4F7-STATISTICAL SIGNAL ANALYSIS

Examples Paper Solutions

Exercise 1. The ARMA(2,2) is

$$\begin{aligned} X_n &= a_1 X_{n-1} + a_2 X_{n-2} + b_0 W_n + b_1 W_{n-1} \\ &= a_1 X_{n-1} + b_0 W_n + Z_{n-1}. \end{aligned}$$

where $Z_{n-1} = a_2 X_{n-2} + b_1 W_{n-1}$. The state equation is

$$\begin{bmatrix} X_n \\ Z_n \end{bmatrix} = \begin{bmatrix} a_1 & 1 \\ a_2 & 0 \end{bmatrix} \begin{bmatrix} X_{n-1} \\ Z_{n-1} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} W_n.$$

Note the hidden state is the vector

$$\begin{bmatrix} X_n \\ Z_n \end{bmatrix}.$$

The observation equation is

$$Y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} X_n \\ Z_n \end{bmatrix}.$$

Exercise 2. The Gaussian AR(P) (or the ARMA(P,0)) model is

$$X_n = a_1 X_{n-1} + \cdots + a_P X_{n-P} + b W_n$$

where $\{W_n\}$ are independent and identically distributed Gaussian random variables with mean 0 and variance 1. In state-space form, the state equation is

$$\begin{bmatrix} X_n \\ \vdots \\ X_{n-P+1} \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_P \\ 1 & 0 & \cdots \\ & & \end{bmatrix} \begin{bmatrix} X_{n-1} \\ \vdots \\ X_{n-P} \end{bmatrix} + \begin{bmatrix} b \\ 0 \\ \vdots \end{bmatrix} W_n.$$

The observation equation is

$$Y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} X_n \\ Z_n \end{bmatrix}.$$

When the model is in stationarity, the joint probability density function of $(X_{n-P+1}, \ldots, X_n)$ is Gaussian and its mean m and covariance R does not change with time n. Taking the expectation of both sides of the state equation gives

$$m = \Lambda m + \begin{bmatrix} b \\ 0 \\ \vdots \end{bmatrix} 0$$

so its mean m is 0 assuming $\sum_{i=1}^{P} a_i \neq 1$. Computing the covariance of both sides of the state equation yields

$$R = \Lambda R \Lambda^{\mathrm{T}} + \begin{bmatrix} b \\ 0 \\ \vdots \end{bmatrix} \begin{bmatrix} b \\ 0 \\ \vdots \end{bmatrix}^{\mathrm{T}}$$

which must be solved to find all the elements of R.

When $P = 2$, $\Lambda = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix}$ and solve

$$\begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} \begin{bmatrix} a_1 & 1 \\ a_2 & 0 \end{bmatrix} + \begin{bmatrix} b^2 & 0 \\ 0 & 0 \end{bmatrix}$$

subject to $r_{1,2} = r_{2,1}$. This gives

$$r_{1,1} = \left(1 - a_1^2 - \frac{2a_1^2 a_2}{1 - a_2} - a_2^2\right)^{-1} b^2$$
$$r_{1,2} = \frac{a_1}{1 - a_2} r_{1,1}.$$

Also $r_{2,2} = r_{1,1}$.

Exercise 3. Proving

$$K\left[aX + bU + c \mid Y_{1:n}\right] = aK\left[X \mid Y_{1:n}\right] + bK\left[U \mid Y_{1:n}\right] + c.$$

The lecture notes gives the following fact: let $\mathbf{p} = (\mathrm{Cov}(X, Y_1), \ldots, \mathrm{Cov}(X, Y_n))^{\mathrm{T}}$, let Σ be the square matrix with elements $[\Sigma]_{i,j} = \mathrm{Cov}(Y_i, Y_j)$ and $\mathbf{h} = (h_1, \ldots, h_n)^{\mathrm{T}}$. Let $(h_1, \ldots, h_n)^{\mathrm{T}}$ satisfy $\Sigma\mathbf{h} = \mathbf{p}$ then

$$\hat{X} = K\left[X \mid Y_{1:n}\right] = \mathbb{E}(X) + h_1(Y_1 - \mathbb{E}Y_1) + \ldots + h_n(Y_n - \mathbb{E}Y_n).$$

The optimal vector $\mathbf{h}$ that solves $K\left[aX \mid Y_{1:n}\right]$ must satisfy $\Sigma\mathbf{h} = a\mathbf{p}$. From this fact, it is apparent that $K\left[aX \mid Y_{1:n}\right] = aK\left[X \mid Y_{1:n}\right]$. Thus all we need to do is to prove the result

$$K\left[W + V \mid Y_{1:n}\right] = K\left[W \mid Y_{1:n}\right] + K\left[V \mid Y_{1:n}\right]$$

where W and V are random variables. The solution to $K\left[W+V \mid Y_{1:n}\right]$ must satisfy

$$\Sigma \mathbf{h} = \mathbf{q} + \mathbf{r}$$

where $\mathbf{q} = (\mathrm{Cov}(W, Y_1), \ldots, \mathrm{Cov}(W, Y_n))^{\mathrm{T}}$ and $\mathbf{r} = (\mathrm{Cov}(V, Y_1), \ldots, \mathrm{Cov}(V, Y_n))^{\mathrm{T}}$. (We have used the fact that $\mathrm{Cov}(W+V, Y_i) = \mathrm{Cov}(W, Y_i) + \mathrm{Cov}(V, Y_i)$.) Let constants $c_1, \ldots, c_n$ and $d_1, \ldots, d_n$ satisfy

$$\Sigma \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{q}, \qquad \Sigma \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \mathbf{r}$$

then

$$\begin{aligned} & K\left[W+V \mid Y_{1:n}\right] \\ & = \mathbb{E}(W) + \mathbb{E}(V) + h_1(Y_1 - \mathbb{E}Y_1) + \ldots + h_n(Y_n - \mathbb{E}Y_n) \\ & = \mathbb{E}(W) + \mathbb{E}(V) + (c_1 + d_1)(Y_1 - \mathbb{E}Y_1) + \ldots + (c_n + d_n)(Y_n - \mathbb{E}Y_n) \\ & = K\left[W \mid Y_{1:n}\right] + K\left[V \mid Y_{1:n}\right]. \end{aligned}$$

Proving

$$K\left[X \mid Y_{1:n}\right] = K\left[X \mid Y_{1:n-1}\right] + K\left[X \mid Y_n\right] - \mathbb{E}(X),$$

when $\mathrm{Cov}(Y_i, Y_n) = 0$ for $i < n$.

Σ is the $n \times n$ square matrix with elements $[\Sigma]_{i,j} = \mathrm{Cov}(Y_i, Y_j)$. Let S be the $n-1 \times n-1$ square matrix with elements $[S]_{i,j} = \mathrm{Cov}(Y_i, Y_j)$

for $i < n$ and $j < n$. Then

$$\Sigma = \begin{bmatrix} S & 0 \\ 0 & \operatorname{Cov}(Y_n, Y_n) \end{bmatrix}$$

and the vector $(h_1, \ldots, h_n)^{\mathrm{T}}$ that solves $\Sigma \mathbf{h} = \mathbf{p} = (p_1, \ldots, p_n)^{\mathrm{T}}$ also solves

$$S \begin{bmatrix} h_1 \\ \vdots \\ h_{n-1} \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_{n-1} \end{bmatrix} \quad \text{and} \quad h_n \operatorname{Cov}(Y_n, Y_n) = p_n.$$

Thus

$$\begin{aligned}
& K\left[X \mid Y_{1:n}\right] \\
=& \mathbb{E}(X) + h_1(Y_1 - \mathbb{E}Y_1) + \ldots + h_n(Y_n - \mathbb{E}Y_n) \\
=& \mathbb{E}(X) + h_1(Y_1 - \mathbb{E}Y_1) + \ldots + h_{n-1}(Y_{n-1} - \mathbb{E}Y_{n-1}) \\
& + \mathbb{E}(X) + h_n(Y_n - \mathbb{E}Y_n) \\
& - \mathbb{E}(X) \\
=& K\left[X \mid Y_{1:n-1}\right] + K\left[X \mid Y_n\right] - \mathbb{E}(X).
\end{aligned}$$

Exercise 4. Consider the state-space model

$$\begin{aligned}
X_n &= X_{n-1} \\
Y_n &= X_n + V_n
\end{aligned}$$

where $\{V_n\}_n \sim \mathrm{WN}(0, r)$, $\mathbb{E}(X_1) = 0$, $\mathbb{E}(X_1^2) = \sigma$. Moreover, X_1 and $\{V_n\}_n$ are uncorrelated. Find $K\left[X_n \mid Y_{1:n}\right]$ and compare the mean square error of this estimate to that of the sample average. Find the limiting mean square error of $K\left[X_n \mid Y_{1:n}\right]$ as $n \to \infty$.

The Kalman prediction and update equations for this specific state-space model: let $\hat{X}_n = K\left[X_n \mid Y_{1:n}\right]$, $\sigma_n = \mathbb{E}\left\{\left(\hat{X}_n - X_n\right)^2\right\}$. The prediction is

$$\bar{X}_{n+1} = K\left[X_{n+1} \mid Y_{1:n}\right] = \hat{X}_n$$

$$\bar{\sigma}_{n+1} = \mathbb{E}\left\{\left(\bar{X}_{n+1} - X_{n+1}\right)^2\right\} = \sigma_n$$

and the update is

$$\hat{X}_{n+1} = \hat{X}_n + \frac{\sigma_n}{\sigma_n + r}\left(Y_{n+1} - \hat{X}_n\right)$$

$$\sigma_{n+1} = \sigma_n\left(1 - \frac{\sigma_n}{\sigma_n + r}\right).$$

Initialise the mean square error calculation $\sigma_0 = \sigma$ so that σ_1 will the mean square error for $K\left[X_1 \mid Y_1\right]$.

Re-arrange the mean square error expression to

$$\sigma_{n+1} = \sigma_n\left(\frac{r}{\sigma_n + r}\right)$$

and $\sigma_{n+1} < \sigma_n$. The limiting mean square error is thus 0.

Since $X_n = X_{n-1} = \cdots = X_1$, $Y_n = X_1 + V_n$. Averaging gives $n^{-1}(Y_1 + \ldots + Y_n)$. The average is clearly unbiased, i.e.

$$\mathbb{E}\left\{n^{-1}(Y_1 + \ldots + Y_n)\right\} = \mathbb{E}\left(X_1\right).$$

Its mean square error can be calculated directly:

$$\begin{aligned}
&\mathbb{E}\left\{\left[n^{-1}(Y_1+\ldots+Y_n)-X_1\right]^2\right\}\\
=&\mathbb{E}\left\{\left[n^{-1}(Y_1-X_1+\ldots+Y_n-X_1)\right]^2\right\}\\
=&\mathbb{E}\left\{\left[n^{-1}(V_1+\ldots+V_n)\right]^2\right\}\\
=&r/n.
\end{aligned}$$

The $K\left[X_n \mid Y_{1:n}\right]$ of should be less than r/n since $K\left[X_n \mid Y_{1:n}\right]$ is optimally weights the data. For example, this is clear for $n = 1$ since $\sigma_1 = r\sigma_1/\left(\sigma_1 + r\right) < r$. Now show that if $\sigma_n < r/n$ then $\sigma_{n+1} < r/(n+1)$ to confirm the Kalman filter always defeats sample average:

$$\begin{aligned}
\sigma_{n+1} &= \sigma_n\left(\frac{r}{\sigma_n+r}\right)\\
&= r\left(\frac{\sigma_n}{\sigma_n+r}\right)\\
&< r\left(\frac{r/n}{r/n+r}\right)\\
&= \frac{r}{n+1}.
\end{aligned}$$

Exercise 5. The ARMA(2,2) model has been expressed as a state-space model with a vector valued hidden state at time n which is

$$\begin{bmatrix} X_n \\ Z_n \end{bmatrix}$$

while the observation equation was

$$Y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} X_n \\ Z_n \end{bmatrix}.$$

Note only the first component of the hidden state is observed. This is a Gaussian state-space model since the state is being driven by Gaussian noise: the fact that the observation is noiseless for the component of the state that is observed does not matter, i.e. it is still a Gaussian state-space model.

The lectures only presents the Kalman equations for a scalar valued state and observation process. Moreover, the Kalman filter also calculates $p(y_{k+1} \mid y_0, \ldots, y_k)$ sequentially. Thus $p(y_0, \ldots, y_n)$ can be calculated using the output of the Kalman filter as follow:

$$p(y_0, \ldots, y_n) = p(y_n \mid y_0, \ldots, y_{n-1}) \ldots p(y_1 \mid y_0) p(y_0).$$

You only need to make the remark that the Kalman equations for a vector valued hidden state process could be similarly applied to get $p(y_0, \ldots, y_n)$.

Exercise 6. The hidden Markov model that describes the series of outcomes $\{Y_1, Y_2, \ldots\}$ observed by the player: The state process is $X_n \in \{1, 2\}$ where 1 indicates a fair dice. The state transition probability matrix is

$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{bmatrix}.$$

Let the probability mass function of X_1 be $\lambda = [\lambda_1, \lambda_2]^{\mathrm{T}}$ where λ_1 is the probability that $X_1 = 1$.

The observation process is a sequence of discrete random variables Y_n where $g(x_n, y_n)$ is a probability mass function of Y_n given $X_n = x_n$. That is, $g(1,1) = \cdots = g(1,6) = 1/6$ and $g(2,1) = \cdots = g(2,5) = 0.1$ and $g(2,6) = 0.5$. This completes the description of the hidden Markov model.

Exercise 7. The probability of observing the sequence $(x_1, y_1, \ldots, x_T, y_T)$ is

$$\lambda_{x_1} g(x_1, y_1) P_{x_1,x_2} g(x_2, y_2) \cdots P_{x_{T-1},x_T} g(x_T, y_T).$$

Exercise 8. The prediction is

$$\begin{aligned}
&p(x_{n+1} \mid y_{1:n}) \\
&\quad = \sum_{x_n=1}^{2} p(x_n, x_{n+1} \mid y_{1:n}) \\
&\quad = \sum_{x_n=1}^{2} p(x_{n+1} \mid x_n, y_{1:n}) p(x_n \mid y_{1:n}) \\
&\quad = \sum_{x_n=1}^{2} p(x_n \mid y_{1:n}) P_{x_n,x_{n+1}} \\
&\quad = \left(\pi_n^{\mathrm{T}} P\right)_{x_{n+1}}.
\end{aligned}$$

The update is

$$p(x_{n+1} \mid y_{1:n+1}) = \frac{p(x_{n+1} \mid y_{1:n}) g(x_{n+1}, y_{n+1})}{\sum_{x_{n+1}=1}^{2} p(x_{n+1} \mid y_{1:n}) g(x_{n+1}, y_{n+1})}.$$

The update can be written as

$$\pi_{n+1}^{\mathrm{T}} = \pi_n^{\mathrm{T}} P B_{n+1} / \left(\pi_n^{\mathrm{T}} P B_{n+1} \mathbf{1}\right)$$

when B_{n+1} is the diagonal matrix

$$B_{n+1} = \begin{bmatrix} g(1, y_{n+1}) & 0 \\ 0 & g(2, y_{n+1}) \end{bmatrix}.$$

Exercise 9. Let $\beta_n(x_n) = p(y_{n+1}, \ldots, y_T \mid x_n)$, for $n \leq T-1$, finding $\beta_{n-1}(x_{n-1})$:

$$\begin{aligned} p(y_n, \ldots, y_T \mid x_{n-1}) &= \sum_{x_n} p(x_n, y_n, \ldots, y_T \mid x_{n-1}) \\ &= \sum_{x_n} p(y_n, \ldots, y_T \mid x_{n-1}, x_n) p(x_n \mid x_{n-1}) \\ &= \sum_{x_n} p(y_{n+1}, \ldots, y_T \mid x_{n-1}, x_n, y_n) p(y_n \mid x_{n-1}, x_n) p(x_n \mid x_{n-1}) \\ &= \sum_{x_n} p(y_{n+1}, \ldots, y_T \mid x_n) p(y_n \mid x_n) p(x_n \mid x_{n-1}) \\ &= \sum_{x_n} \beta_n(x_n) g(x_n, y_n) P_{x_{n-1}, x_n}. \end{aligned}$$

$$\beta_{n-1}(x_{n-1}) = \left(P B_n \beta_n\right)_{x_{n-1}}$$

where $\beta_n = [\beta_n(1), \beta_n(2)]^{\mathrm{T}}$. Define $\beta_T = [1, 1]^{\mathrm{T}}$.

Exercise 10. The smoother is

$$\begin{aligned}
p(x_n \mid y_{1:T}) &= \frac{p(x_n, y_{1:T})}{p(y_{1:T})} \\
&= \frac{p(y_{n+1:T} \mid x_n, y_{1:n}) p(x_n \mid y_{1:n}) p(y_{1:n})}{p(y_{1:T})} \\
&= \frac{p(y_{n+1:T} \mid x_n) p(x_n \mid y_{1:n}) p(y_{1:n})}{p(y_{n+1:T} \mid y_{1:n}) p(y_{1:n})} \\
&= \frac{\beta_n(x_n)\pi_n(x_n)}{\beta_n^{\mathrm{T}}\pi_n}.
\end{aligned}$$

Note that the definition of $\beta_T = [1,1]^{\mathrm{T}}$ensures $p(x_T \mid y_{1:T}) = \beta_T(x_T)\pi_T(x_T)/\beta_T^{\mathrm{T}}\pi_T$.

Exercise 11. Let X_k be a Markov chain with values in the finite set $\{1, \ldots, n\}$ and let the probability mass function of X_1 be $\lambda = (\lambda_1, \ldots, \lambda_n)^{\mathrm{T}}$ where λ_i is the probability $X_1 = i$. The transition probability matrix is an $n \times n$ matrix P with elements $P_{i,j}$.

The process Y_k takes values in the finite set $\{1, \ldots, m\}$. Condition on $X_k = x_k$, the probability mass function of Y_k is

$$g(x_k, 1), \ldots, g(x_k, m).$$

This completes the description of the hidden Markov model .

The derivation of the filter and smoother is unchanged for this more general finite state and finite observation valued hidden Markov model when the diagonal matrix B_k is

$$B_k = \begin{bmatrix} g(1, y_k) & & & \\ & g(2, y_k) & & \\ & & \ddots & \\ & & & g(n, y_k) \end{bmatrix}.$$

S.S. Singh, Department of Engineering, University of Cambridge, Cambridge, CB1 7AT, UK

Email address: sss40@cam.ac.uk