Derivation of the Hierarchical EM algorithm for Dynamic Textures

Antoni B. Chan
Dept. of Computer Science
City University of Hong Kong

Emanuele Coviello   Gert. R. G. Lanckriet
Dept. of Electrical and Computer Engineering
University of California, San Diego

Abstract

This is the supplemental material for the CVPR 2010 paper, "Clustering Dynamic Textures with the Hierarchical EM Algorithm" [1]. It contains the derivation of the HEM-DTM algorithm and the associated E-step computations, including sensitivity analysis for the Kalman smoothing filter.

1 Derivation of HEM for dynamic textures

In this section we derive the HEM algorithm for dynamic textures. We begin with the derivation of the $\mathcal{Q}$ function, followed by the E- and M-steps.

1.1 $\mathcal{Q}$ function for HEM-DTM

In the E-step, the $\mathcal{Q}$ function is obtained by taking the conditional expectation, with respect to the hidden variables $\{X, Z\}$, of the complete-data likelihood in (7)

$$\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) = \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \mathbb{E}_{X,Z|Y,\hat{\Theta}^{(r)}} \left[\mathbf{z}_{i,j} \log \pi_j^{(r)} + \mathbf{z}_{i,j} \log p(Y_i, X_i | \Theta_j^{(r)}) \right] \tag{S.1}$$

$$= \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} + \hat{\mathbf{z}}_{i,j} \mathbb{E}_{X_i|Y_i,\hat{\Theta}_j^{(r)}} \left[\log p(Y_i, X_i | \Theta_j^{(r)}) \right], \tag{S.2}$$

where

$$\hat{\mathbf{z}}_{i,j} = \mathbb{E}_{Z|Y,\hat{\Theta}^{(r)}}[\mathbf{z}_{i,j}] = p(z_i^{(r)} = j | Y_i, \hat{\Theta}^{(r)}) \tag{S.3}$$

$$= \frac{\pi_j^{(r)} p(Y_i | \hat{\Theta}_j^{(r)})}{\sum_{j'=1}^{K^{(r)}} \pi_{j'}^{(r)} p(Y_i | \hat{\Theta}_{j'}^{(r)})} \tag{S.4}$$

is the probability that sample set Y_i is assigned to component j in $\Theta^{(r)}$. For the likelihood of the virtual samples, $p(Y_i|\hat{\Theta}_j^{(r)})$, we have

$$\log p(Y_i | \hat{\Theta}_j^{(r)}) = \sum_{m=1}^{N_i} \log p(y_{1:\tau}^{(i,m)} | \hat{\Theta}_j^{(r)}) \tag{S.5}$$

$$= N_i \left[\frac{1}{N_i} \sum_{m=1}^{N_i} \log p(y_{1:\tau}^{(i,m)} | \hat{\Theta}_j^{(r)}) \right] \tag{S.6}$$

$$\approx N_i \mathbb{E}_{\Theta_i^{(b)}} \left[\log p(y_{1:\tau} | \hat{\Theta}_j^{(r)}) \right] \tag{S.7}$$

where (S.7) follows from the law of large numbers [2] (as $N_i \to \infty$). Substituting into (S.4), we get the expression for $\hat{\mathbf{z}}_{i,j}$ similar to the one derived in [2],

$$\hat{\mathbf{z}}_{i,j} = \frac{\pi_j^{(r)} \exp\left(N_i \mathbb{E}_{\Theta_i^{(b)}}[\log p(y_{1:\tau}|\hat{\Theta}_j^{(r)})] \right)}{\sum_{j'=1}^{K^{(r)}} \pi_{j'}^{(r)} \exp\left(N_i \mathbb{E}_{\Theta_i^{(b)}}[\log p(y_{1:\tau}|\hat{\Theta}_{j'}^{(r)})] \right)}. \tag{S.8}$$

For the last term in (S.2), we have

$$\begin{aligned} \mathbb{E}_{X_i|Y_i,\hat{\Theta}_j^{(r)}}[\log p(Y_i, X_i|\Theta_j^{(r)})] &= \sum_{m=1}^{N_i} \mathbb{E}_{x_{1:\tau}^{(i,m)}|y_{1:\tau}^{(i,m)},\hat{\Theta}_j^{(r)}} \log p(y_{1:\tau}^{(i,m)}, x_{1:\tau}^{(i,m)}|\Theta_j^{(r)}) \\ &\approx N_i \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\log p(y_{1:\tau}, x_{1:\tau}|\Theta_j^{(r)})] \right], \end{aligned} \tag{S.9}$$

where, again, (S.9) follows from the law of large numbers. Hence, the $\mathcal{Q}$ function is given by

$$\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) = \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} + \hat{\mathbf{z}}_{i,j} N_i \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\log p(y_{1:\tau}, x_{1:\tau}|\Theta_j^{(r)})] \right]. \tag{S.10}$$

Note that the form of (S.10) is very similar to the $\mathcal{Q}$ function for the EM algorithm for DTM [3], where each base DT $\Theta_i^{(b)}$ in (S.10) corresponds to a single sample sequence y_i in EM for DTM. For the $\mathcal{Q}$ function of EM, the second term is the conditional expectation of the hidden states conditioned on the observation y_i, i.e.,, $\mathbb{E}_{x|y_i,\hat{\Theta}_j^{(r)}}[\cdot]$. In the case of HEM, the double expectation $\mathbb{E}_{y|\Theta_i^{(b)}}[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\cdot]]$ averages this conditional expectation over all possible observations y_i generated from the base DT $\Theta_i^{(b)}$. There is also an additional weighting of $N_i = \pi_i^{(b)} N$ on these expectations, which accounts for the prior probabilities of the base components.

1.2 E-step

Substituting for the joint likelihood of each DT component into (S.9), and defining $\hat{\mathbf{w}}_{i,j} = \hat{\mathbf{z}}_{i,j} N_i$,

$$\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) = \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} + \hat{\mathbf{w}}_{i,j} \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\log p(y_{1:\tau}|x_{1:\tau}, \Theta_j^{(r)}) p(x_{1:\tau}|\Theta_j^{(r)})] \right] \tag{S.11}$$

$$\begin{aligned} = \; & \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} \\ & + \hat{\mathbf{w}}_{i,j} \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\log \prod_{t=1}^{\tau} p(y_t|x_t, \Theta_j^{(r)}) p(x_1|\Theta_j^{(r)}) \prod_{t=2}^{\tau} p(x_t|x_{t-1}, \Theta_j^{(r)})] \right] \end{aligned} \tag{S.12}$$

$$\begin{aligned} = \; & \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} \\ & + \hat{\mathbf{w}}_{i,j} \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[\sum_{t=1}^{\tau} \log p(y_t|x_t, \Theta_j^{(r)}) + \log p(x_1|\Theta_j^{(r)}) + \sum_{t=2}^{\tau} \log p(x_t|x_{t-1}, \Theta_j^{(r)})] \right]. \end{aligned} \tag{S.13}$$

The conditional distributions are Gaussian, and hence

$$
\begin{aligned}
&\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) \\
&= \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} + \hat{\mathbf{w}}_{i,j} \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}} \left[-\frac{1}{2} \sum_{t=1}^{\tau} \left(\|(y_t - \bar{y}_j) - C_j x_t\|_{R_j}^2 + \log|R_j| \right) \right.\right. \\
&\quad \left.\left. - \frac{1}{2} \left(\|x_1 - \mu_j\|_{S_j}^2 + \log|S_j| \right) - \frac{1}{2} \sum_{t=2}^{\tau} \left(\|x_t - A_j x_{t-1}\|_{Q_j}^2 + \log|Q_j| \right) \right] \right],
\end{aligned}
\tag{S.14}
$$

where we have dropped the constant terms that do not affect the maximization of $\mathcal{Q}$ later. For convenience, we define the following expectations,

$$
\hat{x}_{t|j}^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[x_t] \right], \tag{S.15}
$$

$$
\hat{P}_{t,t|j}^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[x_t x_t^T] \right], \tag{S.16}
$$

$$
\hat{P}_{t,t-1|j}^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} \left[\mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[x_t x_{t-1}^T] \right], \tag{S.17}
$$

$$
\hat{W}_{t|j}^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} \left[(y_t - \bar{y}_j) \mathbb{E}_{x|y,\hat{\Theta}_j^{(r)}}[x_t]^T \right], \tag{S.18}
$$

$$
\hat{U}_{t|j}^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} \left[(y_t - \bar{y}_j)(y_t - \bar{y}_j)^T \right], \tag{S.19}
$$

$$
\hat{u}_t^{(i)} = \mathbb{E}_{y|\Theta_i^{(b)}} [y_t]. \tag{S.20}
$$

Taking the trace of each norm term in (S.14) and substituting yields,

$$
\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) = \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{z}}_{i,j} \log \pi_j^{(r)} \tag{S.21}
$$

$$
\begin{aligned}
&+ \sum_{i=1}^{K^{(b)}} \sum_{j=1}^{K^{(r)}} \hat{\mathbf{w}}_{i,j} \left[-\frac{1}{2} \sum_{t=1}^{\tau} \operatorname{tr}\left(R_j^{-1} \left(\hat{U}_{t|j}^{(i)} - \hat{W}_{t|j}^{(i)} C_j^T - C_j (\hat{W}_{t|j}^{(i)})^T + C_j \hat{P}_{t|j}^{(i)} C_j^T \right) \right) - \frac{\tau}{2} \log|R_j| \right. \qquad \text{(S.22)} \\
&- \frac{1}{2} \operatorname{tr}\left(S_j^{-1} \left(\hat{P}_{1,1|j}^{(i)} - \hat{x}_{1|j}^{(i)} \mu_j^T - \mu_j (\hat{x}_{1|j}^{(i)})^T + \mu_j \mu_j^T \right) \right) - \frac{1}{2} \log|S_j| \\
&\left. - \frac{1}{2} \sum_{t=2}^{\tau} \operatorname{tr}\left(Q_j^{-1} \left(\hat{P}_{t|j}^{(i)} - \hat{P}_{t,t-1|j}^{(i)} A_j^T - A_j (\hat{P}_{t,t-1|j}^{(i)})^T + A_j \hat{P}_{t-1|j}^{(i)} A_j^T \right) \right) - \frac{\tau - 1}{2} \log|Q_j| \right].
\end{aligned}
$$

Finally, we define the expectations, aggregated over the base components,

$$
\begin{aligned}
\hat{N}_j &= \textstyle\sum_i \hat{\mathbf{z}}_{i,j}, & \Phi_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=1}^{\tau} \hat{P}_{t,t|j}^{(i)}, \\
\hat{M}_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j}, & \Psi_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=2}^{\tau} \hat{P}_{t,t-1|j}^{(i)}, \\
\xi_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \hat{x}_{1|j}^{(i)}, & \varphi_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=2}^{\tau} \hat{P}_{t,t|j}^{(i)}, \\
\eta_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \hat{P}_{1,1|j}^{(i)}, & \phi_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=2}^{\tau} \hat{P}_{t-1,t-1|j}^{(i)}, \\
\gamma_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=1}^{\tau} \hat{u}_t^{(i)}, & \Lambda_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=1}^{\tau} \hat{U}_{t|j}^{(i)}, \\
\beta_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=1}^{\tau} \hat{x}_{t|j}^{(i)}, & \Gamma_j &= \textstyle\sum_i \hat{\mathbf{w}}_{i,j} \sum_{t=1}^{\tau} \hat{W}_{t|j}^{(i)},
\end{aligned}
\tag{S.23}
$$

yielding the $\mathcal{Q}$ function,

$$\mathcal{Q}(\Theta^{(r)}, \hat{\Theta}^{(r)}) = \sum_{j=1}^{K^{(r)}} \hat{N}_j \log \pi_j^{(r)} \tag{S.24}$$
$$- \frac{1}{2} \sum_{j=1}^{K^{(r)}} \Big[\mathrm{tr} \left(R_j^{-1} \left(\Lambda_j - \Gamma_j C_j^T - C_j \Gamma_j^T + C_j \Phi_j^{(i)} C_j^T \right) \right) + \tau \hat{M}_j \log |R_j|$$
$$+ \mathrm{tr} \left(S_j^{-1} \left(\eta_j - \xi_j \mu_j^T - \mu_j \xi_j^T + \hat{M}_j \mu_j \mu_j^T \right) \right) + \hat{M}_j \log |S_j|$$
$$+ \mathrm{tr} \left(Q_j^{-1} \left(\varphi_j - \Psi_j A_j^T - A_j \Psi_j^T + A_j \phi_j A_j^T \right) \right) + (\tau - 1) \hat{M}_j \log |Q_j| \Big] .$$

Hence, the E-step consists of computing the aggregate statistics in (S.23), which are based on the expectations in (S.15-S.20) and (S.8).

1.3 M-step

In the M-step, the parameters $\Theta^{(r)}$ are updated by maximizing the $\mathcal{Q}$ function in (S.24). Given the aggregate expectations defined in (S.23), the $\mathcal{Q}$ functions for HEM and EM for DTM are identical. Hence, the parameters for each component DT are updated according to

$$\begin{array}{ll} C_j^* = \Gamma_j \Phi_j^{-1}, & R_j^* = \frac{1}{\tau \hat{M}_j} (\Lambda_j - C_j^* \Gamma_j), \\ A_j^* = \Psi_j \phi_j^{-1}, & Q_j^* = \frac{1}{(\tau-1)\hat{M}_j} (\varphi_j - A_j^* \Psi_j^T), \\ \mu_j^* = \frac{1}{\hat{M}_j} \xi_j, & S_j^* = \frac{1}{\hat{M}_j} \eta_j - \mu_j^* (\mu_j^*)^T, \\ \pi_j^* = \frac{\hat{N}_j}{K^{(b)}}, & \bar{y}_j^* = \frac{1}{\tau \hat{M}_j} (\gamma_j - C_j^* \beta_j). \end{array} \tag{S.25}$$

2 Computing the E-step for HEM-DTM

The E-step of HEM-DTM requires computing expectations of the form,

$$\begin{array}{ll} \hat{x}_t^{(b)} = \mathbb{E}_{y|\Theta_b} \left[\mathbb{E}_{x|y,\Theta_r} [x_t] \right], & \hat{P}_t^{(b)} = \mathbb{E}_{y|\Theta_b} \left[\mathbb{E}_{x|y,\Theta_r} [x_t x_t^T] \right], \\ \hat{P}_{t,t-1}^{(b)} = \mathbb{E}_{y|\Theta_b} \left[\mathbb{E}_{x|y,\Theta_r} [x_t x_{t-1}^T] \right], & \hat{W}_t^{(b)} = \mathbb{E}_{y|\Theta_b} \left[(y_t - \bar{y}_r) \mathbb{E}_{x|y,\Theta_r} [x_t]^T \right], \\ \hat{U}_t^{(b)} = \mathbb{E}_{y|\Theta_b} \left[(y_t - \bar{y}_r)(y_t - \bar{y}_r)^T \right], & \hat{u}_t^{(b)} = \mathbb{E}_{y|\Theta_b} [y_t], \end{array} \tag{S.26}$$

where $\Theta_b = \{A_b, C_b, Q_b, R_b, \mu_b, S_b, \bar{y}_b\}$ and $\Theta_r = \{A_r, C_r, Q_r, R_r, \mu_r, S_r, \bar{y}_r\}$ are the dynamic texture parameters of a base mixture component and a reduced mixture component, respectively. The inner expectations, $\mathbb{E}_{x|y,\Theta_r}[\cdot]$, are the conditional state mean and second moment of the Kalman smoothing filter,

$$\hat{x}_{t|\tau}^{(r)} = \mathbb{E}_{x|y,\Theta_r}[x_t] \tag{S.27}$$
$$\hat{V}_{t|\tau}^{(r)} = \mathrm{cov}_{x|y,\Theta_r}(x_t) \tag{S.28}$$
$$\hat{V}_{t,t-1|\tau}^{(r)} = \mathrm{cov}_{x|y,\Theta_r}(x_t, x_{t-1}) \tag{S.29}$$

where we use the subscript notation "${}_{t|s}$" as short-hand for the expectation at time t, conditioned on sequence $y_{1:s} = \{y_1, \cdots, y_s\}$. Rewriting the first four expectations of (S.26) in terms of the Kalman smoothing filter, we have

$$\hat{x}_t^{(b)} = \mathbb{E}_{y|\Theta_b}\left[\hat{x}_{t|\tau}^{(r)}\right], \tag{S.30}$$

$$\hat{P}_t^{(b)} = \mathbb{E}_{y|\Theta_b}\left[\hat{V}_{t|\tau}^{(r)} + \hat{x}_{t|\tau}^{(r)}(\hat{x}_{t|\tau}^{(r)})^T\right] = \hat{V}_t^{(b)} + \hat{\chi}_t^{(b)} + \hat{x}_t^{(b)}(\hat{x}_t^{(b)})^T, \tag{S.31}$$

$$\hat{P}_{t,t-1}^{(b)} = \mathbb{E}_{y|\Theta_b}\left[\hat{V}_{t,t-1|\tau}^{(r)} + \hat{x}_{t|\tau}^{(r)}(\hat{x}_{t-1|\tau}^{(r)})^T\right] = \hat{V}_{t,t-1}^{(b)} + \hat{\chi}_{t,t-1}^{(b)} + \hat{x}_t^{(b)}(\hat{x}_{t-1}^{(b)})^T, \tag{S.32}$$

$$\hat{W}_t^{(b)} = \mathbb{E}_{y|\Theta_b}\left[(y_t - \bar{y}_r)(\hat{x}_{t|\tau}^{(r)})^T\right] = \hat{\kappa}_t^{(b)} + (\hat{u}_t^{(b)} - \bar{y}_r)(\hat{x}_t^{(b)})^T, \tag{S.33}$$

where we define

$$\hat{V}_t^{(b)} = \mathbb{E}_{y|\Theta_b}[\hat{V}_{t|\tau}^{(r)}], \tag{S.34}$$

$$\hat{V}_{t,t-1}^{(b)} = \mathbb{E}_{y|\Theta_b}[\hat{V}_{t,t-1|\tau}^{(r)}], \tag{S.35}$$

$$\hat{\chi}_t^{(b)} = \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}), \tag{S.36}$$

$$\hat{\chi}_{t,t-1}^{(b)} = \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t-1|\tau}^{(r)}), \tag{S.37}$$

$$\hat{\kappa}_t^{(b)} = \text{cov}_{y|\Theta_b}(y_t, \hat{x}_{t|\tau}^{(r)}). \tag{S.38}$$

The quantities in (S.30) and (S.34-S.38) are the output of the Kalman filter for Θ_r, when the input is from source Θ_b. These quantities can be computed using sensitivity analysis for the Kalman smoothing filter, derived in the remainder of the section. The final two expectations in (S.26) are computed from the marginal statistics of Θ_b,

$$\hat{U}_t^{(b)} = \mathbb{E}_{y|\Theta_b}\left[(y_t - \bar{y}_r)(y_t - \bar{y}_r)^T\right] \tag{S.39}$$

$$= \mathbb{E}_{y|\Theta_b}\left[(y_t - \hat{u}_t^{(b)} + \hat{u}_t^{(b)} - \bar{y}_r)(y_t - \hat{u}_t^{(b)} + \hat{u}_t^{(b)} - \bar{y}_r)^T\right] \tag{S.40}$$

$$= \text{cov}_{y|\Theta_b}(y_t, y_t) + (\hat{u}_t^{(b)} - \bar{y}_r)(\hat{u}_t^{(b)} - \bar{y}_r)^T \tag{S.41}$$

$$= C_b \text{cov}_{x|\Theta_b}(x_t) C_b^T + R_b + (\hat{u}_t^{(b)} - \bar{y}_r)(\hat{u}_t^{(b)} - \bar{y}_r)^T, \tag{S.42}$$

$$\hat{u}_t^{(b)} = \mathbb{E}_{y|\Theta_b}[y_t] = C_b \mathbb{E}_{x|\Theta_b}[x_t] + \bar{y}_b. \tag{S.43}$$

We next look at the Kalman smoothing filter.

2.1 Kalman smoothing filter

The Kalman filter computes the mean and covariance of the state x_t of an LDS, conditioned on the observed sequence $y_{1:t-1} = \{y_1, \dots, y_{t-1}\}$. The conditional expectations

$$\hat{x}_{t|t-1} = \mathbb{E}_{x|y_{1:t-1}}[x_t], \tag{S.44}$$

$$\hat{V}_{t|t-1} = \text{cov}_{x|y_{1:t-1}}(x_t), \tag{S.45}$$

are calculated via a set of forward recursive equations [4,5], for $t = \{1, \cdots, \tau\}$

$$\hat{V}_{t|t-1} = A\hat{V}_{t-1|t-1}A^T + Q, \tag{S.46}$$

$$K_t = \hat{V}_{t|t-1}C^T(C\hat{V}_{t|t-1}C^T + R)^{-1}, \tag{S.47}$$

$$\hat{V}_{t|t} = (I - K_t C)\hat{V}_{t|t-1}, \tag{S.48}$$

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}, \tag{S.49}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - C\hat{x}_{t|t-1} - \bar{y}), \tag{S.50}$$

with initial conditions $\hat{x}_{1|0} = \mu$ and $\hat{V}_{1|0} = S$. The conditional expectations, (S.27-S.29), of the Kalman smoothing filter are obtained with backward recursion. For $t = \tau, \cdots, 1$,

$$\begin{aligned} J_{t-1} &= \hat{V}_{t-1|t-1} A^T (\hat{V}_{t|t-1})^{-1}, && \text{(S.51)} \\ \hat{x}_{t-1|\tau} &= \hat{x}_{t-1|t-1} + J_{t-1}(\hat{x}_{t|\tau} - A\hat{x}_{t-1|t-1}), && \text{(S.52)} \\ \hat{V}_{t-1|\tau} &= \hat{V}_{t-1|t-1} + J_{t-1}(\hat{V}_{t|\tau} - \hat{V}_{t|t-1})J_{t-1}^T, && \text{(S.53)} \\ \hat{V}_{t-1,t-2|\tau} &= \hat{V}_{t-1|t-1}J_{t-2}^T + J_{t-1}(\hat{V}_{t,t-1|\tau} - A\hat{V}_{t-1|t-1})J_{t-2}^T, && \text{(S.54)} \end{aligned}$$

with initial conditions $\hat{x}_{\tau|\tau}$ and $\hat{V}_{\tau|\tau}$ computed by the forward recursions above, and $\hat{V}_{\tau,\tau-1|\tau} = (I - K_\tau C)A\hat{V}_{\tau-1|\tau-1}$. To apply sensitivity analysis in the next section, it will be convenient to rewrite the state estimators as functions only of $\hat{x}_{t|t-1}$ and $\hat{x}_{t|\tau}$,

$$\begin{aligned} G_{t-1} &= AK_{t-1}, && \text{(S.55)} \\ F_{t-1} &= A - AK_{t-1}C, && \text{(S.56)} \\ H_{t-1} &= A^{-1} - J_{t-1}, && \text{(S.57)} \\ \hat{x}_{t|t-1} &= F_{t-1}\hat{x}_{t-1|t-2} + G_{t-1}(y_{t-1} - \bar{y}), && \text{(S.58)} \\ \hat{x}_{t-1|\tau} &= H_{t-1}\hat{x}_{t|t-1} + J_{t-1}\hat{x}_{t|\tau}. && \text{(S.59)} \end{aligned}$$

Finally, note that the conditional covariances $\hat{V}_{t|\tau}$ and $\hat{V}_{t,t-1|\tau}$ are not functions of the observed sequence $y_{1:\tau}$. Hence, for the expectations in (S.34) and (S.35), we have

$$\begin{aligned} \hat{V}_t^{(b)} &= \mathbb{E}_{y|\Theta_b}[\hat{V}_{t|\tau}^{(r)}] = \hat{V}_{t|\tau}^{(r)}, && \text{(S.60)} \\ \hat{V}_{t,t-1}^{(b)} &= \mathbb{E}_{y|\Theta_b}[\hat{V}_{t,t-1|\tau}^{(r)}] = \hat{V}_{t,t-1|\tau}^{(r)}. && \text{(S.61)} \end{aligned}$$

We next analyze the sensitivity of the Kalman smoothing filter when the input distribution does not match the parameters of the filter.

2.2 Sensitivity analysis of the Kalman smoothing filter

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

$$\begin{aligned} \hat{x}_t^{(b)} &= \mathbb{E}_{y|\Theta_b}\left[\hat{x}_{t|\tau}^{(r)}\right], && \text{(S.62)} \\ \hat{\chi}_t^{(b)} &= \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}), && \text{(S.63)} \\ \hat{\chi}_{t,t-1}^{(b)} &= \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t-1|\tau}^{(r)}), && \text{(S.64)} \\ \hat{\kappa}_t^{(b)} &= \mathrm{cov}_{y|\Theta_b}(y_t, \hat{x}_{t|\tau}^{(r)}). && \text{(S.65)} \end{aligned}$$

We next derive an efficient algorithm for sensitivity analysis of the discrete time Kalman smoothing filter.

2.2.1 Forward recursion

We first need to analyze the sensitivity of the forward recursions of the Kalman filter [5]. From (S.58), the Kalman filters for Θ_b and Θ_r are recursively defined by

$$\begin{bmatrix} \hat{x}_{t|t-1}^{(b)} \\ \hat{x}_{t|t-1}^{(r)} \end{bmatrix} = \begin{bmatrix} F_{t-1}^{(b)}\hat{x}_{t-1|t-2}^{(b)} + G_{t-1}^{(b)}(y_{t-1}^{(b)} - \bar{y}_b) \\ F_{t-1}^{(r)}\hat{x}_{t-1|t-2}^{(r)} + G_{t-1}^{(r)}(y_{t-1}^{(b)} - \bar{y}_r) \end{bmatrix}, \qquad \text{(S.66)}$$

where $\{y_t^{(b)}\}$ are the observations from source Θ_b. Substituting $y_{t-1}^{(b)} = C_b x_{t-1}^{(b)} + w_{t-1}^{(b)} + \bar{y}_b$, we have

$$\begin{aligned} \begin{bmatrix} x_t^{(b)} \\ \hat{x}_{t|t-1}^{(b)} \\ \hat{x}_{t|t-1}^{(r)} \end{bmatrix} &= \begin{bmatrix} A_b x_{t-1}^{(b)} + v_t^{(b)} \\ F_{t-1}^{(b)} \hat{x}_{t-1|t-2}^{(b)} + G_{t-1}^{(b)} (C_b x_{t-1}^{(b)} + w_{t-1}^{(b)}) \\ F_{t-1}^{(r)} \hat{x}_{t-1|t-2}^{(r)} + G_{t-1}^{(r)} (C_b x_{t-1}^{(b)} + w_{t-1}^{(b)} + \bar{y}_b - \bar{y}_r) \end{bmatrix} && \text{(S.67)} \\ &= \mathbf{A}_{t-1} \begin{bmatrix} x_{t-1}^{(b)} \\ \hat{x}_{t-1|t-2}^{(b)} \\ \hat{x}_{t-1|t-2}^{(r)} \end{bmatrix} + \mathbf{B} v_t^{(b)} + \mathbf{C}_{t-1} w_{t-1}^{(b)} + \mathbf{D}_{t-1} (\bar{y}_b - \bar{y}_r), && \text{(S.68)} \end{aligned}$$

where

$$\mathbf{A}_{t-1} = \begin{bmatrix} A_b & 0 & 0 \\ G_{t-1}^{(b)} C_b & F_{t-1}^{(b)} & 0 \\ G_{t-1}^{(r)} C_b & 0 & F_{t-1}^{(r)} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{C}_{t-1} = \begin{bmatrix} 0 \\ G_{t-1}^{(b)} \\ G_{t-1}^{(r)} \end{bmatrix}, \ \mathbf{D}_{t-1} = \begin{bmatrix} 0 \\ 0 \\ G_{t-1}^{(r)} \end{bmatrix}. \tag{S.69}$$

We define the block vector $\hat{\mathbf{x}}_t$ as

$$\hat{\mathbf{x}}_t = \begin{bmatrix} \mathbb{E}_{x|\Theta_b}[x_t^{(b)}] \\ \mathbb{E}_{y|\Theta_b}[\hat{x}_{t|t-1}^{(b)}] \\ \mathbb{E}_{y|\Theta_b}[\hat{x}_{t|t-1}^{(r)}] \end{bmatrix}. \tag{S.70}$$

Taking the expectation of (S.68) with respect to $x, y|\Theta_b$, $\hat{\mathbf{x}}_t$ can be computed recursively according to

$$\hat{\mathbf{x}}_t = \mathbf{A}_{t-1} \hat{\mathbf{x}}_{t-1} + \mathbf{D}_{t-1} (\bar{y}_b - \bar{y}_r), \quad \hat{\mathbf{x}}_1 = \begin{bmatrix} \mu_b \\ \mu_b \\ \mu_r \end{bmatrix}. \tag{S.71}$$

For the covariance of the state estimators, we define the block covariance matrix $\hat{\mathbf{V}}_t$ as

$$\hat{\mathbf{V}}_t = \begin{bmatrix} \mathrm{cov}_{x|\Theta_b}(x_t^{(b)}) & \mathrm{cov}_{x,y|\Theta_b}(x_t^{(b)}, \hat{x}_{t|t-1}^{(b)}) & \mathrm{cov}_{x,y|\Theta_b}(x_t^{(b)}, \hat{x}_{t|t-1}^{(r)}) \\ \mathrm{cov}_{x,y|\Theta_b}(\hat{x}_{t|t-1}^{(b)}, x_t^{(b)}) & \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(b)}) & \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(b)}, \hat{x}_{t|t-1}^{(r)}) \\ \mathrm{cov}_{x,y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}, x_t^{(b)}) & \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}, \hat{x}_{t|t-1}^{(b)}) & \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}) \end{bmatrix}. \tag{S.72}$$

From (S.68), the covariance is computed recursively as

$$\hat{\mathbf{V}}_t = \mathbf{A}_{t-1} \hat{\mathbf{V}}_{t-1} \mathbf{A}_{t-1}^T + \mathbf{B} Q_b \mathbf{B}^T + \mathbf{C}_{t-1} R_b \mathbf{C}_{t-1}^T, \quad \hat{\mathbf{V}}_1 = \begin{bmatrix} S_b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{S.73}$$

where we have used the fact that $\{v_t^{(b)}, w_{t-1}^{(b)}\} \perp\!\!\!\perp \{x_{t-1}^{(b)}, \hat{x}_{t-1|t-2}^{(b)}, \hat{x}_{t-1|t-2}^{(r)}\}$, and $v_t^{(b)} \perp\!\!\!\perp w_t^{(b)}$. We will use the notation $\mathbf{V}^{[i,j]}$ to refer to the (i, j) matrix in the block matrix $\mathbf{V}$, and $\mathbf{x}^{[i]}$ to refer to the ith vector in the block vector $\mathbf{x}$.

2.2.2 Backward recursion

Taking the expectation of (S.59) yields a recursive definition for $\hat{x}_t^{(b)}$,

$$\begin{aligned}\hat{x}_{t-1}^{(b)} &= \mathbb{E}_{y|\Theta_b}[\hat{x}_{t-1|\tau}^{(r)}] && \text{(S.74)}\\ &= H_{t-1}^{(r)}\mathbb{E}_{y|\Theta_b}[\hat{x}_{t|t-1}^{(r)}] + J_{t-1}^{(r)}\mathbb{E}_{y|\Theta_b}[\hat{x}_{t|\tau}^{(r)}] && \text{(S.75)}\\ &= H_{t-1}^{(r)}\mathbb{E}_{y|\Theta_b}[\hat{x}_{t|t-1}^{(r)}] + J_{t-1}^{(r)}\hat{x}_t^{(b)} && \text{(S.76)}\\ &= H_{t-1}^{(r)}\hat{\mathbf{x}}_t^{[3]} + J_{t-1}^{(r)}\hat{x}_t^{(b)}, && \text{(S.77)}\end{aligned}$$

where we use the notation $\hat{\mathbf{x}}_t^{[3]}$ to indicate the 3rd entry in the block vector $\hat{\mathbf{x}}_t$, and the initial condition is

$$\hat{x}_\tau^{(b)} = \mathbb{E}_{y|\Theta_b}[\hat{x}_{\tau|\tau}^{(r)}] = A_r^{-1}\mathbb{E}_{y|\Theta_b}[\hat{x}_{\tau+1|\tau}^{(r)}] = A_r^{-1}\hat{\mathbf{x}}_{\tau+1}^{[3]}. \quad \text{(S.78)}$$

For the covariance of $\hat{x}_{t|\tau}^{(r)}$, we have

$$\begin{aligned}\hat{\chi}_{t-1}^{(b)} &= \text{cov}_{y|\Theta_b}(\hat{x}_{t-1|\tau}^{(r)}) = \text{cov}_{y|\Theta_b}(H_{t-1}^{(r)}\hat{x}_{t|t-1}^{(r)} + J_{t-1}^{(r)}\hat{x}_{t|\tau}^{(r)}) && \text{(S.79)}\\ &= \begin{bmatrix} H_{t-1}^{(r)} & J_{t-1}^{(r)} \end{bmatrix} \begin{bmatrix} \text{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}) & \text{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}, \hat{x}_{t|\tau}^{(r)}) \\ \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t|t-1}^{(r)}) & \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}) \end{bmatrix} \begin{bmatrix} (H_{t-1}^{(r)})^T \\ (J_{t-1}^{(r)})^T \end{bmatrix} && \text{(S.80)}\\ &= \begin{bmatrix} H_{t-1}^{(r)} & J_{t-1}^{(r)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_t^{[3,3]} & \omega_t^T \\ \omega_t & \hat{\chi}_t^{(b)} \end{bmatrix} \begin{bmatrix} (H_{t-1}^{(r)})^T \\ (J_{t-1}^{(r)})^T \end{bmatrix}, && \text{(S.81)}\end{aligned}$$

and

$$\begin{aligned}\hat{\chi}_{t,t-1}^{(b)} &= \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t-1|\tau}^{(r)}) && \text{(S.82)}\\ &= \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, H_{t-1}^{(r)}\hat{x}_{t|t-1}^{(r)} + J_{t-1}^{(r)}\hat{x}_{t|\tau}^{(r)}) && \text{(S.83)}\\ &= \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t|t-1}^{(r)})(H_{t-1}^{(r)})^T + \hat{\chi}_t^{(b)}(J_{t-1}^{(r)})^T, && \text{(S.84)}\\ &= \omega_t(H_{t-1}^{(r)})^T + \hat{\chi}_t^{(b)}(J_{t-1}^{(r)})^T, && \text{(S.85)}\end{aligned}$$

where $\hat{\mathbf{V}}_t^{[3,3]}$ is the $(3,3)$ entry of the block covariance matrix $\hat{\mathbf{V}}_t$, and we define the cross-covariance,

$$\omega_t = \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t|t-1}^{(r)}). \quad \text{(S.86)}$$

The initial condition is given by

$$\hat{\chi}_\tau^{(b)} = \text{cov}_{y|\Theta_b}(\hat{x}_{\tau|\tau}^{(r)}) = \text{cov}_{y|\Theta_b}(A_r^{-1}\hat{x}_{\tau+1|\tau}^{(r)}) = A_r^{-1}\text{cov}_{y|\Theta_b}(\hat{x}_{\tau+1|\tau}^{(r)})A_r^{-T}. \quad \text{(S.87)}$$

All that is required now is an efficient method for computing the cross-covariance term ω_t and $\hat{\kappa}_t^{(b)}$.

2.3 Calculating the cross-covariance terms

We next look at calculating the cross-covariance terms,

$$\begin{aligned}\omega_t &= \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t|t-1}^{(r)}), && \text{(S.88)}\\ \rho_t &= (\hat{\kappa}_t^{(b)})^T = \text{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, y_t). && \text{(S.89)}\end{aligned}$$

Filling in the backward recursion of the state estimate of the Kalman smoothing filter, we have

$$\begin{aligned}
\hat{x}^{(r)}_{t|\tau} &= H^{(r)}_t \hat{x}^{(r)}_{t+1|t} + J^{(r)}_t \hat{x}^{(r)}_{t+1|\tau} && \text{(S.90)}\\
&= H^{(r)}_t \hat{x}^{(r)}_{t+1|t} + J^{(r)}_t (H^{(r)}_{t+1} \hat{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1} \hat{x}^{(r)}_{t+2|\tau}) && \text{(S.91)}\\
&= H^{(r)}_t \hat{x}^{(r)}_{t+1|t} + J^{(r)}_t (H^{(r)}_{t+1} \hat{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1} (\cdots + J^{(r)}_{\tau-2} (H^{(r)}_{\tau-1} \hat{x}^{(r)}_{\tau|\tau-1} + J^{(r)}_{\tau-1} \hat{x}^{(r)}_{\tau|\tau}) \cdots)) && \text{(S.92)}\\
&= H^{(r)}_t \hat{x}^{(r)}_{t+1|t} + J^{(r)}_t (H^{(r)}_{t+1} \hat{x}^{(r)}_{t+2|t+1} + J^{(r)}_{t+1} (\cdots + J^{(r)}_{\tau-2} (H^{(r)}_{\tau-1} \hat{x}^{(r)}_{\tau|\tau-1} + J^{(r)}_{\tau-1} A_r^{-1} \hat{x}^{(r)}_{\tau+1|\tau}) \cdots)) && \text{(S.93)}\\
&= H^{(r)}_t \hat{x}^{(r)}_{t+1|t} + \sum_{s=t+2}^{\tau} (\prod_{i=t}^{s-2} J^{(r)}_i) H^{(r)}_{s-1} \hat{x}^{(r)}_{s|s-1} + (\prod_{i=t}^{\tau-1} J^{(r)}_i) A_r^{-1} \hat{x}^{(r)}_{\tau+1|\tau} && \text{(S.94)}\\
&= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \hat{x}^{(r)}_{s|s-1}, && \text{(S.95)}
\end{aligned}$$

where we define $\hat{H}_t = \begin{cases} H^{(r)}_t & ,t < \tau \\ A_r^{-1} & ,t = \tau \end{cases}$, and $\mathbf{J}_{t,s} = \begin{cases} I & ,t > s \\ J^{(r)}_t J^{(r)}_{t+1} \cdots J^{(r)}_s & ,t \le s \end{cases}$. Next, we rewrite the Kalman filter terms $\hat{x}^{(r)}_{s|s-1}$, where $s > t$, as a function of $\hat{x}^{(r)}_{t|t-1}$ 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, we have

$$\begin{aligned}
\hat{x}^{(r)}_{t+1|t} &= F^{(r)}_t \hat{x}_{t|t-1} + G^{(r)}_t y_t, && \text{(S.96)}\\
\hat{x}^{(r)}_{t+2|t+1} &= F^{(r)}_{t+1} \hat{x}_{t+1|t} + G^{(r)}_{t+1} y_{t+1} && \text{(S.97)}\\
&= F^{(r)}_{t+1} (F^{(r)}_t \hat{x}_{t|t-1} + G^{(r)}_t y_t) + G^{(r)}_{t+1} y_{t+1}, && \text{(S.98)}\\
\hat{x}^{(r)}_{t+3|t+2} &= F^{(r)}_{t+2} \hat{x}_{t+2|t+1} + G^{(r)}_{t+2} y_{t+2} && \text{(S.99)}\\
&= F^{(r)}_{t+2} (F^{(r)}_{t+1} (F^{(r)}_t \hat{x}_{t|t-1} + G^{(r)}_t y_t) + G^{(r)}_{t+1} y_{t+1}) + G^{(r)}_{t+2} y_{t+2}, && \text{(S.100)}
\end{aligned}$$

or in general, for $s > t$,

$$\hat{x}^{(r)}_{s|s-1} = \underbrace{\mathbf{F}_{s-1,t+1} (F^{(r)}_t \hat{x}_{t|t-1} + G^{(r)}_t y_t)}_{\alpha_s} + \underbrace{\sum_{j=t+1}^{s-1} \mathbf{F}_{s-1,j+1} G^{(r)}_j y_j}_{\beta_s}, \qquad \text{(S.101)}$$

where we define $\mathbf{F}_{s,t} = \begin{cases} I & ,s < t \\ F_s F_{s-1} \cdots F_t & ,s \ge t \end{cases}$, and the quantities α_s and β_s as above. Note that $\beta_{t+1} = \mathbf{0}$. We now substitute each term of $\hat{x}^{(r)}_{s|s-1} = \alpha_s + \beta_s$ into (S.95). Substituting the α_s terms into (S.95),

$$\begin{aligned}
\hat{\alpha}_t &= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \alpha_s && \text{(S.102)}\\
&= \underbrace{\sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,t+1}}_{L_t} (F^{(r)}_t \hat{x}_{t|t-1} + G^{(r)}_t y_t). && \text{(S.103)}
\end{aligned}$$

L_t can be computed recursively,

$$
\begin{aligned}
L_t &= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,t+1} && \text{(S.104)} \\
&= \mathbf{J}_{t,t-1} \hat{H}_t \mathbf{F}_{t,t+1} + \sum_{s=t+2}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,t+1} && \text{(S.105)} \\
&= \hat{H}_t + \sum_{s=t+2}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,t+1} && \text{(S.106)} \\
&= \hat{H}_t + J_t^{(r)} \left(\sum_{s=t+2}^{\tau+1} \mathbf{J}_{t+1,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,t+2} \right) F_{t+1}^{(r)} && \text{(S.107)} \\
&= H_t^{(r)} + J_t^{(r)} L_{t+1} F_{t+1}^{(r)}, && \text{(S.108)}
\end{aligned}
$$

where in (S.105) we have separated the first term of the summation ($s = t + 1$) , and in (S.107) we have used $\mathbf{J}_{t,s-2} = J_t^{(r)} \mathbf{J}_{t+1,s-2}$ and $\mathbf{F}_{s-1,t+1} = \mathbf{F}_{s-1,t+2} F_{t+1}^{(r)}$, when $s \geq t$. The initial condition for the L_t backward recursion is given by

$$
L_\tau = \mathbf{J}_{\tau,\tau-1} \hat{H}_\tau \mathbf{F}_{\tau,\tau+1} = \hat{H}_\tau = A_r^{-1}. \qquad \text{(S.109)}
$$

Next, we look at substituting the β_s terms of (S.101) into (S.95),

$$
\begin{aligned}
\hat{\beta}_t &= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \beta_s && \text{(S.110)} \\
&= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \sum_{j=t+1}^{s-1} \mathbf{F}_{s-1,j+1} G_j^{(r)} y_j && \text{(S.111)} \\
&= \sum_{s=t+2}^{\tau+1} \sum_{j=t+1}^{s-1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,j+1} G_j^{(r)} y_j && \text{(S.112)} \\
&= \sum_{j=t+1}^{\tau} \sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} \mathbf{F}_{s-1,j+1} G_j^{(r)} y_j, && \text{(S.113)}
\end{aligned}
$$

where in (S.113) we have collected the $G_j^{(r)} y_j$ terms by switching the double summation. Finally, using (S.108), we can rewrite the Kalman smoothing filter of (S.95) as

$$
\begin{aligned}
\hat{x}_{t|\tau}^{(r)} &= \sum_{s=t+1}^{\tau+1} \mathbf{J}_{t,s-2} \hat{H}_{s-1} (\alpha_s + \beta_s) && \text{(S.114)} \\
&= \hat{\alpha}_t + \hat{\beta}_t && \text{(S.115)} \\
&= L_t (F_t^{(r)} \hat{x}_{t|t-1}^{(r)} + G_t^{(r)} y_t) + \hat{\beta}_t. && \text{(S.116)}
\end{aligned}
$$

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

Next, we will derive an expression for $\omega_t = \mathrm{cov}_{y|\Theta_b}(\hat{x}^{(r)}_{t|\tau}, \hat{x}^{(r)}_{t|t-1})$. We first note that, for $k \geq 0$,

$$\begin{aligned}
\mathrm{cov}_{y|\Theta_b}(y^{(b)}_{t+k}, \hat{x}^{(r)}_{t|t-1}) &= \mathrm{cov}_{\Theta_b}(C_b x^{(b)}_{t+k} + w^{(b)}_{t+k} + \bar{y}_b, \hat{x}^{(r)}_{t|t-1}) && \text{(S.117)}\\
&= \mathrm{cov}_{\Theta_b}(C_b(A_b^k x^{(b)}_t + \sum_{l=1}^{k} A_b^{k-l} v^{(b)}_{t+l}) + w^{(b)}_{t+k} + \bar{y}_b, \hat{x}^{(r)}_{t|t-1}) && \text{(S.118)}\\
&= \mathrm{cov}_{x_t, y_{1:t-1}|\Theta_b}(C_b A_b^k x^{(b)}_t, \hat{x}^{(r)}_{t|t-1}) && \text{(S.119)}\\
&= C_b A_b^k \mathrm{cov}_{y_{1:t-1}|\Theta_b}(\mathbb{E}_{x_t|y_{1:t-1}}[x^{(b)}_t], \hat{x}^{(r)}_{t|t-1}) && \text{(S.120)}\\
&= C_b A_b^k \mathrm{cov}_{y|\Theta_b}(\hat{x}^{(b)}_{t|t-1}, \hat{x}^{(r)}_{t|t-1}) && \text{(S.121)}\\
&= C_b A_b^k \hat{\mathbf{V}}^{[2,3]}_t, && \text{(S.122)}
\end{aligned}$$

where in (S.118) 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 (S.119) we have used the fact that the future state noise is independent of the current state (i.e.,for $l \geq 1$, $v^{(b)}_{t+l} \perp\!\!\!\perp \hat{x}^{(r)}_{t|t-1}$) and the current observation noise is independent of the current state and past observations (i.e.,for $k \geq 0$, $w^{(b)}_{t+k} \perp\!\!\!\perp \hat{x}^{(r)}_{t|t-1}$), and in (S.120), we have used $\mathrm{cov}_{x,y}(x, y) = \mathrm{cov}_y(\mathbb{E}_{x|y}[x], y)$. We also have the property

$$\begin{aligned}
\mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, \hat{x}^{(r)}_{t|t-1}) &= \mathrm{cov}_{y|\Theta_b}(\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j y_j, \hat{x}^{(r)}_{t|t-1}) && \text{(S.123)}\\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j \mathrm{cov}_{y|\Theta_b}(y_j, \hat{x}^{(r)}_{t|t-1}) && \text{(S.124)}\\
&= \underbrace{\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t}}_{M_t} \hat{\mathbf{V}}^{[2,3]}_t, && \text{(S.125)}
\end{aligned}$$

where we define M_t as above, and in (S.124) we have used (S.122). M_t can also be computed with a backward recursion,

$$\begin{aligned}
M_t &= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t} && \text{(S.126)}\\
&= J^{(r)}_t \left(\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-t-1}\right) A_b && \text{(S.127)}\\
&= J^{(r)}_t \left(\left[\sum_{s=t+2}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,t+2}G^{(r)}_{t+1} C_b\right]\right. && \text{(S.128)}\\
&\quad \left. + \sum_{j=t+2}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t+1,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G^{(r)}_j C_b A_b^{j-(t+1)}\right) A_b\\
&= J^{(r)}_t (L_{t+1} G^{(r)}_{t+1} C_b + M_{t+1}) A_b, && \text{(S.129)}
\end{aligned}$$

where in (S.129) we have separated the first term of the summation ($j = t + 1$), and in (S.129) we have used the definition of L_{t+1} and M_{t+1}. The initial condition is $M_\tau = \mathbf{0}$. Finally using (S.116), the cross-covariance is

$$\begin{aligned}
\omega_t &= \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, \hat{x}_{t|t-1}^{(r)}) && \text{(S.130)}\\
&= \mathrm{cov}_{y|\Theta_b}(L_t(F_t^{(r)}\hat{x}_{t|t-1}^{(r)} + G_t^{(r)}y_t) + \hat{\beta}_t, \hat{x}_{t|t-1}^{(r)}) && \text{(S.131)}\\
&= L_tF_t^{(r)}\mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}) + L_tG_t^{(r)}\mathrm{cov}_{y|\Theta_b}(y_t, \hat{x}_{t|t-1}^{(r)}) + \mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, \hat{x}_{t|t-1}^{(r)}) && \text{(S.132)}\\
&= L_tF_t^{(r)}\hat{\mathbf{V}}_t^{[3,3]} + L_tG_t^{(r)}C_b\hat{\mathbf{V}}_t^{[2,3]} + M_t\hat{\mathbf{V}}_t^{[2,3]} && \text{(S.133)}\\
&= L_tF_t^{(r)}\hat{\mathbf{V}}_t^{[3,3]} + (L_tG_t^{(r)}C_b + M_t)\hat{\mathbf{V}}_t^{[2,3]}, && \text{(S.134)}
\end{aligned}$$

where in (S.133) we have used (S.122) and (S.125).

To compute $\rho_t = \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, y_t)$, we first note that, for $k > 0$,

$$\begin{aligned}
\mathrm{cov}_{y|\Theta_b}(y_{t+k}^{(b)}, y_t^{(b)}) &= \mathrm{cov}_{\Theta_b}(C_bx_{t+k}^{(b)} + w_{t+k}^{(b)} + \bar{y}_b, C_bx_t^{(b)} + w_t^{(b)} + \bar{y}_b) && \text{(S.135)}\\
&= C_b\mathrm{cov}_{\Theta_b}(x_{t+k}^{(b)}, x_t^{(b)})C_b^T + \mathrm{cov}_{\Theta_b}(w_{t+k}^{(b)}, w_t^{(b)}) && \text{(S.136)}\\
&= C_b\mathrm{cov}_{\Theta_b}(A_b^kx_t^{(b)} + \sum_{l=1}^{k} A_b^{k-l}v_{t+l}^{(b)}, x_t^{(b)})C_b^T && \text{(S.137)}\\
&= C_bA_b^k\mathrm{cov}_{x|\Theta_b}(x_t^{(b)})C_b^T && \text{(S.138)}\\
&= C_bA_b^k\hat{\mathbf{V}}_t^{[1,1]}C_b^T, && \text{(S.139)}
\end{aligned}$$

where in (S.136) we have used the fact that the observation noise is independent of the hidden-state (i.e.,for $k \geq 0$, $x_t^{(b)} \perp\!\!\!\perp w_{t+k}^{(b)}$ and $x_{t+k}^{(b)} \perp\!\!\!\perp w_t^{(b)}$), and in (S.138) we have used the fact that future state noise is independent of the current state (i.e.,for $l \geq 1$, $x_t^{(b)} \perp\!\!\!\perp v_{t+l}^{(b)}$). We then can derive the property,

$$\begin{aligned}
\mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, y_t) &= \mathrm{cov}_{y|\Theta_b}(\sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G_j^{(r)}y_j, y_t) && \text{(S.140)}\\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G_j^{(r)}\mathrm{cov}_{y|\Theta_b}(y_j, y_t) && \text{(S.141)}\\
&= \sum_{j=t+1}^{\tau}\sum_{s=j+1}^{\tau+1} \mathbf{J}_{t,s-2}\hat{H}_{s-1}\mathbf{F}_{s-1,j+1}G_j^{(r)}C_bA_b^{j-t}\hat{\mathbf{V}}_t^{[1,1]}C_b^T && \text{(S.142)}\\
&= M_t\hat{\mathbf{V}}_t^{[1,1]}C_b^T, && \text{(S.143)}
\end{aligned}$$

where in (S.141) we have used (S.139). Finally, using (S.116), the cross-covariance is given by

$$\begin{aligned}
\rho_t &= \mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|\tau}^{(r)}, y_t) && \text{(S.144)}\\
&= \mathrm{cov}_{y|\Theta_b}(L_t(F_t^{(r)}\hat{x}_{t|t-1}^{(r)} + G_t^{(r)}y_t) + \hat{\beta}_t, y_t) && \text{(S.145)}\\
&= L_tF_t^{(r)}\mathrm{cov}_{y|\Theta_b}(\hat{x}_{t|t-1}^{(r)}, y_t) + L_tG_t^{(r)}\mathrm{cov}_{y|\Theta_b}(y_t) + \mathrm{cov}_{y|\Theta_b}(\hat{\beta}_t, y_t) && \text{(S.146)}\\
&= L_tF_t^{(r)}\hat{\mathbf{V}}_t^{[3,2]}C_b^T + L_tG_t^{(r)}(C_b\hat{\mathbf{V}}_t^{[1,1]}C_b^T + R_b) + M_t\hat{\mathbf{V}}_t^{[1,1]}C_b^T && \text{(S.147)}\\
&= \left(L_tF_t^{(r)}\hat{\mathbf{V}}_t^{[3,2]} + (L_tG_t^{(r)}C_b + M_t)\hat{\mathbf{V}}_t^{[1,1]}\right)C_b^T + L_tG_t^{(r)}R_b && \text{(S.148)}
\end{aligned}$$

where (S.147) follows by using (S.118), (S.143), and $\mathrm{cov}_{y|\Theta_b}(y_t) = C_b\mathrm{cov}_{y|\Theta_b}(x_t)C_b^T + R_b$.

2.4 Expected log-likelihood term

In the E-step, the soft assignments $\hat{\mathbf{z}}_{i,j}$ depend on computing the expected log-likelihoods of the form, $\mathbb{E}_{y|\Theta_b}[\log p(y|\Theta_r)]$. For dynamic textures, this can be computed efficiently using the Kalman filter. The observation log-likelihood of the DT can be expressed in "innovation" form [6],

$$\begin{aligned}
\log p(y|\Theta_r) &= \sum_{t=1}^{\tau} \log p(y_t|y_{1:t-1},\Theta_r) && \text{(S.149)}\\
&= \sum_{t=1}^{\tau} \log G(y_t, C_r\hat{x}^{(r)}_{t|t-1} + \bar{y}_r, C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r) && \text{(S.150)}\\
&= \sum_{t=1}^{\tau} -\frac{1}{2}\left\|y_t - C_r\hat{x}^{(r)}_{t|t-1} - \bar{y}_r\right\|^2_{C_r\hat{V}^{(r)}_{t|t-1}C_r^T+R_r} && \text{(S.151)}\\
&\quad -\frac{1}{2}\log\left|C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r\right| - \frac{m}{2}\log(2\pi)\\
&= \sum_{t=1}^{\tau} -\frac{1}{2}\mathrm{tr}\left((C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r)^{-1}(y_t - \bar{y}_r - C_r\hat{x}^{(r)}_{t|t-1})(y_t - \bar{y}_r - C_r\hat{x}^{(r)}_{t|t-1})^T\right)\\
&\quad -\frac{1}{2}\log\left|C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r\right| - \frac{m}{2}\log(2\pi). && \text{(S.152)}
\end{aligned}$$

Taking the expectation of (S.152), and noting that $\hat{V}^{(r)}_{t|t-1}$ is not a function of the observations y_t,

$$\begin{aligned}
\ell &= \mathbb{E}_{y|\Theta_b}[\log p(y|\Theta_r)] && \text{(S.153)}\\
&= \sum_{t=1}^{\tau} -\frac{1}{2}\mathrm{tr}\left((C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r)^{-1}(\hat{U}^{(b)}_t - \hat{\lambda}^{(b)}_t C_r^T - C_r(\hat{\lambda}^{(b)}_t)^T + C_r\hat{\Lambda}^{(b)}_t C_r^T)\right)\\
&\quad -\frac{1}{2}\log\left|C_r\hat{V}^{(r)}_{t|t-1}C_r^T + R_r\right| - \frac{m}{2}\log(2\pi), && \text{(S.154)}
\end{aligned}$$

where

$$\begin{aligned}
\hat{\Lambda}^{(b)}_t &= \mathbb{E}_{y|\Theta_b}[\hat{x}^{(r)}_{t|t-1}(\hat{x}^{(r)}_{t|t-1})^T] = \mathrm{cov}_{y|\Theta_b}(\hat{x}^{(r)}_{t|t-1}) + \mathbb{E}_{y|\Theta_b}[\hat{x}^{(r)}_{t|t-1}]\mathbb{E}_{y|\Theta_b}[\hat{x}^{(r)}_{t|t-1}]^T && \text{(S.155)}\\
&= \hat{\mathbf{V}}^{[3,3]}_t + \hat{\mathbf{x}}^{[3]}_t(\hat{\mathbf{x}}^{[3]}_t)^T, && \text{(S.156)}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\lambda}^{(b)}_t &= \mathbb{E}_{y|\Theta_b}[(y_t - \bar{y}_r)(\hat{x}^{(r)}_{t|t-1})^T] && \text{(S.157)}\\
&= \mathrm{cov}_{y|\Theta_b}(y_t - \bar{y}_r, \hat{x}^{(r)}_{t|t-1}) + \mathbb{E}_{y|\Theta_b}[y_t - \bar{y}_r]\mathbb{E}_{y|\Theta_b}[\hat{x}^{(r)}_{t|t-1}]^T && \text{(S.158)}\\
&= C_b\hat{\mathbf{V}}^{[2,3]}_t + (C_b\hat{\mathbf{x}}^{[1]}_t + \bar{y}_b - \bar{y}_r)(\hat{\mathbf{x}}^{[3]}_t)^T. && \text{(S.159)}
\end{aligned}$$

2.5 Summary

The E-step for HEM-DTM is summarized in Algorithm 1. The Kalman smoothing filter is summarized in Algorithm 2. Computing the E-step requires running sensitivity analysis on the Kalman and Kalman smoothing filters, which are summarized in Algorithms 3 and 4.

Algorithm 1 Expectations for HEM-DTM

1: **Input**: DT parameters Θ_b and Θ_r, length τ.

2: Run Kalman smoothing filter (Algorithm 2) on Θ_b and Θ_r to obtain filter matrices $\{\hat{V}^{(b)}_{t|t-1}, \hat{V}^{(b)}_{t|t}, \hat{V}^{(b)}_{t|\tau}, \hat{V}^{(b)}_{t,t-1|\tau}, G^{(b)}_t, F^{(b)}_t\}$ and $\{\hat{V}^{(r)}_{t|t-1}, \hat{V}^{(r)}_{t|t}, \hat{V}^{(r)}_{t|\tau}, \hat{V}^{(r)}_{t,t-1|\tau}, G^{(r)}_t, F^{(r)}_t, H^{(r)}_t, J^{(r)}_t\}$.

3: Run sensitivity analysis for the Kalman filters, Θ_b and Θ_r, (Algorithm 3) to obtain $\{\hat{\mathbf{x}}_t, \hat{\mathbf{V}}_t\}$.

4: Run sensitivity analysis for the Kalman smoothing filters, Θ_b and Θ_r, (Algorithm 4) to obtain $\{\hat{x}^{(b)}_t, \hat{\chi}^{(b)}_t, \hat{\chi}^{(b)}_{t,t-1}, \hat{\kappa}^{(b)}_t\}$.

5: Compute E-step expectations, for $t = \{1, \cdots, \tau\}$:

$$\hat{u}^{(b)}_t = C_b \hat{\mathbf{x}}^{[1]}_t + \bar{y}_b, \tag{S.160}$$

$$\hat{U}^{(b)}_t = C_b \hat{\mathbf{V}}^{[1,1]}_t C_b^T + R_b + (\hat{u}^{(b)}_t - \bar{y}_r)(\hat{u}^{(b)}_t - \bar{y}_r)^T, \tag{S.161}$$

$$\hat{P}^{(b)}_t = \hat{V}^{(r)}_{t|\tau} + \hat{\chi}^{(b)}_t + \hat{x}^{(b)}_t (\hat{x}^{(b)}_t)^T, \tag{S.162}$$

$$\hat{P}^{(b)}_{t,t-1} = \hat{V}^{(r)}_{t,t-1|\tau} + \hat{\chi}^{(b)}_{t,t-1} + \hat{x}^{(b)}_t (\hat{x}^{(b)}_{t-1})^T, \tag{S.163}$$

$$\hat{W}^{(b)}_t = \hat{\kappa}^{(b)}_t + (\hat{u}^{(b)}_t - \bar{y}_r)(\hat{x}^{(b)}_t)^T. \tag{S.164}$$

6: Compute expected log-likelihood ℓ:

$$\begin{aligned}\ell = & \sum_{t=1}^{\tau} -\frac{1}{2}\mathrm{tr}\left((C_r \hat{V}^{(r)}_{t|t-1} C_r^T + R_r)^{-1} (\hat{U}^{(b)}_t - \hat{\lambda}^{(b)}_t C_r^T - C_r (\hat{\lambda}^{(b)}_t)^T + C_r \hat{\Lambda}^{(b)}_t C_r^T)\right) \\ & - \frac{1}{2}\log\left|C_r \hat{V}^{(r)}_{t|t-1} C_r^T + R_r\right| - \frac{m}{2}\log(2\pi),\end{aligned} \tag{S.165}$$

where $\hat{\Lambda}^{(b)}_t = \hat{\mathbf{V}}^{[3,3]}_t + \hat{\mathbf{x}}^{[3]}_t (\hat{\mathbf{x}}^{[3]}_t)^T$, and $\hat{\lambda}^{(b)}_t = C_b \hat{\mathbf{V}}^{[2,3]}_t + (C_b \hat{\mathbf{x}}^{[1]}_t + \bar{y}_b - \bar{y}_r)(\hat{\mathbf{x}}^{[3]}_t)^T$.

7: **Output**: $\{\hat{x}^{(b)}_t, \hat{P}^{(b)}_t, \hat{P}^{(b)}_{t,t-1}, \hat{W}^{(b)}_t, \hat{U}^{(b)}_t, \hat{u}^{(b)}_t\}, \ell$.

References

[1] A. B. Chan, E. Coviello, and G. Lanckriet, "Clustering dynamic textures with the hierarchical EM algorithm," in *Intl. Conference on Computer Vision and Pattern Recognition*, 2010.

[2] N. Vasconcelos and A. Lippman, "Learning mixture hierarchies," in *Neural Information Processing Systems*, 1998.

[3] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE TPAMI*, vol. 30, no. 5, pp. 909–926, May 2008.

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[5] A. Gelb, *Applied Optimal Estimation*. MIT Press, 1974.

[6] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.

Algorithm 2 Kalman smoothing filter

Input: DT parameters $\Theta = \{A, C, Q, R, \mu, S, \bar{y}\}$, length τ.
Initialize: $\hat{x}_{1|0} = \mu_b,\ \hat{V}_{1|0} = S_b$.
for $t = \{1, \cdots, \tau\}$ **do**
    $\{$Kalman filter – forward recursion$\}$

$$\begin{array}{l} \hat{V}_{t|t-1} = A\hat{V}_{t-1|t-1}A^T + Q, \\ K_t = \hat{V}_{t|t-1}C^T(C\hat{V}_{t|t-1}C^T + R)^{-1}, \\ \hat{V}_{t|t} = (I - K_tC)\hat{V}_{t|t-1}, \\ G_t = AK_t, \\ F_t = A - AK_tC. \end{array} \tag{S.166}$$

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

$$\begin{array}{l} J_{t-1} = \hat{V}_{t-1|t-1}A^T(\hat{V}_{t|t-1})^{-1}, \\ H_{t-1} = A^{-1} - J_{t-1}, \\ \hat{V}_{t-1|\tau} = \hat{V}_{t-1|t-1} + J_{t-1}(\hat{V}_{t|\tau} - \hat{V}_{t|t-1})J_{t-1}^T, \\ \hat{V}_{t-1,t-2|\tau} = \hat{V}_{t-1|t-1}J_{t-2}^T + J_{t-1}(\hat{V}_{t,t-1|\tau} - A\hat{V}_{t-1|t-1})J_{t-2}^T. \end{array} \tag{S.167}$$

end for
Output: Kalman filter matrices $\{\hat{V}_{t|t-1}, \hat{V}_{t|t}, \hat{V}_{t|\tau}, \hat{V}_{t,t-1|\tau}, G_t, F_t, H_t\}$.

Algorithm 3 Sensitivity Analysis of Kalman filter

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.
Initialize: $\hat{\mathbf{x}}_1 = \begin{bmatrix} \mu_b \\ \mu_b \\ \mu_r \end{bmatrix}, \quad \hat{\mathbf{V}}_1 = \begin{bmatrix} S_b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
for $t = \{2, \cdots, \tau + 1\}$ **do**
    Form block matrices:

$$\mathbf{A}_{t-1} = \begin{bmatrix} A_b & 0 & 0 \\ G_{t-1}^{(b)}C_b & F_{t-1}^{(b)} & 0 \\ G_{t-1}^{(r)}C_b & 0 & F_{t-1}^{(r)} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{C}_{t-1} = \begin{bmatrix} 0 \\ G_{t-1}^{(b)} \\ G_{t-1}^{(r)} \end{bmatrix}, \ \mathbf{D}_{t-1} = \begin{bmatrix} 0 \\ 0 \\ G_{t-1}^{(r)} \end{bmatrix}. \tag{S.168}$$

    Update means and covariances:

$$\hat{\mathbf{x}}_t = \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{D}_{t-1}(\bar{y}_b - \bar{y}_r), \tag{S.169}$$

$$\hat{\mathbf{V}}_t = \mathbf{A}_{t-1}\hat{\mathbf{V}}_{t-1}\mathbf{A}_{t-1}^T + \mathbf{B}Q_b\mathbf{B}^T + \mathbf{C}_{t-1}R_b\mathbf{C}_{t-1}^T. \tag{S.170}$$

end for
Output: $\hat{\mathbf{x}}_t = \mathbb{E}_{\Theta_b}\begin{bmatrix} x_t^{(b)} \\ \hat{x}_{t|t-1}^{(b)} \\ \hat{x}_{t|t-1}^{(r)} \end{bmatrix}, \hat{\mathbf{V}}_t = \mathrm{cov}_{\Theta_b}(\begin{bmatrix} x_t^{(b)} \\ \hat{x}_{t|t-1}^{(b)} \\ \hat{x}_{t|t-1}^{(r)} \end{bmatrix})$.

Algorithm 4 Sensitivity Analysis of Kalman smoothing filter (non-recursive)

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.

Initialize: $\hat{x}_\tau^{(b)} = A_r^{-1}\hat{\mathbf{x}}_{\tau+1}^{[3]}$, $\hat{\chi}_\tau^{(b)} = A_r^{-1}\hat{\mathbf{V}}_{\tau+1}^{[3,3]}A_r^{-T}$, $L_\tau = A_r^{-1}$, $M_\tau = \mathbf{0}$.

for $t = \{\tau, \cdots, 1\}$ **do**

Compute cross-covariance:

$$\rho_t = \left(L_t F_t^{(r)}\hat{\mathbf{V}}_t^{[3,2]} + (L_t G_t^{(r)} C_b + M_t)\hat{\mathbf{V}}_t^{[1,1]}\right) C_b^T + L_t G_t^{(r)} R_b, \tag{S.171}$$

$$\hat{\kappa}_t^{(b)} = \rho_t^T. \tag{S.172}$$

if $t > 1$ **then**

Compute sensitivity:

$$\omega_t = L_t F_t^{(r)}\hat{\mathbf{V}}_t^{[3,3]} + (L_t G_t^{(r)} C_b + M_t)\hat{\mathbf{V}}_t^{[2,3]}, \tag{S.173}$$

$$\hat{x}_{t-1}^{(b)} = H_{t-1}^{(r)}\hat{\mathbf{x}}_t^{[3]} + J_{t-1}^{(r)}\hat{x}_t^{(b)}, \tag{S.174}$$

$$\hat{\chi}_{t-1}^{(b)} = \begin{bmatrix} H_{t-1}^{(r)} & J_{t-1}^{(r)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_t^{[3,3]} & \omega_t^T \\ \omega_t & \hat{\chi}_t^{(b)} \end{bmatrix} \begin{bmatrix} (H_{t-1}^{(r)})^T \\ (J_{t-1}^{(r)})^T \end{bmatrix}, \tag{S.175}$$

$$\hat{\chi}_{t,t-1}^{(b)} = \omega_t (H_{t-1}^{(r)})^T + \hat{\chi}_t^{(b)} (J_{t-1}^{(r)})^T. \tag{S.176}$$

Update matrices:

$$L_{t-1} = H_{t-1}^{(r)} + J_{t-1}^{(r)} L_t F_t^{(r)}, \tag{S.177}$$

$$M_{t-1} = J_{t-1}^{(r)} (L_t G_t^{(r)} C_b + M_t) A_b. \tag{S.178}$$

end if

end for

Output: $\{\hat{x}_t^{(b)}, \hat{\chi}_t^{(b)}, \hat{\chi}_{t,t-1}^{(b)}, \hat{\kappa}_t^{(b)}\}$.