\beta t_k + \gamma')\right\} d\gamma' + \\
&\quad +\mathbb{E}\frac{1}{2\pi}\int_{2\pi}^{2\pi+\beta h} \exp\left\{i\sum_{k=1}^n \lambda_k \alpha \sin(\beta t_k + \gamma')\right\} d\gamma' = \\
&= \mathbb{E}\frac{1}{2\pi}\int_0^{2\pi} \exp\left\{i\sum_{k=1}^n \lambda_k \alpha \sin(\beta t_k + \gamma'')\right\} d\gamma'' \equiv \Phi(\lambda)
\end{aligned}$$

Date: Summer, 2004.

Problem 2.2

(a) $R(k)$ is non negative definite if

$$\sum_{k,m} a_k R(k-m)\bar{a}_m \ge 0 \tag{2.1}$$

for any sequence $\{a_k\}$. Let $S(\lambda)$ be spectral density corresponding to $R(k)$, then

$$R(k-m) = \frac{1}{2\pi}\int_{[-\pi,\pi]} S(\lambda)e^{j(k-m)\lambda}d\lambda$$

and

$$\begin{aligned}\sum_{k,m} a_k R(k-m)\bar{a}_m &= \sum_{k,m} a_k \frac{1}{2\pi}\int_{[-\pi,\pi]} S(\lambda)e^{j(k-m)\lambda}d\lambda \bar{a}_m = \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} S(\lambda)\sum_{k,m} a_k e^{jk\lambda}\bar{a}_m e^{-jm\lambda}d\lambda = \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} S(\lambda)\Big|A(\lambda)\Big|^2 d\lambda \end{aligned} \tag{2.2}$$

where $A(\lambda)$ is the Fourier transform of $\bar{a}_k$. Due to (2.1), (2.2) and arbitrariness of a_k, $S(\lambda)\ge 0$ follows. Starting from $S(\lambda)\ge 0$, by (2.2), we deduce (2.1), which proves the other direction.

(b) Assume that $R(n)$ can be decomposed

$$R(n) = \sum_{k=-\infty}^{\infty} h(k)\bar{h}(k-n)$$

Then

$$\begin{aligned} S(\lambda) &= \sum_m R(m)e^{-j\lambda m} = \sum_m\sum_k h(k)\bar{h}(k-m)e^{-j\lambda m} = \\ &= \sum_k h(k)\sum_\ell \bar{h}(\ell)e^{-j\lambda(k-\ell)} = \Big|H(\lambda)\Big|^2 \ge 0 \end{aligned}$$

for any λ. So, by virtue of (a), $R(n)$ is a non negative definite sequence.

(c) Let X'_n and X''_n be a pair of independent processes with zero mean and correlation functions $R'(k,m)$ and $R''(k,m)$. Introduce $Y_n = X'_n X''_n$ and $Z_n = X'_n + X''_n$. Then

$$\mathbb{E}Y_kY_m = \mathbb{E}X'_kX''_kX'_mX''_m = \mathbb{E}X'_kX'_m\mathbb{E}X''_kX''_m = R'(k,m)R''(k,m)$$

and

$$\mathbb{E}Z_kZ_m = \mathbb{E}(X'_k + X''_k)(X'_m + X''_m) = R'(k,m) + R''(k,m)$$

Problem 2.3

Any symmetric sequence $R(n)$, which satisfies[1]

(i) $R(0)\ge R(m)$, $m\ne 0$
(ii) $R(n)$ is positive definite

[1] Note that these conditions are not necessarily independent

can be an autocorrelation function of some process.

(a) For $R(n) = e^{-n^2}$ (i) is obvious. Verify (ii) using the results of the previous problem

$$\begin{aligned} S(\lambda) &= \sum_n R(n)e^{-jn\lambda} = \sum_n e^{-n^2-jn\lambda} = 1 + 2\sum_{n=1}^{\infty} e^{-n^2}\cos(n\lambda) \geq \\ &\geq 1 - 2\sum_{n=1}^{\infty} e^{-n^2} \geq 1 - 2e^{-1} - 2e^{-4} - 2\sum_{n=3}^{\infty} e^{-n} = \\ &= 1 - 2e^{-1} - 2e^{-4} - 2e^{-3}/(1-e^{-1}) > 0, \quad \forall\lambda \end{aligned}$$

so that (ii) holds as well.

(b) No: $S(\lambda) = 1 + 1.4\cos(\lambda)$ is negative for λ on some interval (e.g. around $\lambda = \pi$)

(c) Note that $R(n) = h(n) \star h(-n)$ where $h(n) = I(0 \leq n < N)$, so by virtue of (b) from the previous problem, $R(n)$ is non negative definite.

Problem 2.4

(a) Note that

$$\lambda_k = \frac{v_k^* R_x v_k}{v_k^* v_k} \tag{2.3}$$

where v_k is the eigenvector corresponding to λ_k. Denote by $v_{k,\ell}$ the ℓ-th component of the k-the eigenvector. Then

$$v_k^* R_x v_k = \sum_{\ell=1}^{N}\sum_{m=1}^{N} v_{k,\ell} R_x(\ell, m) v_{k,m} = \sum_{\ell=1}^{N}\sum_{m=1}^{N} v_{k,\ell} r_x(\ell - m) v_{k,m}$$

where $r_x(\ell - m) = \mathbb{E}X(\ell)X(m)$ is the autocorrelation sequence of the process. Using the representation

$$r_x(\ell - m) = \frac{1}{2\pi}\int_{[-\pi,\pi]} S_x(\lambda) e^{j\lambda(\ell-m)} d\lambda$$

obtain

$$\begin{aligned} v_k^* R_x v_k &= \frac{1}{2\pi}\int_{[-\pi,\pi]} S_x(\lambda)\left\{\sum_\ell v_{k,\ell} e^{j\lambda\ell} \sum_m v_{k,m} e^{-j\lambda m}\right\} d\lambda = \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} S_x(\lambda)|V_k(\lambda)|^2 d\lambda \end{aligned}$$

Similarly

$$v_k^* v_k = \frac{1}{2\pi}\int_{[-\pi,\pi]} |V_k(\lambda)|^2 d\lambda$$

so that by (2.3)

$$\lambda_k = \frac{\int_{[-\pi,\pi]} S_x(\lambda)|V_k(\lambda)|^2 d\lambda}{\int_{[-\pi,\pi]} |V_k(\lambda)|^2 d\lambda}$$

which in turn implies

$$\min_\lambda S_x(\lambda) \leq \lambda_k \leq \max_\lambda S_x(\lambda)$$

for all k.

(b) Introduce

$$\gamma_n = \frac{\mathbb{E}\big(\sum_{k=0}^{N-1} X_{n-k}a_k\big)^2}{\mathbb{E}\big(\sum_{k=0}^{N-1} \xi_{n-k}a_k\big)^2}$$

Define vectors $X^n = [X_n, ..., X_{n-N+1}]^*$, $\xi^n = [\xi_n, ..., \xi_{n-N+1}]^*$ and $a = [a_0, ..., a_{N-1}]^*$ so that

$$\gamma_n \equiv \gamma = \frac{\mathbb{E}(X^{n*}a)^2}{\mathbb{E}(\xi^{n*}a)^2} = \frac{a^* R_x a}{\sigma^2 a^* a} = \sigma^{-2}\frac{a^* U\Lambda U^* a}{a^* U U^* a}$$

where U is an orthogonal matrix with v_k as columns and Λ is a diagonal matrix with $\Lambda_{jj} = \lambda_j$. Set $\widetilde{a} = U^* a$, then

$$\gamma = \sigma^{-2}\frac{\widetilde{a}^*\Lambda\widetilde{a}}{\widetilde{a}^*\widetilde{a}} = \sigma^{-2}\frac{\sum_{j=0}^{N-1}\widetilde{a}_j^2\lambda_j}{\sum_{j=0}^{N-1}\widetilde{a}_j^2} \leq \lambda_{\max}/\sigma^2$$

where the equality holds when $a = v_{\max}$, the eigenvector corresponding to $\lambda_{\max} = \max_j \lambda_j$.

3. Linear estimation of stationary sequences

Problem 3.1

(a)

$$\widehat{X}_n = \sum_{k=-\infty}^{\infty} Y_k \tilde{a}_{n-k} = \sum_{k=-\infty}^{\infty} Y_{n-k}\tilde{a}_k$$

By orthogonality principle

$$\mathbb{E}(X_n - \widehat{X}_n)Y_{n-\ell} = 0, \quad \ell = ..., -1, 0, 1, ...$$

which implies:

$$R_{xy}(\ell) - \sum_k R_y(\ell - k)\tilde{a}_k = 0, \quad \ell = ..., -1, 0, 1, ...$$

This version of Wiener-Hopf equation can be solved in the domain of Fourier transform:

$$\begin{aligned} S_{xy}(\lambda) \quad &:= \quad \sum_\ell R_{xy}(\ell)e^{-j\lambda\ell} = \sum_k\sum_\ell R_y(\ell-k)\tilde{a}_k e^{-j\lambda\ell} = \\ &= \quad \sum_k \tilde{a}_k e^{-j\lambda k}\sum_\ell R_y(\ell)e^{-j\lambda\ell} = \tilde{A}(\lambda)S_y(\lambda) \end{aligned}$$

Assuming that $S_y(\lambda) > 0$, we obtain the expression for the filter in terms of spectral densities

$$\tilde{A}(\lambda) = \frac{S_{xy}(\lambda)}{S_y(\lambda)}$$

The mean square error is:

$$\begin{aligned} &\mathbb{E}(X_n - \widehat{X}_n)^2 = \mathbb{E}X_n^2 - \mathbb{E}X_n\widehat{X}_n = \\ &= R_x(0) - \sum_k \mathbb{E}X_nY_{n-k}\tilde{a}_k = R_x(0) - \sum_k R_{xy}(k)\tilde{a}_k = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(\lambda)d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{xy}(\lambda)\sum_k \tilde{a}_k e^{j\lambda k}d\lambda = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(S_x(\lambda) - S_{xy}(\lambda)\bar{\tilde{A}}(\lambda)\Big)d\lambda = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(S_x(\lambda) - \frac{|S_{xy}(\lambda)|^2}{S_y(\lambda)}\right)d\lambda \end{aligned}$$

(b) By orthogonality property:

$$\mathbb{E}(X_n - \sum_{k=0}^{\infty} Y_{n-k}\tilde{a}_k)Y_\ell = 0, \quad \ell \le n$$

and

$$R_{xy}(n-\ell) - \sum_{k=0}^{\infty} R_y(n-\ell-k)\tilde{a}_k = 0, \quad \ell \le n$$

or

$$R_{xy}(m) - \sum_{k=0}^{\infty} R_y(m-k)\tilde{a}_k = 0, \quad m \ge 0 \tag{3.1}$$

Z-transform of the left hand side of (3.1) reads:

$$S_{xy}(z) - S_y(z)\tilde{A}(z)$$

but only non-positive powers of z of the latter expression obey (3.1), namely:

$$\Big\lfloor S_{xy}(z) - S_y(z)\tilde{A}(z) \Big\rfloor_+ = 0$$

where $\lfloor \psi(z) \rfloor_+$ denotes non-positive powers of the series expansion of $\psi(z)$. Since $S_y(z)$ can be factored:

$$\Big\lfloor S_{xy}(z) - \tilde{A}(z)B(z)B(1/z) \Big\rfloor_+ = 0$$

where, say, $B(z)$ is the transform of casual sequence (i.e. its Z transform has only non-positive powers).

$$\left\lfloor B(1/z)\left(\frac{S_{xy}(z)}{B(1/z)} - B(z)\tilde{A}(z)\right) \right\rfloor_+ = 0$$

Since $B(1/z)$ is the transform of anti-casual sequence, the only way this equation can be satisfied is when $S_{xy}(z)/B(1/z) - \tilde{A}(z)B(z)$ is the transform of anti-casual sequence as well, by other words:

$$\Big\lfloor S_{xy}(z)/B(1/z) - B(z)\tilde{A}(z) \Big\rfloor_+ = 0$$

But $\tilde{A}(z)B(z)$ corresponds to a casual sequence, that is

$$\Big\lfloor \tilde{A}(z)B(z) \Big\rfloor_+ = \tilde{A}(z)B(z),$$

so the response of the optimal casual filter can be calculated from:

$$\tilde{A}(z) = \frac{1}{B(z)} \left\lfloor \frac{S_{xy}(z)}{B(z^{-1})} \right\rfloor_+ \tag{3.2}$$

The mean square error can be calculated as in the previous case.

(c) Again orthogonality implies

$$R_{xy}(m) - \sum_{k=0}^{p} R_y(m-k)\tilde{a}_k = 0, \quad 0 \le m \le p \tag{3.3}$$

Define the vectors:

$$\rho_{xy} = \begin{bmatrix} R_{xy}(0) \\ R_{xy}(1) \\ \vdots \\ R_{xy}(p) \end{bmatrix} \quad \tilde{a} = \begin{bmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_p \end{bmatrix}$$

and the correlation matrix R^y, so that:

$$R^y(i,j) = R_y(i-j), \quad 0 \le i,j \le p$$

Now (3.3) has the vector formulation:

$$R^y \tilde{a} = \rho_{xy}$$

and assuming $R^y > 0$, one can obtain the optimal filter:

$$\tilde{a} = \left[R^y\right]^{-1} \rho_{xy}$$

The mean square error can be also calculated using these vector notations. Let Y^n denote the vector of $(p+1)$ last samples of Y_n, i.e.

$$Y^n = \begin{bmatrix} Y_n \\ Y_{n-1} \\ \vdots \\ Y_{n-p} \end{bmatrix}$$

$$\begin{aligned}\mathbb{E}(X_n - a^* Y^n)^2 &= R_x(0) - \rho_{xy}^* \tilde{a} - \tilde{a}^* \rho_{xy} + \tilde{a}^* R^y \tilde{a} = \\ &= R_x(0) - \rho_{xy}^* \big[R^y\big]^{-1} \rho_{xy}\end{aligned}$$

Problem 3.2

(a) Consider the sequence $(X_n)_{n\in\mathbb{Z}}$, given by:

$$X_n = \sum_{k=-\infty}^{n} a^{n-k}\varepsilon_k.$$

These series are convergent (for any fixed n, in $\mathbb{L}^2$) since

$$\xi_m^{(n)} = \sum_{k=-m}^{n} a^{n-k}\varepsilon_k$$

is a Cauchy sequence and $\mathbb{L}^2$ is a complete space. Indeed (for, say, $m \ge \ell$)

$$E\big(\xi_m^{(n)} - \xi_\ell^{(n)}\big)^2 = E\Big(\sum_{k=-m}^{-\ell} a^{n-k}\varepsilon_k\Big)^2 = \sum_{k=-m}^{-\ell} a^{2(n-k)} \le \sum_{k=-\infty}^{-\ell} a^{2(n-k)} =$$

$$a^{2n}\sum_{k=\ell}^{\infty} a^{2k} = a^{2(n+\ell)}/(1-a^2) \xrightarrow{\ell\to\infty} 0.$$

Clearly X satisfies $X_n = aX_{n-1} + \varepsilon_n$, $n \in \mathbb{Z}$ and it is stationary. Indeed $EX_n = 0$ for all n and

$$R_x(0) = EX_n^2 = E\Big(\sum_{k=-\infty}^{n} a^{n-k}\varepsilon_k\Big)^2 = \sum_{k=-\infty}^{n} a^{2(n-k)} = \sum_{k=0}^{\infty} a^{2k} = \frac{1}{1-a^2}, \quad \forall n.$$

Then

$$EX_nX_{n+1} = EX_n(aX_n + \varepsilon_{n+1}) = aR_x(0)$$

and by induction $EX_nX_{n+m} = a^{|m|}R_x(0)$, that is the covariance function depends only on the time shift.

(b) The pair (X, Y) is stationary as well. Clearly $EY_n = EX_n = 0$, and $R_y(k) := EY_nY_{n+k} = R_x(k) + \sigma^2\delta(k)$ and $R_{xy}(k) := EX_nY_{n+k} = R_x(k)$.

(c) Find the spectral density of X

$$S_x(\lambda) = \sum_{\ell=-\infty}^{\infty} \frac{a^{|\ell|}}{1-a^2} e^{-j\lambda\ell} = ... = \frac{1}{1 - 2a\cos\lambda + a^2}$$

and

$$S_y(\lambda) = S_x(\lambda) + 1, \quad S_{xy}(\lambda) = S_x(\lambda)$$

and using the formulas from the previous problem we obtain:

$$A(\lambda) = \frac{S_x}{S_x+1} = \frac{1}{1+1-2a\cos\lambda+a^2} = \frac{1}{2-2a\cos\lambda+a^2}$$

The minimal mean square error is readily calculated:

$$\mathbb{E}(X_n-\widehat{X}_n)^2 = R_x(0) - \sum_k R_{xy}(k)a_k = \frac{1}{2\pi}\int\left[S_x(\lambda) - \frac{|S_{xy}(\lambda)|^2}{S_y(\lambda)}\right]d\lambda$$
$$= \frac{1}{2\pi}\int\frac{S_x(\lambda)}{S_x(\lambda)+1}d\lambda = \frac{1}{2\pi}\int\frac{1}{2-2a\cos\lambda+a^2}d\lambda = \frac{1}{\sqrt{4+a^4}}$$

(d) Using the formula from the previous problem:

$$\tilde{A}(z) = \frac{1}{B(z)}\left\lfloor\frac{S_{xy}(z)}{B(z^{-1})}\right\rfloor_+ \tag{3.4}$$

where $B(z)$ is the casual term in the factorization of

$$S_y(z) = B(z)B(z^{-1})$$

In this case

$$S_y(z) = S_x(z)+1 = \frac{1}{(1-az)(1-az^{-1})}+1 = \frac{a}{\gamma}\frac{(1-\gamma z^{-1})(1-\gamma z)}{(1-az^{-1})(1-az)}$$

where

$$\gamma := \frac{2+a^2-\sqrt{4+a^4}}{2a}$$

Note that $|\gamma|<1$ for $|a|<1$. So $B(z)$ is identified as:

$$B(z) := \sqrt{\frac{a}{\gamma}\frac{1-\gamma z^{-1}}{1-az^{-1}}}$$

Substitute this into (3.4):

$$\tilde{A}(z) = \sqrt{\frac{\gamma}{a}\frac{1-az^{-1}}{1-\gamma z^{-1}}}\left\lfloor\frac{\sqrt{\gamma/a}(1-az)/(1-\gamma z)}{(1-az)(1-az^{-1})}\right\rfloor_+ =$$
$$= \sqrt{\frac{\gamma}{a}\frac{1-az^{-1}}{1-\gamma z^{-1}}}\left\lfloor\frac{\sqrt{\gamma/a}}{1-a\gamma}\left(\frac{1}{1-az^{-1}} - \frac{1}{1-\gamma^{-1}z^{-1}}\right)\right\rfloor_+ =$$
$$= \frac{\gamma}{a}\frac{1}{1-a\gamma}\frac{1-az^{-1}}{1-\gamma z^{-1}}\frac{1}{1-az^{-1}} = \frac{\gamma}{a(1-a\gamma)}\frac{1}{1-\gamma z^{-1}} =$$
$$= \frac{2+a^2-\sqrt{4+a^4}}{a^2(\sqrt{4+a^4}-a^2)}\frac{1}{1-\gamma z^{-1}} = \frac{a^2-2+\sqrt{4+a^4}}{2a^2}\frac{1}{1-\gamma z^{-1}}$$

The filtering error can be calculated directly, using the formulas similar to the previous case. It also equals the steady state error of the Kalman filter (why?).

(e) Recall that

$$R_y(m) = R_x(m) + 1\cdot\delta(m), \quad R_{xy}(m) = R_x(m)$$

hence ($\tilde{a}$ now denotes a 2-by-1 vector)

$$\tilde{a} = \begin{pmatrix} R_x(0)+1 & R_x(1) \\ R_x(1) & R_x(0)+1 \end{pmatrix}^{-1} \begin{pmatrix} R_x(0) \\ R_x(1) \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{1-a^2}+1 & \frac{a}{1-a^2} \\ \frac{a}{1-a^2} & \frac{1}{1-a^2}+1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{1-a^2} \\ \frac{a}{1-a^2} \end{pmatrix} = ... = \begin{pmatrix} 2 \\ a \end{pmatrix} \frac{1}{4-a^2}$$

The corresponding error is:

$$\mathbb{E}(X_n - \widehat{X}_n)^2 = R_x(0) - \rho_{xy}^* \tilde{a} = \frac{1}{1-a^2} - \left(\frac{1}{1-a^2}, \frac{a}{1-a^2} \right) \tilde{a}$$

$$= ... = \frac{2}{4-a^2}$$

(f) The Kalman filter equations are

$$\begin{aligned} \widehat{X}_n &= a\widehat{X}_{n-1} + P_n(Y_n - a\widehat{X}_{n-1}) \\ P_n &= \frac{a^2 P_{n-1}+1}{a^2 P_{n-1}+2}, \quad n \ge 1 \end{aligned} \tag{3.5}$$

subject to $\widehat{X}_0 = 0$ and $P_0 = 1/(1-a^2)$.

(g) First note that $P_n \in [0,1]$, since by optimality $P_n \le E(Y_n - X_n)^2 = E\xi_n^2 = 1$. Let P_∞ be the unique nonnegative solution of

$$P_\infty = \frac{a^2 P_\infty + 1}{a^2 P_\infty + 2}, \tag{3.6}$$

which is (the other solution is always negative)

$$P_\infty = \frac{a^2 - 2 + \sqrt{4+a^4}}{2a^2}.$$

The sequence $D_n := |P_n - P_\infty|$ satisfies

$$D_n = \left| -\frac{1}{a^2 P_{n-1}+2} + \frac{1}{a^2 P_\infty + 2} \right| = \frac{a^2 D_{n-1}}{(a^2 P_{n-1}+2)(a^2 P_\infty + 2)} \le$$

$$\frac{a^2 D_{n-1}}{2\big(a^2(a^2 - 2 + \sqrt{4+a^4})/(2a^2) + 2\big)} = \frac{a^2 D_{n-1}}{a^2 + 2 + \sqrt{4+a^4}} \le \frac{1}{2} D_{n-1}$$

and thus $\lim_{n\to\infty} D_n = 0$.

The "steady state" filter is then

$$\begin{aligned} \widehat{X}_n &= a\widehat{X}_{n-1} + \frac{a^2 - 2 + \sqrt{4+a^4}}{2a^2}(Y_n - a\widehat{X}_{n-1}) = \\ &= a\left(1 - \frac{a^2 - 2 + \sqrt{4+a^4}}{2a^2}\right)\widehat{X}_{n-1} + \frac{a^2 - 2 + \sqrt{4+a^4}}{2a^2} Y_n = \\ &= \underbrace{\frac{a^2 + 2 - \sqrt{4+a^4}}{2a}}_{\equiv \gamma} \widehat{X}_{n-1} + \frac{a^2 - 2 + \sqrt{4+a^4}}{2a^2} Y_n. \end{aligned}$$

Note that this recursion is exactly the one which was obtained via Kolmogorov-Wiener approach in the appropriate setup.

(h) The best estimate is obtained via optimal smoothing in (c); next is the filter, based on all the observations till n in (d). The Kalman filter in (f) is inferior to the latter filter for any fixed n, but is asymptotically equivalent to it as $n \to \infty$. The worst is of course the filter in (e) that takes into account only two observations. Note that for $a = 0$ (i.e. the signal X_n is an i.i.d. sequence (*white noise*), all the estimates attain the same error $P = 1/2$.

(i) The error recursion for the Kalman filter becomes:

$$P_n = a^2 P_{n-1} - \frac{(a^2 P_{n-1})^2}{a^2 P_{n-1} + 1} = \frac{a^2 P_{n-1}}{a^2 P_{n-1} + 1}$$

This can be explicitly solved (define e.g. $Q_n = 1/P_n$ and obtain a linear recursion for Q_n) and verified that $\lim_{n\to\infty} P_n = 0$.

Problem 3.3

Recall that

$$X_n = \begin{cases} X_{n-1}, & \text{with prob. } p \\ -X_{n-1}, & \text{with prob. } 1-p \end{cases}$$

with $\mathbb{P}\{X_0 = \ell\} = \mathbb{P}\{X_0 = -\ell\} = 1/2$. Let $(\xi_n)_{n\ge 1}$ be an i.i.d. binary sequence of r.v. with

$$\mathbb{P}\{\xi_n = 1\} = 1 - \mathbb{P}\{\xi_n = 0\} = p.$$

Clearly $(X_n)_{n\ge 1}$ can be generated by:

$$X_n \quad = \quad (2\xi_n - 1)X_{n-1}, \quad \text{subject to } X_0$$

Rewrite this equation as:

$$X_n = (2\mathbb{E}\xi_n - 1)X_{n-1} + 2X_{n-1}(\xi_n - \mathbb{E}\xi_n) = (2p-1)X_{n-1} + 2X_{n-1}(\xi_n - p)$$

Define $\eta_n = 2X_{n-1}(\xi_n - p)$, then:

$$\mathbb{E}\eta_n = 2\mathbb{E}X_{n-1}\mathbb{E}(\xi_n - p) = 0$$

and (say $n > m$)

$$\mathbb{E}\eta_n\eta_m = 4\mathbb{E}X_{n-1}X_{m-1}(\xi_m - p)\mathbb{E}(\xi_n - p) = 0$$

$$\mathbb{E}\eta_n^2 = 4\mathbb{E}X_{n-1}^2\mathbb{E}(\xi_n - p)^2 = 4\ell^2(1-p)p$$

Moreover for $k < n$, X_k and η_n are uncorrelated.

Introduce an auxiliary pair of processes (X, Y), generated by

$$\begin{aligned} \widetilde{X}_n &= (2p-1)\widetilde{X}_{n-1} + \widetilde{\eta}_n, \quad \text{subject to } X_0 \\ \widetilde{Y}_n &= \widetilde{X}_n + \varepsilon_n. \end{aligned} \tag{3.7}$$

where $\widetilde{\eta}$ is a white noise sequence with the same mean and variance as η.

The orthogonal projection $\widehat{X}_n = \widehat{E}(\widetilde{X}_n|\widetilde{Y}_1^n)$ is generated by the Kalman filter

$$\begin{aligned} \widehat{X}_n &= (2p-1)\widehat{X}_{n-1} + P_n(\widetilde{Y}_n - (2p-1)\widehat{X}_{n-1}), \quad n \ge 1 \\ P_n &= \frac{(2p-1)^2 P_{n-1} + 4\ell^2 p(1-p)}{(2p-1)^2 P_{n-1} + 4\ell^2 p(1-p) + 1} \end{aligned} \tag{3.8}$$

subject to $\widehat{X}_0 = 0$ and $P_0 = \ell^2$. If these equation are applied to the original observations process Y, the obtained linear functional $\widehat{X}_n(Y_1^n)$ can be considered as an estimate for X_n. Does the obtained filter realizes the orthogonal projection $\widehat{E}(X_n|Y_1^n)$ for the original model?

Let $L(Y_1^n)$ denote any linear functional of $\{Y_1, ..., Y_n\}$, then

$$E\big(X_n - L(Y_1^n)\big)^2 = E\big(\widetilde{X}_n - L(\widetilde{Y}_1^n)\big)^2 \geq \\ E\big(\widetilde{X}_n - \widehat{X}_n(\widetilde{Y}_1^n)\big)^2 = E\big(X_n - \widehat{X}_n(Y_1^n)\big)^2, \quad (3.9)$$

where the equalities hold, since (X, Y) and $\widetilde{X}, \widetilde{Y})$ have the same correlation structure by construction. The inequality (3.9) implies that $\widehat{X}_n(Y_1^n)$ is optimal and hence realizes the orthogonal projection.

Problem 3.4

Denote $\mu = E\eta_n$. Rewrite the eq. for Y_n as:

$$Y_n = \mu X_{n-1} + \xi_n + (\eta_n - \mu)X_{n-1}$$

Set $\widetilde{\xi}_n := \xi_n + (\eta_n - \mu)X_{n-1}$. Then:

$$\mathbb{E}\widetilde{\xi}_n = 0, \quad \mathbb{E}\widetilde{\xi}_n\widetilde{\xi}_k = \delta_{n-k}(\sigma_\xi^2 + \sigma_\eta^2 V_{n-1})$$

where $V_n = \mathbb{E}X_n^2$ satisfies $(n \geq 1)$

$$V_n = a^2 V_{n-1} + \sigma_\varepsilon^2, \quad \text{subject to } V_0 = 1$$

Moreover $\widetilde{\xi}_n$ is uncorrelated with X_m, $m < n$. Consider the model:

$$\begin{aligned} X_n &= aX_{n-1} + \varepsilon_n \\ Y_n &= \mu X_{n-1} + \widetilde{\xi}_n, \quad \text{s.t. } X_0 \end{aligned} \quad (3.10)$$

The optimal linear estimate is given by the Kalman filter

$$\begin{aligned} \widehat{X}_n &= a\widehat{X}_{n-1} + \frac{a\mu P_{n-1}}{\mu^2 P_{n-1} + \sigma_\xi^2 + \sigma_\eta^2 V_{n-1}}(Y_n - \mu\widehat{X}_{n-1}) \\ P_n &= a^2 P_{n-1} + \sigma_\varepsilon^2 - \frac{[a\mu P_{n-1}]^2}{\mu^2 P_{n-1} + \sigma_\xi^2 + \sigma_\eta^2 V_{n-1}} \end{aligned}$$

subject to $\widehat{X}_0 = 0$, $P_0 = 1$.

Problem 3.5

Simple Solution:

Define an augmented state vector $\vartheta_n \in \mathbb{R}^{(p+q)\times 1}$

$$\vartheta_n = \begin{bmatrix} \theta_n \\ \theta_{n-1} \\ \vdots \\ \theta_{n-p+1} \\ \varepsilon_n \\ \vdots \\ \varepsilon_{n-q+1} \end{bmatrix}$$

Introduce $A \in \mathbb{R}^{(p+q)\times(p+q)}$ and $B, C \in \mathbb{R}^{(p+q)\times 1}$:

$$A = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_p & b_1 & \cdots & b_q \\ 1 & 0 & 0 & \cdots & & & 0 \\ 0 & 1 & 0 & \cdots & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & \cdots & & & 0 \\ 0 & 0 & & \cdots & 1 & & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & & \cdots & & 1 & 0 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Consider the vector difference equations ($n \ge p$):

$$\begin{aligned} \vartheta_n &= A\vartheta_{n-1} + B\varepsilon_n \\ \xi_n &= C^*\vartheta_n + \upsilon_n \end{aligned}$$

where ϑ_{p-1} is a vector of initial conditions (θ_0^{p-1} and ε_0^{q-1}). Clearly the first component of the vector ϑ_n coincides with θ_n for any $n \ge p$, i.e. $\vartheta_n(1) = \theta_n$. Note that $\mathbb{E}\vartheta_k\varepsilon_n = 0$, $k < n$ and hence the obtained model suits the Kalman filter setting: let $\widehat{\vartheta}_n = \widehat{\mathbb{E}}(\vartheta_n|\xi_0^n)$ and $\widehat{\theta}_n = \widehat{\mathbb{E}}(\theta_n|\xi_0^n)$, then ($n \ge p$):

$$\begin{aligned} \widehat{\vartheta}_n &= A\widehat{\vartheta}_{n-1} + \frac{(AP_{n-1}A^* + BB^*\sigma^2)C(\xi_n - C^*A\widehat{\vartheta}_{n-1})}{C^*AP_{n-1}A^*C + C^*BB^*C\sigma^2 + \sigma_\upsilon^2} \\ P_n &= AP_{n-1}A^* + \sigma^2 BB^* - \\ &\quad - \frac{(AP_{n-1}A^* + BB^*\sigma^2)CC^*(AP_{n-1}A^* + BB^*\sigma^2)}{C^*AP_{n-1}A^*C + C^*BB^*C\sigma^2 + \sigma_\upsilon^2} \\ \widehat{\theta}_n &= C^*\widehat{\vartheta}_n \end{aligned}$$

subject to $\widehat{\vartheta}_{p-1} = 0$ and [2] $P_{p-1} = I\sigma^2$.

Note that the estimates of $\{\theta_{n-1}, ..., \theta_{n-p+1}\}$ and also of the driving noise $\{\varepsilon_n, ..., \varepsilon_{n-q+1}\}$ are obtained as a byproduct.

Advanced Solution[3]: In the previous solution version to generate $\widehat{\theta}_n$ one has to propagate $(p+q)$-dimensional vector recursion. More delicate arguments lead to a

[2] θ_0^{p-1} and ε_0^{q-1} are assumed to form a vector of i.i.d. components with zero mean and variance σ^2

filter of lower dimensions. Consider a sequence

$$\theta_n = -\sum_{k=1}^{p} a_k \theta_{n-k} + \sum_{k=0}^{p-1} b_k \varepsilon_{n-k} \tag{3.11}$$

where $(\varepsilon_n)_{n\geq 0}$ is standard white noise sequence. Note that the original model of the problem (i.e. $q \leq p$) is obtained by setting appropriate b_k's to zero.

Below we derive a state space model of order p, which generates the same sequence.

Lemma 3.1. *Let η_n be a vector process generated by the recursion:*

$$\eta_n = A\eta_{n-1} + B\varepsilon_n, \quad n \geq 0 \tag{3.12}$$

where $(\varepsilon_n)_{n\geq 0}$ is an i.i.d. scalar standard Gaussian sequence and

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ \vdots & & & ... & \vdots \\ 0 & 0 & & ... & 1 \\ -a_n & -a_{n-1} & & ... & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \\ \beta_n \end{pmatrix}$$

$$\begin{aligned} \beta_1 &= b_0 \\ \beta_j &= b_{j-1} - \sum_{\ell=1}^{j-1} a_{j-\ell}\beta_\ell, \quad j = 2, ..., n \end{aligned}$$

Then $\theta_n \equiv \eta_n(1)$.

Proof. Verify the equivalence between (3.11) and (3.12) as maps, i.e. show that both generate the same output $y(t), t = 0, 1, ...$ for the same input $x(t), t = 0, 1, ...$ Use the Z-transform system representation:

Starting from (3.12):

$$\begin{aligned} y_i(t) &= y_{i+1}(t-1) + \beta_i x(t), \quad i = 1, ..., n-1 \\ y_n(t) &= -\sum_{k=0}^{n-1} a_{n-k} y_{k+1}(t-1) + \beta_n x(t) \end{aligned}$$

or

$$\begin{aligned} Y_i(z) &= z^{-1} Y_{i+1}(z) + \beta_i X(z), \quad i = 1, ..., n-1 \\ Y_n(z) &= -z^{-1} \sum_{k=0}^{n-1} a_{n-k} Y_{k+1}(z) + \beta_n X(z) \end{aligned}$$

Then[4] for $i = 1, ..., n-1$

$$Y_{i+1}(z) = z\big[Y_i(z) - \beta_i X(z)\big] = ... = z^i Y_1(z) - \sum_{j=1}^{i} z^{i-j+1} \beta_j X(z) \tag{3.13}$$

---

[3]This solution is for advanced reading.

[4]the convention $\sum_1^0 = 0$ is followed

On the other hand:

$$\begin{aligned} Y_n(z) &= -z^{-1}\sum_{k=0}^{n-1} a_{n-k}Y_{k+1}(z) + \beta_n X(z) = \qquad (3.14) \\ &= -z^{-1}\sum_{k=0}^{n-1} a_{n-k}z^k Y_1(z) + z^{-1}\sum_{k=0}^{n-1} a_{n-k}\sum_{j=1}^{k} z^{k-j+1}\beta_j X(z) + \\ &\quad +\beta_n X(z) \end{aligned}$$

Equating the (3.13) with $i = n-1$ and (3.14) we arrive at:

$$z^n Y_1(z) + \sum_{k=0}^{n-1} a_{n-k}z^k Y_1(z) = \sum_{j=1}^{n-1} z^{n-j+1}\beta_j X(z) + \\ + \sum_{k=0}^{n-1} a_{n-k}\sum_{j=1}^{k} z^{k-j+1}\beta_j X(z) + z\beta_n X(z)$$

or ($a_0 := 1$)

$$\sum_{k=0}^{n} a_{n-k}z^k Y_1(z) = \sum_{j=1}^{n-1} z^{n-j+1}\beta_j X(z) + \\ + \sum_{k=0}^{n-1} a_{n-k}\sum_{j=1}^{k} z^{k-j+1}\beta_j X(z) + z\beta_n X(z)$$

and

$$Y_1(z)\sum_{k=0}^{n} a_k z^{-k} = \sum_{j=1}^{n-1} z^{-j+1}\beta_j X(z) + \\ + \sum_{k=1}^{n-1} a_{n-k}\sum_{j=n-k+1}^{n} z^{-j+1}\beta_{j-n+k}X(z) + z^{-(n-1)}\beta_n X(z) \qquad (3.15)$$

Equating the right hand side of (3.15) to $X(z)P_{n-1}(z)$ and comparing powers of z we obtain the desired result:

$$\begin{aligned} z^0 &: \beta_1 = b_0 \\ z^{-1} &: \beta_2 + a_1\beta_1 = b_1 \\ \vdots &: \vdots \\ z^{-(n-1)} &: \sum_{k=1}^{n-1} a_{n-k}\beta_k + \beta_n = b_{n-1} \end{aligned}$$

□

Let us demonstrate the latter approach:

Example

Let ξ_n be a stationary random process with zero mean and the spectrum density:

$$f(\lambda) = \left| \frac{1 + e^{-j\lambda}}{1 + 1/2e^{-j\lambda} + 1/2e^{-2j\lambda}} \right|^2$$

Find the optimal linear extrapolation estimate of ξ_t on the basis of $\{\xi_0, ..., \xi_s\}$, $m(t,s) = \widehat{\mathbb{E}}(\xi_t|\xi_0^s)$.

Find the state space representation for ξ_n. Here $b_0 = 1, b_1 = 1$ and $a_0 = 1, a_1 = 1/2, a_2 = 1/2$ and thus $\beta_1 = b_0 = 1$ and $\beta_2 = b_1 - 1/2 \cdot 1 = 1/2$. Let $(\eta_1(t), \eta_2(t))$ be generated by

$$\begin{aligned} \eta_1(t) &= \eta_2(t-1) + \varepsilon(t) \\ \eta_2(t) &= -1/2\eta_1(t-1) - 1/2\eta_2(t-1) + 1/2\varepsilon(t) \end{aligned}$$

where $\varepsilon(t)$ is a standard i.i.d. Gaussian sequence.

Set $\xi_t = \eta_1(t)$ and $\theta_t = \eta_2(t)$. Then ξ_t has the spectral density $f(\lambda)$ and:

$$\begin{aligned} \xi_t &= \theta_{t-1} + \varepsilon(t) \\ \theta_t &= -1/2\theta_{t-1} - 1/2\xi_{t-1} + 1/2\varepsilon(t) \end{aligned} \tag{3.16}$$

And thus $(t > s)$

$$\begin{aligned} m(t,s) &= \mu(t-1,s) \\ \mu(t,s) &= -1/2\mu(t-1,s) - 1/2m(t-1,s) \end{aligned}$$

subject to $m(s,s) = \xi_s$ and $\mu(s,s) = \mu(s) = \widehat{\mathbb{E}}(\theta_s|\xi_0^s)$.

The filtering estimate $\mu(s)$ satisfies $(k \le s)$:

$$\begin{aligned} \mu_k &= -1/2\mu_{k-1} - 1/2\xi_{k-1} + \frac{1/2 - 1/2P_{k-1}}{P_{k-1}+1}(\xi_k - \mu_{k-1}) \\ P_k &= 1/4P_{k-1} + 1/4 - \frac{(1/2 - 1/2P_{k-1})^2}{P_{k-1}+1} = \frac{P_{k-1}}{P_{k-1}+1} \end{aligned} \tag{3.17}$$

The initial conditions for this filter can be recovered due to stationarity assumptions. Let $d_{11} = \mathbb{E}\theta_t^2$, $d_{12} = \mathbb{E}\xi_t\theta_t$ and $d_{22} = \mathbb{E}\xi_t^2$. From (3.16):

$$\begin{aligned} d_{22} &= d_{11} + 1 \\ d_{11} &= 1/4d_{11} + 1/4d_{22} + 1/4 + 1/2d_{12} \\ d_{12} &= -1/2d_{11} - 1/2d_{12} + 1/2 \end{aligned}$$

so that:

$$d_{11} = 1, \quad d_{12} = 0, \quad d_{22} = 2$$

and the initial condition for the filter (3.17):

$$\mu_0 = 0, \quad P(0) = 1$$

Problem 3.6

The Riccati equation of the Kalman filter is transformed by Matrix Inversion Lemma into:

$$\begin{aligned} P_{n+1} &= aP_na^* + bb^* - aP_nA^*(AP_nA^* + BB^*)^{-1}AP_na^* = \\ &= bb^* + a\left\{P_n - P_nA^*(AP_nA^* + BB^*)^{-1}AP_n\right\}a^* = \\ &= bb^* + a\Gamma_n^{-1}a^* \end{aligned}$$

where

$$\Gamma_n = P_n^{-1} + A^*(BB^*)^{-1}A = J_n + A^*(BB^*)^{-1}A$$

$$\begin{aligned} J_{n+1} &:= & P_{n+1}^{-1} = \Big\{bb^* + \big(a^{-*}\Gamma_n a^{-1}\big)^{-1}\Big\}^{-1} = \\ &= & F_n - F_n b(I + b^* F_n b)^{-1} b^* F_n \end{aligned}$$

where $F_n := a^{-*}\Gamma_n a^{-1}$. Summarizing all the equations, J_n can be propogated by:

$$\begin{aligned} J_{n+1} &= F_n - F_n b(I + b^* F_n b)^{-1} b^* F_n \\ F_n = &= a^{-*}(J_n + A^*(BB^*)^{-1}A)a^{-1} \end{aligned}$$

The validity of the Matrix Inversion Lemma is verified directly:

$$\begin{aligned} AA^{-1} &= (B^{-1} + CD^{-1}C^*)(B - BC(D + C^*BC)^{-1}C^*B) = \\ &= I + CD^{-1}C^*B - C(D + C^*BC)^{-1}C^*B - \\ &\quad -CD^{-1}C^*BC(D + C^*BC)^{-1}C^*B = \\ &= I + CD^{-1}C^*B - C\left\{I + D^{-1}C^*BC\right\}(D + C^*BC)^{-1}C^*B = \\ &= I + CD^{-1}C^*B - CD^{-1}C^*B \equiv I \end{aligned}$$

Problem 3.7

Clearly x_t are the orthogonal projections of a standard random vector x on $\{y_1, ..., y_t\}$, where

$$y_{t+1} = a_{t+1}x + \sqrt{\alpha}\varepsilon_{t+1}$$

with ε_t being standard white noise, independent of x. So x_k is the orthogonal projection of x on $y = Ax + \sqrt{\alpha}\varepsilon$, where ε is a standard random vector.

Then

$$Q = \mathbb{E}(xy^*)\big(\mathbb{E}(yy^*)\big)^{-1} = A^*\big(\alpha I + AA^*\big)^{-1} = \big(\alpha I + A^*A\big)^{-1}A^*$$

since

$$A^*(\alpha I + AA^*) = (\alpha I + A^*A)A^*$$

The second statement follows form the fact that $\gamma_k = \mathbb{E}(x - x_k)(x - x_k)^*$

$$\begin{aligned} &\gamma_k = \mathbb{E}(xx^*) - \mathbb{E}(xy^*)\mathbb{E}^{-1}(yy^*)\mathbb{E}(yx^*) = I - A^*(\alpha I + AA^*)^{-1}A = \\ &= (I + A^*A/\alpha)^{-1} = (I\alpha + A^*A)^{-1}\alpha \end{aligned}$$