Homework #4

Chapter 6, Shumway and Stoffer

These are outlines of the solutions. If you would like to fill in other details, please come see me during office hours.

6.1 State-space representation of AR(1) plus noise

(a) The equations are almost in state-space form as given; you just have to watch the time lag in the state equation. As noted in class, there are many ways to represent the process as a state-space model. This form is probably the simplest:

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & -0.9 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix} .$$

The observation equation is then $y_t = (1,0)x_t + v_t$. (That is, $H' = (1,0)$.)

(b) Since the elements of the state vector are observations of the AR(2) process, set $\sigma_0^2 = \sigma_1^2 = \text{Var}(x_t)$. Since the process is stationary with just one coefficient, we can solve this one as in an AR(1) model, obtaining $\text{Var}(x_t) = \sigma_w^2/(1-0.9^2)$.

(c) and (d) The simulation is as done in the class notes for the Kalman filter: generate the specified AR(2) model and add independent noise. The PACF shows much less of the cut-off characteristic "AR(2) signature" when hidden in more noise. The higher noise level obscures the dependence when most of the variance is associated with the white noise errors.

6.2 Innovations in AR(1) plus noise

The innovations are uncorrelated; these are the "residuals" of projecting Y_t on the prior observations $Y_{t-1}, Y_{t-2}, \ldots$. Notice that for any random variables X and Y that

$$\mathbb{E}\left(X(Y - \mathbb{E}\left(Y|X\right))\right) = \mathbb{E}_x \mathbb{E}_{y|x}(Y - \mathbb{E}\left(Y|X\right)) = \mathbb{E}_x X \mathbb{E}\left(Y - \mathbb{E}\left(Y|X\right)\right) = 0$$

Hence, for any $s < t$, with $Y = Y_t$ and $X = (Y_1, \ldots, Y_{t-1})$ and $\epsilon_t = y_t - \mathbb{E}\left(y_t|y_1, \ldots, y_{t-1}\right)$ that $\mathbb{E}\left(\epsilon_s \epsilon_t\right) = 0$. If $s = t$, then the variance of the innovations is the term seen in expressions for the gain:

$$\text{Var}(\epsilon_t) = \text{Var}(H\tilde{x}_t + v_t) = HP_{t|t-1}H' + \sigma_v^2 = P_{t|t-1} + \sigma_v^2$$

since $H = 1$ in that example.

6.3 Simulation of AR(1) plus noise, with error bands

This is a nice style for plots that show a sequence of estimates of the unknown state.

6.5 Projection theorem derivation of Kalman smoother

The exercise uses H_t for what we called the gain K_t. I will use the notation from class in this answer, and call their H by A. The space $\mathcal{L}_k = \overline{\text{sp}}\{y_1, \ldots, y_k\} = \overline{\text{sp}}\{\tilde{y}_1, \ldots, \tilde{y}_k\}$ where $\tilde{y}_t$ are the innovations. I've used the following expressions for the state and observation equations, as in the class notes (Lecture 15):

$$\text{State:} \quad x_t = F\,x_{t-1} + v_t \tag{1}$$

$$\text{Observation:} \quad y_t = H\,x_t + w_t \tag{2}$$

I'll also drop the t subscripts from F and H.

(a) The matrix A_{k+1} is the regression coefficient of x_k on the innovation $\tilde{y}_{k+1} = y_{k+1} - \hat{y}_{k+1|k}$,

$$\begin{aligned}
A_{k+1} &= \text{Cov}(x_k, \tilde{y}_{k+1})\,\text{Var}(\tilde{y}_{k+1})^{-1} \\
&= \text{Cov}(x_k, w_{k+1} + Hx_{k+1} - H\hat{x}_{k+1|k})\,\text{Var}(\tilde{y}_{k+1})^{-1} \\
&= \text{Cov}(x_k, H(Fx_k + v_{k+1}) - HF\hat{x}_{k|k})\,\text{Var}(\tilde{y}_{k+1})^{-1} \\
&= \text{Cov}(\tilde{x}_k + \hat{x}_{k|k}, H(Fx_k + v_{k+1}) - HF\hat{x}_{k|k})\,\text{Var}(\tilde{y}_{k+1})^{-1} \\
&= P_{k|k}F'H'(HP_{k+1|k}H' + R)^{-1}
\end{aligned}$$

where $\tilde{x}_k = x_k - \hat{x}_{k|k}$ which is hence orthogonal to the innovations $\tilde{y}_j$, $j = k, k-1, \ldots$

(b) Equating the two expressions given in the exercise means that we must show that $A_{k+1} = JK_{k+1}$. Postmultipling both by $\text{Var}(\tilde{y}_{k+1})$ leaves the expression $Pk|kF'H' = JP_{k+1|k}H'$. For this to hold for all H implies that

$$J = P_{k|k}F'P^{-1}_{k+1|k}\,.$$

(c) As in part (a) we again need a regression coefficient. Solving for A_{k+2} follows the same script as in (a); A_{k+2} is the regression coefficient of x_k on the innovation $y_{k+2} - \hat{y}_{k+2|k+1}$. It is useful to recognize that the gain K_t times the innovation gives the update to the estimate of the state,

$$K_t(y_t - \hat{y}_{t|t-1}) = \hat{x}_{t|t} - \hat{x}_{t|t-1}$$

Patiently back-substiting from the definition of the filter and innovations gives

$$\begin{aligned}
A_{k+2} &= \text{Cov}(x_k, \tilde{y}_{k+2})\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, w_{k+2} + H(x_{k+2} - \hat{x}_{k+2|k+1})\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, Fx_{k+1} + v_{k+2} - F\hat{x}_{k+1|k+1})H'\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, x_{k+1} - \hat{x}_{k+1|k} - K_{k+1}\tilde{y}_{k+1})F'H'\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, x_{k+1} - \hat{x}_{k+1|k} - K_{k+1}(w_{k+1} + H(x_{k+1} - \hat{x}_{k+1|k}))F'H'\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, (I - K_{k+1}H)(x_{k+1} - \hat{x}_{k+1|k}))F'H'\,\text{Var}(\tilde{y}_{k+2})^{-1} \\
&= \text{Cov}(x_k, v_{k+1} + F(x_k - \hat{x}_{k|k}))(I - KH)'F'H'\,\text{Var}(\tilde{y}_{k+2})^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{Cov}(\tilde{x}_{k|k}, \tilde{x}_{k|k}) F'(I - KH)' F' H' \operatorname{Var}(\tilde{y}_{k+2})^{-1} \\
&= P_{k|k} F'(I - KH)' F' H' \operatorname{Var}(\tilde{y}_{k+2})^{-1}
\end{aligned}
$$

The two expressions in the exercise are the same if

$$
J(\hat{x}_{k+1|k+2} - \hat{x}_{k+1|k}) - J(\hat{x}_{k+1|k+1} - \hat{x}_{k+1|k}) = J(\hat{x}_{k+1|k+2} - \hat{x}_{k+1|k+1}) = A_{k+2} \tilde{y}_{k+2}
$$

This must hold for all innovations $\tilde{y}_{k+2}$, implying that

$$
J P_{k+1|k+1} F' H' \operatorname{Var}(\tilde{y}_{k+2})^{-1} = A_{k+2}
$$

and hence that

$$
J P_{k+1|k+1} F' H' = P_{k|k} F'(I - KH)' F' H'
$$

Again arguing that this must hold for all H and F we get

$$
J P_{k+1|k+1} = P_{k|k} F'(I - KH)' \, .
$$

Now substitute for J from part (b), and we need to show that

$$
P_{k|k} F' (P_{k+1|k})^{-1} P_{k+1|k+1} = P_{k|k} F'(I - KH)'
$$

Multiply by $P_{k|k}^{-1}$ and again drop the common F gives

$$
(P_{k+1|k})^{-1} P_{k+1|k+1} = (I - KH)'
$$

Now multipy by $P_{k+1|k}$ to obtain

$$
P_{k+1|k+1} = P_{k+1|k}(I - KH)'
$$

Since the covariance matrices are symmetric, it holds that

$$
P_{k+1|k+1} = (I - KH) P_{k+1|k}
$$

This is the update equation for $P_{k|k}$.

(d) To prove the claim by induction, the initial "$n = 1$" statement is shown in part (a). The induction step requires that one mimic the argument used above. That is, given that

$$
\hat{x}_{k|k+m} = \hat{x}_{k|k} + J_k(\hat{x}_{k+1|k+m} - \hat{x}_{k+1|k})
$$

then show that

$$
\hat{x}_{k|k+m+1} = \hat{x}_{k|k} + J_k(\hat{x}_{k+1|k+m+1} - \hat{x}_{k+1|k})
$$

Part (c) shows this in the special case with $m = 1$. The general case works similarly. Just keep careful track of the indices.

6.6 Random walk plus noise fit to glacial varve

(a) To show that y_t is IMA(1,1), show that the differences $z_t = y_t - y_{t-1}$ have the covariances of a first order moving average. Direct substitution shows that $z_t = w_t + (v_t - v_{t-1})$ so that $\gamma_z(0) = \text{Var} z_t = \sigma_w^2 + 2\sigma_v^2$, $\gamma_z(1) = -\sigma_v^2$, and $\gamma_z(h) = 0$ for $h = 2, 3, \ldots$. The resulting first correlation is

$$\rho(1) = -\sigma_v^2/(\sigma_w^2 + 2\sigma_v^2) < \tfrac{1}{2} \; .$$

(b) Compared to the results in Example 3.31 (or 3.32 in 3rd ed), you can fit an IMA using the ARIMA estimation routine. To use the Kalman filter for the estimation, follow a script like those we used in the examples with R in class. Roughly the estimates should be similar to $\sigma_w \approx 0.11$ and $\hat\sigma_v \approx 0.425$. Together these give $\hat\rho(1) \approx -0.48$ and $\hat\theta \approx -0.77$. Here's some R code you can use...

```
 y <- log(varve)
 n <- length(y)
 mu0 <- y[1];  sigma0 <- var(y[1:10]) # rough starting values
 # function to pass to optimizer
 like <- function(para){
       cQ <- para[1]; cR <- para[2];
       kf <- Kfilter0(n, y, 1, mu0, sigma0, 1, cQ, cR)
       kf$like }
init.par <- c(.1,.1)
est <- optim(init.par, like, NULL, method=''BFGS'', hessian=TRUE,
                control=list(trace=1,REPORT=1)))
se <- sqrt(diag(solve(est$hessian)))
# Summary of estimation
c(sig.w=est$par[1], sig.v=est$par[2])
```

6.13 Missing data

The role of the observation equation in this example is solely to handle the missing data. The example also requires that the observation matrix H and variance matrix R have time subscripts (H_t, R_t). (It's a little odd discussing x_0^n since we set $x_0 = 0$, but it does show you how the smoothing filter reverses the data. It would have made more sense I think to set $A_1 = 0$ and go from there as if the first observation were missing.)

Perhaps the "easiest" way to answer this question is to argue that the KF is just a way to compute estimates $\mathbb{E}\,(x_t|y_t)$ from processes with a given covariance formula, in this case AR(1).

From this point of view, the best prediction of x_0 uses the reverse regression, obtaining $\phi x_1 = \phi y_1$ with error variance σ_w^2. For the interpolation problem, the Markovian nature of the process gives

$$\mathbb{E}\,(x_m | y_1, \ldots, y_{m-1}) = \phi y_{m-1}$$

and

$$\mathbb{E}\,(x_m | y_{m+1}, \ldots, y_n) = \phi y_{m+1}$$

Since the spaces are not orthogonal, however, we cannot partition the expected value $\mathbb{E}\,(x_m | y_{t \neq m})$ as the sum of these. So, we have to do the projection simultaneously,

$$\mathbb{E}\,(x_m | y_{t \neq m}) = \mathbb{E}\,(x_m | y_{m-1}, y_{m+1})$$

This is regression with coefficients (see the regression summary in 6.4)

$$\begin{pmatrix} \gamma_0 & \gamma_2 \\ \gamma_2 & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \phi \\ \phi \end{pmatrix} = \begin{pmatrix} \phi/(1+\phi^2) \\ \phi/(1+\phi^2) \end{pmatrix} .$$

You need to remember how to invert a 2x2 matrix for this!

To get the variance, use the regression expression for the conditional variance of one normal r.v. given another. Some remarkable cancellation happens:

$$\begin{aligned} \mathrm{Var}(Y|X) &= \mathrm{Var}(Y) - \beta' \,\mathrm{Cov}(X,Y) \\ &= \sigma_w^2 \left(\frac{1}{1-\phi^2} - \frac{\phi}{1+\phi^2} \frac{2\phi}{1-\phi^2} \right) \\ &= \frac{\sigma_w^2}{1+\phi^2} \,. \end{aligned}$$