1. Suppose that we have collected data $\boldsymbol{y}_k$ at times $k = 1, 2, \ldots, T$ and that the measurement model in (1)–(5) in handout `Kalman` holds. Then, smoothing requires the computation of the following posterior predictive pdf (which is, as expected, Gaussian):

$$f(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:T}) = \mathcal{N}\Big(\boldsymbol{\beta}_k \,\Big|\, \widehat{\boldsymbol{\beta}}(k \,|\, T), P(k \,|\, T)\Big).$$

Show that we can obtain $\widehat{\boldsymbol{\beta}}(k \,|\, T)$ and $P(k \,|\, T)$ *recursively* as follows (for $k = T-1, T-2, \ldots$):

$$\widehat{\boldsymbol{\beta}}(k \,|\, T) = \widehat{\boldsymbol{\beta}}(k \,|\, k) + G_k \, [\widehat{\boldsymbol{\beta}}(k+1 \,|\, T) - H \, \widehat{\boldsymbol{\beta}}(k \,|\, k)] \quad (1)$$

$$P(k \,|\, T) = P(k \,|\, k) - G_k \, [P(k+1 \,|\, k) - P(k+1 \,|\, T)] \, G_k^T \quad (2)$$

where

$$G_k \triangleq P(k \,|\, k) \, H^T \, P(k+1 \,|\, k)^{-1} \quad (3)$$

and $P(k \,|\, k)$ and $P(k+1 \,|\, k)$ have been defined in handout `Kalman`. Observe that the above equations are recursive: to find $\widehat{\boldsymbol{\beta}}(k \,|\, T)$ and $P(k \,|\, T)$, we need $\widehat{\boldsymbol{\beta}}(k+1 \,|\, T)$ and $P(k+1 \,|\, T)$. Therefore, you may apply induction.

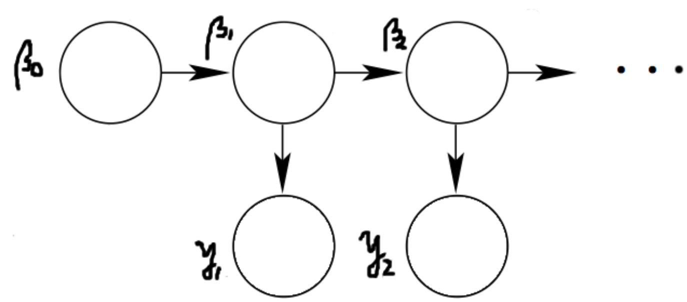
Hints:

It may help to consider the scalar case and derive scalar version of the above equations (you will lose some credit, but not much).

I suggest starting with

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:T}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:T}) = \int f_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_{k+1} \,|\, \boldsymbol{y}_{1:T}}(\boldsymbol{\beta}_k, \boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:T}) \, d\boldsymbol{\beta}$$

and continuing similarly to our Kalman-filter derivation. Clearly, you need to use conditional-independence properties that follow from the HMM graph:



To obtain results (1)–(3), you may need to use the matrix-inversion lemma

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

and the identity

$$(A + BCD)^{-1}BC = A^{-1}B(C^{-1} + DA^{-1}B)^{-1}.$$

see handout `Kalman`. (Again, I suggest trying out the scalar case if you have trouble with matrices.) You will also need the "useful fact" in handout `Kalman` — make sure that you determine the marginal and conditional pdfs properly. Finally, here is my (not carefully checked) version of the "raw" smoothing equations (prior to being polished by the matrix-inversion identities):

$$\widehat{\boldsymbol{\beta}}(k\,|\,T) = [P(k\,|\,k)^{-1} + H^T(JQJ^T)^{-1}H]^{-1}\,[P(k\,|\,k)^{-1}\widehat{\boldsymbol{\beta}}(k\,|\,k) + H^T(JQJ^T)^{-1}\widehat{\boldsymbol{\beta}}(k+1\,|\,T)]$$

$$P(k\,|\,T) = [P(k\,|\,k)^{-1} + H^T(JQJ^T)^{-1}H]^{-1}\,H^T(JQJ^T)^{-1}P(k+1\,|\,T)(JQJ^T)^{-1}H\,[P(k\,|\,k)^{-1} + H^T(JQJ^T)^{-1}H]^{-1} + [P(k\,|\,k)^{-1} + H^T(JQJ^T)^{-1}H]^{-1}.$$

2. We now drop the Gaussian assumption and show optimality properties of the Kalman filter in the linear minimum mean-square error (LMMSE) sense. Show that the Kalman filtering estimate $\widehat{\boldsymbol{\beta}}(k\,|\,k)$ and its error $\boldsymbol{\beta}(k) - \widehat{\boldsymbol{\beta}}(k\,|\,k)$ are orthogonal, i.e.

$$\mathrm{E}_{\boldsymbol{\beta}(k),\boldsymbol{Y}_{1:k}}\{\widehat{\boldsymbol{\beta}}(k\,|\,k)\,[\boldsymbol{\beta}(k) - \widehat{\boldsymbol{\beta}}(k\,|\,k)]^T\} = 0.$$