Clustering Dynamic Textures with the Hierarchical EM Algorithm for Modeling Video

Adeel Mumtaz, Emanuele Coviello, Gert. R. G. Lanckriet, Antoni B. Chan

APPENDIX A
SENSITIVITY ANALYSIS OF THE KALMAN SMOOTHING FILTER

In this section, we derive a novel efficient algorithm for performing *sensitivity analysis* of the Kalman smoothing filter, which is used to compute the E-step expectations. We begin with a summary of the Kalman smoothing filter, followed by the derivation of sensitivity analysis of the Kalman filter and Kalman smoothing filter. Finally, we derive an efficient algorithm for computing the expected log-likelihood using these results.

A.1 Kalman smoothing filter

The Kalman filter [36, 26] computes the mean and covariance of the state x_t of an LDS, conditioned on the partially observed sequence $y_{1:t-1} = \{y_1, \ldots, y_{t-1}\}$,

$$\tilde{x}_{t|t-1} = \mathbb{E}_{x|y_{1:t-1}}[x_t], \quad \tilde{V}_{t|t-1} = \mathrm{cov}_{x|y_{1:t-1}}(x_t), \tag{42}$$

while the Kalman smoothing filter estimates the state conditioned on the fully observed sequence $y_{1:\tau}$,

$$\begin{aligned} \tilde{x}_{t|\tau} &= \mathbb{E}_{x_t|y_{1:\tau}}[x_t], \\ \tilde{V}_{t|\tau} &= \mathrm{cov}_{x_t|y_{1:\tau}}(x_t), \\ \tilde{V}_{t,t-1|\tau} &= \mathrm{cov}_{x_{t-1,t}|y_{1:\tau}}(x_t, x_{t-1}). \end{aligned} \tag{43}$$

Both filters are summarized in Algorithm 2. The Kalman filter consists of a set of forward recursive equations (Alg. 2, line 4), while the Kalman smoothing filter contains an additional backward recursion (Alg. 2, line 8).

For sensitivity analysis, it will be convenient to rewrite the state estimators in (50) and (56) as functions only of $\tilde{x}_{t|t-1}$ and $\tilde{x}_{t|\tau}$,

$$\tilde{x}_{t|t-1} = F_{t-1}\tilde{x}_{t-1|t-2} + G_{t-1}(y_{t-1} - \bar{y}), \tag{44}$$
$$\tilde{x}_{t-1|\tau} = H_{t-1}\tilde{x}_{t|t-1} + J_{t-1}\tilde{x}_{t|\tau}, \tag{45}$$

where $\{F_t, G_t, J_t, H_t\}$ are defined in (52), (53).

Finally, note that the conditional covariances, $\tilde{V}_{t|\tau}$ and $\tilde{V}_{t,t-1|\tau}$ in (54) and (55), and matrices $\{F_t, G_t, J_t, H_t\}$, are not functions of the observed sequence $y_{1:\tau}$. Hence, we have

$$\begin{aligned} \hat{V}_t &= \mathbb{E}_{y|\Theta_b}[\tilde{V}^{(r)}_{t|\tau}] = \tilde{V}^{(r)}_{t|\tau}, \\ \hat{V}_{t,t-1} &= \mathbb{E}_{y|\Theta_b}[\tilde{V}^{(r)}_{t,t-1|\tau}] = \tilde{V}^{(r)}_{t,t-1|\tau}. \end{aligned} \tag{46}$$

- *A. Mumtaz and A. B. Chan are with the Department of Computer Science, City University of Hong Kong.*
  E-mail: adeelmumtaz@gmail.com, abchan@cityu.edu.hk.
- *E. Coviello and G. R. G. Lanckriet are with the Department of Electrical and Computer Engineering, University of California, San Diego.*
  E-mail: emanuetre@gmail.com, gert@ece.ucsd.edu.

Algorithm 2 Kalman filter and Kalman smoothing filter

1: **Input**: DT parameters $\Theta = \{A, Q, C, R, \mu, S, \bar{y}\}$, video $y_{1:\tau}$.
2: Initialize: $\tilde{x}_{1|0} = \mu,\ \tilde{V}_{1|0} = S$.
3: **for** $t = \{1, \cdots, \tau\}$ **do**
4: Kalman filter – forward recursion:

$$\tilde{V}_{t|t-1} = A\tilde{V}_{t-1|t-1}A^T + Q, \tag{47}$$
$$K_t = \tilde{V}_{t|t-1}C^T(C\tilde{V}_{t|t-1}C^T + R)^{-1}, \tag{48}$$
$$\tilde{V}_{t|t} = (I - K_t C)\tilde{V}_{t|t-1}, \tag{49}$$
$$\tilde{x}_{t|t-1} = A\tilde{x}_{t-1|t-1}, \tag{50}$$
$$\tilde{x}_{t|t} = \tilde{x}_{t|t-1} + K_t(y_t - C\tilde{x}_{t|t-1} - \bar{y}), \tag{51}$$
$$G_t = AK_t, \quad F_t = A - AK_tC. \tag{52}$$

5: **end for**
6: Initialize: $\tilde{V}_{\tau,\tau-1|\tau} = (I - K_\tau C)A\tilde{V}_{\tau-1|\tau-1}$.
7: **for** $t = \{\tau, \cdots, 2\}$ **do**
8: Kalman smoothing filter – backward recursion:

$$J_{t-1} = \tilde{V}_{t-1|t-1}A^T\tilde{V}^{-1}_{t|t-1}, \quad H_{t-1} = A^{-1} - J_{t-1}, \tag{53}$$
$$\tilde{V}_{t-1|\tau} = \tilde{V}_{t-1|t-1} + J_{t-1}(\tilde{V}_{t|\tau} - \tilde{V}_{t|t-1})J^T_{t-1}, \tag{54}$$
$$\tilde{V}_{t-1,t-2|\tau} = \tilde{V}_{t-1|t-1}J^T_{t-2} + J_{t-1}(\tilde{V}_{t,t-1|\tau} - A\tilde{V}_{t-1|t-1})J^T_{t-2}, \tag{55}$$
$$\tilde{x}_{t-1|\tau} = \tilde{x}_{t-1|t-1} + J_{t-1}(\tilde{x}_{t|\tau} - A\tilde{x}_{t-1|t-1}). \tag{56}$$

9: **end for**
10: **Output**: Kalman filter matrices $\{\tilde{V}_{t|t-1}, \tilde{V}_{t|\tau}, \tilde{V}_{t,t-1|\tau}, G_t, F_t, H_t\}$, and state estimators $\{\tilde{x}_{t|t-1}, \tilde{x}_{t|t}, \tilde{x}_{t|\tau}\}$.

A.2 Sensitivity Analysis of the Kalman smoothing filter

We consider the two LDS, Θ_b and Θ_r, and their associated Kalman filters $\{F^{(b)}_t, G^{(b)}_t, H^{(b)}_t, \tilde{x}^{(b)}_{t|t-1}, \tilde{x}^{(b)}_{t|\tau}\}$ and $\{F^{(r)}_t, G^{(r)}_t, H^{(r)}_t, \tilde{x}^{(r)}_{t|t-1}, \tilde{x}^{(r)}_{t|\tau}\}$. The goal is to compute the mean and covariance of the Kalman smoothing filter for Θ_r, when the source distribution is Θ_b,

$$\begin{aligned} &\hat{x}_t = \mathbb{E}_{y|\Theta_b}\left[\tilde{x}^{(r)}_{t|\tau}\right], \quad \hat{\kappa}_t = \mathrm{cov}_{y|\Theta_b}(y_t, \tilde{x}^{(r)}_{t|\tau}), \\ &\hat{\chi}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}), \quad \hat{\chi}_{t,t-1} = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, \tilde{x}^{(r)}_{t-1|\tau}). \end{aligned} \tag{57}$$

To achieve this, we first analyze the forward recursion, followed by the backward recursion.

A.2.1 Forward recursion

For the forward recursion, the Kalman filters for Θ_b and Θ_r are recursively defined by (44),

$$\begin{bmatrix} \tilde{x}^{(b)}_{t|t-1} \\ \tilde{x}^{(r)}_{t|t-1} \end{bmatrix} = \begin{bmatrix} F^{(b)}_{t-1}\tilde{x}^{(b)}_{t-1|t-2} + G^{(b)}_{t-1}(y^{(b)}_{t-1} - \bar{y}_b) \\ F^{(r)}_{t-1}\tilde{x}^{(r)}_{t-1|t-2} + G^{(r)}_{t-1}(y^{(b)}_{t-1} - \bar{y}_r) \end{bmatrix},$$

where $\{y^{(b)}_t\}$ are the observations from source Θ_b. Substituting (2) of the base model, *i.e.*, $y^{(b)}_{t-1} = C_b x^{(b)}_{t-1} + w^{(b)}_{t-1} + \bar{y}_b$, and including

Algorithm 3 Sensitivity Analysis of Kalman filter

1: **Input**: DT parameters Θ_b and Θ_r, Kalman filter matrices $\{G_t^{(b)}, F_t^{(b)}, G_t^{(r)}, F_t^{(r)}\}$, length τ.
2: Initialize: $\hat{\mathbf{x}}_1 = \begin{bmatrix}\mu_b\\ \mu_b\\ \mu_r\end{bmatrix}$, $\hat{\mathbf{V}}_1 = \begin{bmatrix}S_b & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$.
3: **for** $t = \{2, \cdots, \tau+1\}$ **do**
4: Form block matrices:

$$\mathbf{A}_{t-1} = \begin{bmatrix} A_b & 0 & 0\\ G_{t-1}^{(b)}C_b & F_{t-1}^{(b)} & 0\\ G_{t-1}^{(r)}C_b & 0 & F_{t-1}^{(r)}\end{bmatrix},$$
$$\mathbf{B} = \begin{bmatrix} I\\0\\0\end{bmatrix},\ \mathbf{C}_{t-1} = \begin{bmatrix}0\\ G_{t-1}^{(b)}\\ G_{t-1}^{(r)}\end{bmatrix},\ \mathbf{D}_{t-1} = \begin{bmatrix}0\\0\\G_{t-1}^{(r)}\end{bmatrix}. \tag{59}$$

5: Update means and covariances:

$$\hat{\mathbf{x}}_t = \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{D}_{t-1}(\bar{y}_b - \bar{y}_r), \tag{60}$$
$$\hat{\mathbf{V}}_t = \mathbf{A}_{t-1}\hat{\mathbf{V}}_{t-1}\mathbf{A}_{t-1}^T + \mathbf{B}Q_b\mathbf{B}^T + \mathbf{C}_{t-1}R_b\mathbf{C}_{t-1}^T. \tag{61}$$

6: **end for**
7: **Output**: $\hat{\mathbf{x}}_t$, $\hat{\mathbf{V}}_t$.

the recursion of the associated state-space $x_{t-1}^{(b)}$ given in (1), we have

$$\begin{bmatrix} x_t^{(b)}\\ \tilde{x}_{t|t-1}^{(b)}\\ \tilde{x}_{t|t-1}^{(r)}\end{bmatrix} = \begin{bmatrix} A_b x_{t-1}^{(b)} + v_t^{(b)}\\ F_{t-1}^{(b)}\tilde{x}_{t-1|t-2}^{(b)} + G_{t-1}^{(b)}(C_b x_{t-1}^{(b)} + w_{t-1}^{(b)})\\ F_{t-1}^{(r)}\tilde{x}_{t-1|t-2}^{(r)} + G_{t-1}^{(r)}(C_b x_{t-1}^{(b)} + w_{t-1}^{(b)} + \bar{y}_b - \bar{y}_r)\end{bmatrix}$$

which can be rewritten succinctly as

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}v_t^{(b)} + \mathbf{C}_{t-1}w_{t-1}^{(b)} + \mathbf{D}_{t-1}(\bar{y}_b - \bar{y}_r), \tag{58}$$

where $\mathbf{x}_t = [(x_t^{(b)})^T, (\tilde{x}_{t|t-1}^{(b)})^T, (\tilde{x}_{t|t-1}^{(r)})^T]^T$, and the block matrices $\{\mathbf{A}_{t-1}, \mathbf{B}, \mathbf{C}_{t-1}, \mathbf{D}_{t-1}\}$ are defined in (59).

Finally, taking the expectation of (58), with respect to $\{x_{1:\tau}, y_{1:\tau}\} \sim \Theta_b$, yields the recursive equations for $\hat{\mathbf{x}}_t$ in (60). Similarly, taking the covariance of (58) yields a recursive equation for $\hat{\mathbf{V}}_t$ in (61), where we have used the fact that $\{v_t^{(b)}, w_{t-1}^{(b)}\} \perp\!\!\!\perp \{x_{t-1}^{(b)}, \tilde{x}_{t-1|t-2}^{(b)}, \tilde{x}_{t-1|t-2}^{(r)}\}$, and $v_t^{(b)} \perp\!\!\!\perp w_t^{(b)}$. The recursive equations for the sensitivity analysis of the Kalman filter are summarized in Algorithm 3.

A.2.2 Backward recursion

Taking the expectation of (45) yields a recursion for $\hat{x}_t$,

$$\begin{aligned}\hat{x}_{t-1} &= \mathbb{E}_{y|\Theta_b}[\tilde{x}_{t-1|\tau}^{(r)}]\\ &= H_{t-1}^{(r)}\mathbb{E}_{y|\Theta_b}[\tilde{x}_{t|t-1}^{(r)}] + J_{t-1}^{(r)}\mathbb{E}_{y|\Theta_b}[\tilde{x}_{t|\tau}^{(r)}]\\ &= H_{t-1}^{(r)}\hat{\mathbf{x}}_t^{[3]} + J_{t-1}^{(r)}\hat{x}_t,\end{aligned} \tag{62}$$

with initial condition

$$\hat{x}_\tau = \mathbb{E}_{y|\Theta_b}[\tilde{x}_{\tau|\tau}^{(r)}] = A_r^{-1}\mathbb{E}_{y|\Theta_b}[\tilde{x}_{\tau+1|\tau}^{(r)}] = A_r^{-1}\hat{\mathbf{x}}_{\tau+1}^{[3]}. \tag{63}$$

Taking the covariance of (45), we obtain the recursion,

$$\begin{aligned}\hat{\chi}_{t-1} &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t-1|\tau}^{(r)})\\ &= \mathrm{cov}_{y|\Theta_b}(H_{t-1}\tilde{x}_{t|t-1} + J_{t-1}\tilde{x}_{t|\tau})\\ &= \begin{bmatrix}H_{t-1}^{(r)} & J_{t-1}^{(r)}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{V}}_t^{[3,3]} & \hat{\omega}_t^T\\ \hat{\omega}_t & \hat{\chi}_t\end{bmatrix}\begin{bmatrix}(H_{t-1}^{(r)})^T\\ (J_{t-1}^{(r)})^T\end{bmatrix},\end{aligned} \tag{64}$$

where $\hat{\omega}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t|t-1}^{(r)}, \tilde{x}_{t|\tau}^{(r)})$, and the initial condition is

$$\begin{aligned}\hat{\chi}_\tau &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{\tau|\tau}^{(r)}) = \mathrm{cov}_{y|\Theta_b}(A_r^{-1}\tilde{x}_{\tau+1|\tau}^{(r)})\\ &= A_r^{-1}\mathrm{cov}_{y|\Theta_b}(\tilde{x}_{\tau+1|\tau}^{(r)})A_r^{-T} = A_r^{-1}\hat{\mathbf{V}}^{[3,3]}A_r^{-T}.\end{aligned} \tag{65}$$

Substituting (45) into the definition of $\hat{\chi}_{t,t-1}$ yields

Algorithm 4 Sensitivity Analysis of Kalman smoothing filter

1: **Input**: DT parameters Θ_b and Θ_r, Kalman smoothing filter matrices $\{G_t^{(b)}, F_t^{(b)}\}$ and $\{G_t^{(r)}, F_t^{(r)}, H_t^{(r)}, J_t^{(r)}\}$, Kalman filter sensitivity analysis $\{\hat{\mathbf{x}}_t, \hat{\mathbf{V}}_t\}$, length τ.
2: Initialize: $\hat{x}_\tau = A_r^{-1}\hat{\mathbf{x}}_{\tau+1}^{[3]}$, $\hat{\chi}_\tau = A_r^{-1}\hat{\mathbf{V}}_{\tau+1}^{[3,3]}A_r^{-T}$, $L_\tau = A_r^{-1}$, $M_\tau = \mathbf{0}$.
3: **for** $t = \{\tau, \cdots, 1\}$ **do**
4: Compute cross-covariance:

$$\hat{\rho}_t = \left(L_t F_t^{(r)}\hat{\mathbf{V}}_t^{[3,2]} + (L_t G_t^{(r)}C_b + M_t)\hat{\mathbf{V}}_t^{[1,1]}\right)C_b^T + L_t G_t^{(r)}R_b. \tag{67}$$

5: **if** $t > 1$ **then**
6: Compute sensitivity:

$$\hat{\omega}_t = L_t F_t^{(r)}\hat{\mathbf{V}}_t^{[3,3]} + (L_t G_t^{(r)}C_b + M_t)\hat{\mathbf{V}}_t^{[2,3]}, \tag{68}$$
$$\hat{x}_{t-1} = H_{t-1}^{(r)}\hat{\mathbf{x}}_t^{[3]} + J_{t-1}^{(r)}\hat{x}_t, \tag{69}$$
$$\hat{\chi}_{t-1} = \begin{bmatrix}H_{t-1}^{(r)} & J_{t-1}^{(r)}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{V}}_t^{[3,3]} & \hat{\omega}_t^T\\ \hat{\omega}_t & \hat{\chi}_t\end{bmatrix}\begin{bmatrix}(H_{t-1}^{(r)})^T\\ (J_{t-1}^{(r)})^T\end{bmatrix}, \tag{70}$$
$$\hat{\chi}_{t,t-1} = \hat{\omega}_t(H_{t-1}^{(r)})^T + \hat{\chi}_t(J_{t-1}^{(r)})^T. \tag{71}$$

7: Update matrices:

$$L_{t-1} = H_{t-1}^{(r)} + J_{t-1}^{(r)}L_t F_t^{(r)}, \tag{72}$$
$$M_{t-1} = J_{t-1}^{(r)}(L_t G_t^{(r)}C_b + M_t)A_b. \tag{73}$$

8: **end if**
9: **end for**
10: **Output**: $\{\hat{x}_t, \hat{\chi}_t, \hat{\chi}_{t,t-1}, \hat{\kappa}_t = \hat{\rho}_t^T\}$.

$$\begin{aligned}\hat{\chi}_{t,t-1} &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t|\tau}^{(r)}, \tilde{x}_{t-1|\tau}^{(r)})\\ &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t|\tau}^{(r)}, H_{t-1}^{(r)}\tilde{x}_{t|t-1}^{(r)} + J_{t-1}^{(r)}\tilde{x}_{t|\tau}^{(r)})\\ &= \hat{\omega}_t(H_{t-1}^{(r)})^T + \hat{\chi}_t(J_{t-1}^{(r)})^T.\end{aligned} \tag{66}$$

Finally, the cross-covariances, $\hat{\omega}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t|t-1}^{(r)}, \tilde{x}_{t|\tau}^{(r)})$ and $\hat{\kappa}_t^T = \hat{\rho}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}_{t|\tau}^{(r)}, y_t)$, are calculated efficiently using the recursion in (67) and (68), where $\{L_t, M_t\}$ are recursively given by (72) and (73). The derivation is quite involved and appears in the Appendix B. The algorithm for sensitivity analysis of the Kalman smoothing filter is summarized in Algorithm 4.

A.3 Expected Log-Likelihood

The expected log-likelihood $\mathbb{E}_{y|\Theta_b}[\log p(y|\Theta_r)]$ can be calculated efficiently using the results from the sensitivity analysis for Kalman filters. First, the observation log-likelihood of the DT is expressed in "innovation" form

$$\begin{aligned}\log p(y|\Theta_r) &= \sum_{t=1}^{\tau}\log p(y_t|y_{1:t-1}, \Theta_r)\\ &= \sum_{t=1}^{\tau}\log\mathcal{N}(y_t|C_r\tilde{x}_{t|t-1}^{(r)} + \bar{y}_r, \Sigma_t) \qquad (74)\\ &= \sum_{t=1}^{\tau}\tfrac{-1}{2}\mathrm{tr}\left[\Sigma_t^{-1}(y_t - \bar{y}_r - C_r\tilde{x}_{t|t-1}^{(r)})(y_t - \bar{y}_r - C_r\tilde{x}_{t|t-1}^{(r)})^T\right]\\ &\quad - \tfrac{1}{2}\log|\Sigma_t| - \tfrac{m}{2}\log(2\pi), \qquad (75)\end{aligned}$$

where $\Sigma_t = C_r\tilde{V}_{t|t-1}^{(r)}C_r^T + R_r$. Taking the expectation of (75), and noting that $\tilde{V}_{t|t-1}^{(r)}$ and Σ_t are not a functions of the observations $y_{1:t-1}$,

$$\begin{aligned}\ell &= \mathbb{E}_{y|\Theta_b}[\log p(y|\Theta_r)]\\ &= \sum_{t=1}^{\tau}\tfrac{-1}{2}\mathrm{tr}\left[\Sigma_t^{-1}(\hat{U}_t - \hat{\lambda}_t C_r^T - C_r\hat{\lambda}_t^T + C_r\hat{\Lambda}_t C_r^T)\right]\\ &\quad - \frac{1}{2}\log|\Sigma_t| - \frac{m}{2}\log(2\pi),\end{aligned} \tag{76}$$

where

$$\begin{aligned}\hat{\Lambda}_t &= \mathbb{E}_{y|\Theta_b}[\tilde{x}^{(r)}_{t|t-1}(\tilde{x}^{(r)}_{t|t-1})^T] \\ &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|t-1}) + \mathbb{E}_{y|\Theta_b}[\tilde{x}^{(r)}_{t|t-1}]\mathbb{E}_{y|\Theta_b}[\tilde{x}^{(r)}_{t|t-1}]^T \\ &= \hat{\mathbf{V}}^{[3,3]}_t + \hat{\mathbf{x}}^{[3]}_t(\hat{\mathbf{x}}^{[3]}_t)^T, \end{aligned} \tag{77}$$

and

$$\begin{aligned}\hat{\lambda}_t &= \mathbb{E}_{y|\Theta_b}[(y_t - \bar{y}_r)(\tilde{x}^{(r)}_{t|t-1})^T] \\ &= \mathrm{cov}_{y|\Theta_b}(y_t - \bar{y}_r, \tilde{x}^{(r)}_{t|t-1}) + \mathbb{E}_{y|\Theta_b}[y_t - \bar{y}_r]\mathbb{E}_{y|\Theta_b}[\tilde{x}^{(r)}_{t|t-1}]^T \\ &= C_b\hat{\mathbf{V}}^{[2,3]}_t + (C_b\hat{\mathbf{x}}^{[1]}_t + \bar{y}_b - \bar{y}_r)(\hat{\mathbf{x}}^{[3]}_t)^T. \end{aligned} \tag{78}$$

Appendix B
Efficient calculation of the cross-covariance terms

In this section, we derive efficient expressions for calculating the cross-covariance terms,

$$\hat{\omega}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, \tilde{x}^{(r)}_{t|t-1}), \quad \hat{\rho}_t = \kappa_t^T = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, y_t).$$

First, we derive an expression of the Kalman smoothing filter $\tilde{x}^{(r)}_{t|\tau}$ as a function of only $\tilde{x}^{(r)}_{t|t-1}$ and observations $y_{t:\tau}$. Filling in the backward recursion of the Kalman smoothing filter in (45),

$$\begin{aligned}\tilde{x}^{(r)}_{t|\tau} &= H^{(r)}_t\tilde{x}^{(r)}_{t+1|t} + J^{(r)}_t\tilde{x}^{(r)}_{t+1|\tau} \\ &= H^{(r)}_t\tilde{x}^{(r)}_{t+1|t} + J^{(r)}_t(H^{(r)}_{t+1}\tilde{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1}\tilde{x}^{(r)}_{t+2|\tau}) \\ &= H^{(r)}_t\tilde{x}^{(r)}_{t+1|t} + J^{(r)}_t(H^{(r)}_{t+1}\tilde{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1}(\cdots \\ &\quad + J^{(r)}_{\tau-2}(H^{(r)}_{\tau-1}\tilde{x}^{(r)}_{\tau|\tau-1} + J^{(r)}_{\tau-1}\tilde{x}^{(r)}_{\tau|\tau})\cdots)) \\ &= H^{(r)}_t\tilde{x}^{(r)}_{t+1|t} + J^{(r)}_t(H^{(r)}_{t+1}\tilde{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1}(\cdots \\ &\quad + J^{(r)}_{\tau-2}(H^{(r)}_{\tau-1}\tilde{x}^{(r)}_{\tau|\tau-1} + J^{(r)}_{\tau-1}A_r^{-1}\tilde{x}^{(r)}_{\tau+1|\tau})\cdots)) \\ &= H^{(r)}_t\tilde{x}^{(r)}_{t+1|t} + \sum_{s=t+2}^{\tau}(\prod_{i=t}^{s-2}J^{(r)}_i)H^{(r)}_{s-1}\tilde{x}^{(r)}_{s|s-1} \\ &\quad + (\prod_{i=t}^{\tau-1}J^{(r)}_i)A_r^{-1}\tilde{x}^{(r)}_{\tau+1|\tau} \\ &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\tilde{x}^{(r)}_{s|s-1}, \end{aligned} \tag{79}$$

where we define $\hat{H}_t = \begin{cases} H^{(r)}_t & ,t<\tau \\ A_r^{-1} & ,t=\tau \end{cases}$, and $\mathbf{J}_{t,s} = \begin{cases} I & ,t>s \\ J^{(r)}_tJ^{(r)}_{t+1}\cdots J^{(r)}_s & ,t\le s \end{cases}$. Next, we rewrite the Kalman filter terms $\tilde{x}^{(r)}_{s|s-1}$, where $s>t$, as a function of $\tilde{x}^{(r)}_{t|t-1}$ and $\{y_t,\cdots,y_{s-1}\}$. Note that we drop the constant $\bar{y}_r$ term since it will not affect the covariance operator later. Substituting the forward recursions in (44), we have

$$\tilde{x}^{(r)}_{t+1|t} = F^{(r)}_t\tilde{x}_{t|t-1} + G^{(r)}_t y_t, \tag{80}$$

$$\begin{aligned}\tilde{x}^{(r)}_{t+2|t+1} &= F^{(r)}_{t+1}\tilde{x}_{t+1|t} + G^{(r)}_{t+1}y_{t+1} \\ &= F^{(r)}_{t+1}(F^{(r)}_t\tilde{x}_{t|t-1} + G^{(r)}_t y_t) + G^{(r)}_{t+1}y_{t+1}, \end{aligned} \tag{81}$$

$$\begin{aligned}\tilde{x}^{(r)}_{t+3|t+2} &= F^{(r)}_{t+2}\tilde{x}_{t+2|t+1} + G^{(r)}_{t+2}y_{t+2} \\ &= F^{(r)}_{t+2}(F^{(r)}_{t+1}(F^{(r)}_t\tilde{x}_{t|t-1} + G^{(r)}_t y_t) \\ &\quad + G^{(r)}_{t+1}y_{t+1}) + G^{(r)}_{t+2}y_{t+2}, \end{aligned} \tag{82}$$

or in general, for $s>t$,

$$\begin{aligned}\tilde{x}^{(r)}_{s|s-1} &= \underbrace{\mathbf{F}_{s-1,t+1}(F^{(r)}_t\tilde{x}_{t|t-1} + G^{(r)}_t y_t)}_{\alpha_s} \\ &\quad + \underbrace{\sum_{j=t+1}^{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j}_{\beta_s} = \alpha_s + \beta_s, \end{aligned} \tag{83}$$

where we define $\mathbf{F}_{s,t} = \begin{cases} I & ,s<t \\ F_sF_{s-1}\cdots F_t & ,s\ge t \end{cases}$, and the quantities α_s and β_s as above. Note that $\beta_{t+1} = \mathbf{0}$.

We now substitute $\tilde{x}^{(r)}_{s|s-1} = \alpha_s + \beta_s$ into (79). First, substituting the α_s terms into (79),

$$\begin{aligned}\hat{\alpha}_t &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\alpha_s \\ &= \underbrace{\sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+1}}_{L_t}(F^{(r)}_t\tilde{x}_{t|t-1} + G^{(r)}_t y_t), \end{aligned} \tag{84}$$

where L_t can be computed recursively,

$$\begin{aligned}L_t &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+1} \\ &= \mathbf{J}_{t,t-1}\hat{H}_t\mathbf{F}_{t,t+1} + \sum_{s=t+2}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+1} && (85) \\ &= \hat{H}_t + \sum_{s=t+2}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+1} \\ &= \hat{H}_t + J^{(r)}_t\left(\sum_{s=t+2}^{\tau+1}\mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+2}\right)F^{(r)}_{t+1} && (86) \\ &= H^{(r)}_t + J^{(r)}_tL_{t+1}F^{(r)}_{t+1}, && (87) \end{aligned}$$

where in (85) we have separated the first term of the summation $(s=t+1)$, and in (86) we have used $\mathbf{J}_{t,s-2} = J^{(r)}_t\mathbf{J}_{t+1,s-2}$ and $\mathbf{F}_{s-1,t+1} = \mathbf{F}_{s-1,t+2}F^{(r)}_{t+1}$, when $s\ge t$. The initial condition for the L_t backward recursion is

$$L_\tau = \mathbf{J}_{\tau,\tau-1}\hat{H}_\tau\mathbf{F}_{\tau,\tau+1} = \hat{H}_\tau = A_r^{-1}. \tag{88}$$

Next, we substitute the β_s terms into (79),

$$\begin{aligned}\hat{\beta}_t &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\beta_s \\ &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\sum_{j=t+1}^{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j \\ &= \sum_{s=t+2}^{\tau+1}\sum_{j=t+1}^{s-1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j \\ &= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j, \end{aligned} \tag{89}$$

where in (89) we have collected the $G^{(r)}_j y_j$ terms by switching the double summation. Finally, using $\{\hat{\alpha}_t, \hat{\beta}_t\}$, we rewrite the Kalman smoothing filter of (79) as a function of $\tilde{x}^{(r)}_{t|t-1}$ and $y_{t+1:\tau}$,

$$\begin{aligned}\tilde{x}^{(r)}_{t|\tau} &= \sum_{s=t+1}^{\tau+1}\mathbf{J}_{t,s-2}\hat{H}_{s-1}(\alpha_s + \beta_s) = \hat{\alpha}_t + \hat{\beta}_t \\ &= L_t(F^{(r)}_t\tilde{x}^{(r)}_{t|t-1} + G^{(r)}_t y_t) + \hat{\beta}_t. \end{aligned} \tag{90}$$

Note that $\hat{\beta}_t$ is a function of $\{y_{t+1}, \cdots, y_\tau\}$.

B.1 Calculating $\hat{\omega}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, \tilde{x}^{(r)}_{t|t-1})$

Next, we will derive an expression for $\hat{\omega}_t$

$$\hat{\omega}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, \tilde{x}^{(r)}_{t|t-1}) = \mathrm{cov}_{y|\Theta_b}(\hat{\alpha}_t + \hat{\beta}_t, \tilde{x}^{(r)}_{t|t-1}). \tag{91}$$

Looking at the covariance with the $\hat{\beta}_t$ term,

$$\begin{aligned}
&\mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, \tilde{x}^{(r)}_{t|t-1}) \\
&= \mathrm{cov}_{y|\Theta_b}(\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j, \tilde{x}^{(r)}_{t|t-1}) \\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j \mathrm{cov}_{y|\Theta_b}(y_j, \tilde{x}^{(r)}_{t|t-1}) \\
&= \underbrace{\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t}}_{M_t} \hat{\mathbf{V}}^{[2,3]}_t, \qquad (92)
\end{aligned}$$

where in the last line we have used (107). M_t can be computed with a backward recursion,

$$\begin{aligned}
M_t &= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t} \\
&= J^{(r)}_t \Bigg[\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1} \\
&\qquad \cdot \mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t-1}\Bigg] A_b \\
&= J^{(r)}_t \Bigg(\Bigg[\sum_{s=t+2}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+2}G^{(r)}_{t+1} C_b\Bigg] \qquad (93) \\
&+ \sum_{j=t+2}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-(t+1)}\Bigg) A_b \\
&= J^{(r)}_t (L_{t+1}G^{(r)}_{t+1}C_b + M_{t+1})A_b, \qquad (94)
\end{aligned}$$

where in (93) we have separated the first term of the summation ($j = t+1$), and in (94) we have used the definition of L_{t+1} and M_{t+1}. The initial condition is $M_\tau = \mathbf{0}$. Finally using (90), the cross-covariance is

$$\begin{aligned}
\hat{\omega}_t &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, \tilde{x}^{(r)}_{t|t-1}) \\
&= \mathrm{cov}_{y|\Theta_b}(L_t(F^{(r)}_t \tilde{x}^{(r)}_{t|t-1} + G^{(r)}_t y_t) + \hat{\beta}_t, \tilde{x}^{(r)}_{t|t-1}) \\
&= L_t F^{(r)}_t \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|t-1}) + L_t G^{(r)}_t \mathrm{cov}_{y|\Theta_b}(y_t, \tilde{x}^{(r)}_{t|t-1}) \\
&\quad + \mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, \tilde{x}^{(r)}_{t|t-1}) \\
&= L_t F^{(r)}_t \hat{\mathbf{V}}^{[3,3]}_t + L_t G^{(r)}_t C_b \hat{\mathbf{V}}^{[2,3]}_t + M_t \hat{\mathbf{V}}^{[2,3]}_t \qquad (95) \\
&= L_t F^{(r)}_t \hat{\mathbf{V}}^{[3,3]}_t + (L_t G^{(r)}_t C_b + M_t)\hat{\mathbf{V}}^{[2,3]}_t, \qquad (96)
\end{aligned}$$

where in (95) we have used (107) and (92).

B.2 Calculating $\hat{\rho}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, y_t)$

We now derive an expression for $\hat{\rho}_t$,

$$\hat{\rho}_t = \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, y_t) = \mathrm{cov}_{y|\Theta_b}(\hat{\alpha}_t + \hat{\beta}_t, y_t). \tag{97}$$

Looking at the covariance with $\hat{\beta}_t$,

$$\begin{aligned}
&\mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, y_t) \\
&= \mathrm{cov}_{y|\Theta_b}(\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j, y_t) \\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j \mathrm{cov}_{y|\Theta_b}(y_j, y_t) \\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t}\hat{\mathbf{V}}^{[1,1]}_t C_b^T \\
&= M_t \hat{\mathbf{V}}^{[1,1]}_t C_b^T. \qquad (98)
\end{aligned}$$

Finally, using (90), the cross-covariance is

$$\begin{aligned}
\hat{\rho}_t &= \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|\tau}, y_t) \\
&= \mathrm{cov}_{y|\Theta_b}(L_t(F^{(r)}_t \tilde{x}^{(r)}_{t|t-1} + G^{(r)}_t y_t) + \hat{\beta}_t, y_t) \\
&= L_t F^{(r)}_t \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(r)}_{t|t-1}, y_t) + L_t G^{(r)}_t \mathrm{cov}_{y|\Theta_b}(y_t) \\
&\quad + \mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, y_t) \\
&= L_t F^{(r)}_t \hat{\mathbf{V}}^{[3,2]}_t C_b^T + L_t G^{(r)}_t (C_b \hat{\mathbf{V}}^{[1,1]}_t C_b^T + R_b) \\
&\quad + M_t \hat{\mathbf{V}}^{[1,1]}_t C_b^T \qquad (99) \\
&= \left(L_t F^{(r)}_t \hat{\mathbf{V}}^{[3,2]}_t + (L_t G^{(r)}_t C_b + M_t)\hat{\mathbf{V}}^{[1,1]}_t\right) C_b^T \\
&\quad + L_t G^{(r)}_t R_b, \qquad (100)
\end{aligned}$$

where (99) follows by using (107), (98), and $\mathrm{cov}_{y|\Theta_b}(y_t) = C_b \mathrm{cov}_{x|\Theta_b}(x_t) C_b^T + R_b$.

B.3 Useful properties

In this section, we derive some properties used in the previous section. Note that in the sequel we remove the mean terms, $\bar{y}_b$ and $\bar{y}_r$, which do not affect the covariance operator. First, we derive the covariance between two observations, for $k > 0$,

$$\begin{aligned}
&\mathrm{cov}_{y|\Theta_b}(y^{(b)}_{t+k}, y^{(b)}_t) \\
&= \mathrm{cov}(C_b x^{(b)}_{t+k} + w^{(b)}_{t+k}, C_b x^{(b)}_t + w^{(b)}_t) \\
&= C_b \mathrm{cov}(x^{(b)}_{t+k}, x^{(b)}_t) C_b^T + \mathrm{cov}_{\Theta_b}(w^{(b)}_{t+k}, w^{(b)}_t) \qquad (101) \\
&= C_b \mathrm{cov}(A_b^k x^{(b)}_t + \sum_{l=1}^{k} A_b^{k-l} v^{(b)}_{t+l}, x^{(b)}_t) C_b^T \qquad (102) \\
&= C_b A_b^k \mathrm{cov}(x^{(b)}_t) C_b^T = C_b A_b^k \hat{\mathbf{V}}^{[1,1]}_t C_b^T, \qquad (103)
\end{aligned}$$

where in (102) we have rewritten $x^{(b)}_{t+k}$ as a function of $x^{(b)}_t$, *i.e.*, $x^{(b)}_{t+k} = A_b^k x^{(b)}_t + \sum_{l=1}^{k} A_b^{k-l} v^{(b)}_{t+l}$, in (101) we have used $x^{(b)}_t \perp\!\!\!\perp w^{(b)}_{t+k}$ and $x^{(b)}_{t+k} \perp\!\!\!\perp w^{(b)}_t$ for $k > 0$, and in (103), $x^{(b)}_t \perp\!\!\!\perp v^{(b)}_{t+l}$ for $l \geq 1$. Next, we derive the covariance between the one-step ahead state estimator and an observation, for $k \geq 0$,

$$\begin{aligned}
&\mathrm{cov}_{y|\Theta_b}(y^{(b)}_{t+k}, \tilde{x}^{(r)}_{t|t-1}) = \mathrm{cov}(C_b x^{(b)}_{t+k} + w^{(b)}_{t+k}, \tilde{x}^{(r)}_{t|t-1}) \\
&= \mathrm{cov}(C_b(A_b^k x^{(b)}_t + \sum_{l=1}^{k} A_b^{k-l} v^{(b)}_{t+l}) + w^{(b)}_{t+k}, \tilde{x}^{(r)}_{t|t-1}) \qquad (104) \\
&= \mathrm{cov}_{x_t, y_{1:t-1}|\Theta_b}(C_b A_b^k x^{(b)}_t, \tilde{x}^{(r)}_{t|t-1}) \qquad (105) \\
&= C_b A_b^k \mathrm{cov}_{y_{1:t-1}|\Theta_b}(\mathbb{E}_{x_t|y_{1:t-1}}[x^{(b)}_t], \tilde{x}^{(r)}_{t|t-1}) \qquad (106) \\
&= C_b A_b^k \mathrm{cov}_{y|\Theta_b}(\tilde{x}^{(b)}_{t|t-1}, \tilde{x}^{(r)}_{t|t-1}) = C_b A_b^k \hat{\mathbf{V}}^{[2,3]}_t, \qquad (107)
\end{aligned}$$

where in (105) we use $v^{(b)}_{t+l} \perp\!\!\!\perp \hat{x}^{(r)}_{t|t-1}$ for $l \geq 1$ and $w^{(b)}_{t+k} \perp\!\!\!\perp \tilde{x}^{(r)}_{t|t-1}$ for $k \geq 0$, and in (106), $\mathrm{cov}_{x,y}(x, y) = \mathrm{cov}_y(\mathbb{E}_{x|y}[x], y)$.