Exercise 1 (Solution to exercise 2)

(a). We have

$$\mathbf{v}^T\mathbf{Z} = \mathbf{v}^T\mathbf{Y} + \mathbf{v}^T\boldsymbol{\eta}$$

Here $\mathbf{v}^T\mathbf{Y}$ is normal since $\mathbf{Y}$ is multivariate normal (from exercise 1). Further, $\mathbf{v}^T\boldsymbol{\eta}$ is a linear combination of independent normals, so is normal. Finally, from the model assumptions, $\mathbf{v}^T\mathbf{Y}$ and $\mathbf{v}^T\boldsymbol{\eta}$ are indpendent giving that $\mathbf{v}^T\mathbf{Z}$ is a linear combination of independent normals, which then is normal itself. Since this is true for all $\mathbf{v}$, we have that $\mathbf{Z}$ is normal.

(b). We have that

$$E[Z_t] = E[E[Z_t|Y_t]] = E[Y_t] = 0$$

where the last equality is from exercise 1 using that $\mu = 0$ in this case.

(c). We have that

$$\begin{aligned}
\text{cov}(Z_t, Z_{t+\tau}) =& E[\text{cov}(Z_t, Z_{t+\tau})|Y_t, Y_{t+\tau}] + \text{cov}(E[Z_t|Y_t, Y_{t+\tau}], E[Z_{t+\tau}|Y_t, Y_{t+\tau}]) \\
=& E[\text{cov}(\eta_t, \eta_{t+\tau})|Y_t, Y_{t+\tau}] + \text{cov}(Y_t, Y_{t+\tau}) \\
=& \tau^2 I(\tau = 0) + \text{cov}(Y_t, Y_{t+\tau})
\end{aligned}$$

where the first term is obtained by the properties of the η_t's.

(d). We have that

$$f(\mathbf{z}; \boldsymbol{\theta}) = f(z_1; \boldsymbol{\theta}) \prod_{t=2}^{n} f(z_t|z_1, ..., z_{t-1}; \boldsymbol{\theta})$$

Now $\mathbf{Z}$ is multivariate Gaussian. This imply that any subset of $\mathbf{Z}$ is (multivariate) Gaussian. That again imply that Z_1 is Gaussian. Further (Z_1, Z_2) is Gaussian which imply that $Z_2|Z_1$ is Gaussian and similarly $Z_t|Z_1, ..., Z_{t-1}$ is Gaussian. The result then follows.

(e). We have that

$$\hat{z}_t \equiv E[Z_t|Z_1, ..., Z_{t-1}] = E[E[Z_t|Z_1, ..., Z_{t-1}, Y_t]] = E[Y_t|Z_1, ..., Z_{t-1}] \equiv \hat{y}_{t|t-1}$$

Further,

$$\begin{aligned}
S_T \equiv& \text{var}(Z_t|Z_1, ..., Z_{t-1}) \\
=& \text{var}(E[Z_t|Z_1, ..., Z_{t-1}, Y_t]) + E[\text{var}(Z_t|Z_1, ..., Z_{t-1}, Y_t)] \\
=& \text{var}(Y_t|Z_1, ..., Z_{t-1}) + E[\tau^2|Z_1, ..., Z_{t-1}] \\
\equiv& P_{t|t-1} + \tau^2
\end{aligned}$$

(f). With a similar argument as in (a) we also have that the combined vector $(\mathbf{Y}, \mathbf{Z})$ is multivariate normal and therefore the conditional distribution in question is normal. The conditional means and variances are given directly from the definition. We further have that

$$\begin{aligned}\text{cov}(Y_t, Z_t|Z_1, ..., Z_{t-1}) =& \text{cov}(E[Y_t|Z_1, ..., Z_{t-1}, Y_t], E[Z_t|Z_1, ..., Z_{t-1}, Y_t]) + \\ & E[\text{cov}(Y_t, Z_t|Z_1, ..., Z_{t-1}, Y_t) \\ =& \text{cov}(Y_t, Y_t|Z_1, ..., Z_{t-1}) + 0 = P_{t|t-1}\end{aligned}$$

where the zero is obtained by that given Y_t Y_t itself is a constant and have zero variance. Based on this we have (using the rules for conditional normals)

$$\begin{aligned}\hat{y}_{t|t} \equiv& E(Y_t|Z_1, ..., Z_{t-1}, Z_t) \\ =& E[Y_t|Z_1, ..., Z_{t-1}] + \frac{\text{cov}[Y_t, Z_t|Z_1, ..., Z_{t-1}]}{\text{var}[Z_t|Z_1, ..., Z_{t-1}]}(Z_t - E[Z_t|Z_1, ..., Z_{t-1}]) \\ =& \hat{y}_{t|t-1} + \frac{P_{t|t-1}}{S_t}(Z_t - \hat{z}_t) = \hat{y}_{t|t-1} + K_t(Z_t - \hat{z}_t) \\ P_{t|t} =& \text{var}(Y_t|Z_1, ..., Z_t) \\ =& \text{var}(Y_t|Z_1, ..., Z_{t-1}) - \frac{[\text{cov}[Y_t, Z_t|Z_1, ..., Z_{t-1})]^2}{\text{var}[Z_t|Z_1, ..., Z_{t-1}]} \\ =& P_{t|t-1} - \frac{P_{t|t-1}^2}{S_t} = P_{t|t-1}[1 - K_t]\end{aligned}$$

(g). We have that

$$\begin{aligned}\hat{y}_{t+1|t} =& E(Y_{t+1}|Z_1, ..., Z_t) \\ =& E[E(Y_{t+1}|Z_1, ..., Z_t, Y_t)] \\ =& E[\alpha Y_t|Z_1, ..., Z_t)] \equiv \alpha\hat{y}_{t|t} \\ P_{t+1|t} =& \text{var}(Y_{t+1}|Z_1, ..., Z_t) \\ =& E[\text{var}(Y_{t+1}|Z_1, ..., Z_t, Y_t) + \text{var}(E[Y_{t+1}|Z_1, ..., Z_t, Y_t]) \\ =& E[\sigma^2] + \text{var}(\alpha Y_t|Z_1, ..., Z_t) \\ =& \sigma^2 + \alpha^2 P_{t|t}\end{aligned}$$

(h). If Z_t is missing, then conditioning on $Z_1, ..., Z_{t-1}$ is equivalent to conditioning on $Z_1, ..., Z_t$ since Z_t contain no information. Thereby the result

(i). From the three equations we can calculate $\hat{z}_t$ and S_t recursively and thereby all the quantities involved for calculating the likelihood is available. Further, at each timestep we can multiply the previous value of the likelihood by $\phi(z_t; \hat{z}_t, S_t)$ making also the calculation of the likelihood recursively.

(j). The predictions seems to follow the observations but are closer to zero due to that the model has a prior prediction of zero on the process. When missing observations, the variance increases.

(k). The maximum value is about 0.895, not far from the true value. Note the smooth and unimodal behaviour of the likelihood function.

(l). The estimate of the observation error is far too small. This is reflected in that the predictions now follow the observations much closer. Further, the effect of missing observations is now much clearer in a larger increase in variance.