Kalman Filtering and Smoothing Calculations

Using the preferred timing for the state space model,

$$\begin{aligned} y_t &= Z_t\alpha_t + d_t + \epsilon_t \qquad & \epsilon_t \sim N(0, H_t) \\ \alpha_t &= T_t\alpha_{t-1} + c_t + R_t\eta_t \qquad & \eta_t \sim N(0, Q_t) \end{aligned}$$

the filtered estimates are given by

$$\begin{aligned} a_{t+1} &= T_{t+1}a_t + c_{t+1} + K_t v_t \qquad & P_{t+1} &= T_{t+1}P_tL_t' + R_{t+1}Q_{t+1}R_{t+1}' \\ v_t &= y_t - Z_ta_t - d_t \qquad & K_t &= T_{t+1}P_tZ_t'F_t^{-1} \\ F_t &= Z_tP_tZ_t' + H_t \qquad & L_t &= T_{t+1} - K_tZ_t \end{aligned}$$

and the smoothed estimates are given by

$$\begin{aligned} \hat{\alpha}_t &= a_t + P_tr_t \qquad & V_t &= P_t - P_tN_tP_t \\ r_t &= Z_t'F_t^{-1}v_t + L_t'r_{t+1} \qquad & N_t &= Z_t'F_t^{-1}Z_t + L_t'N_{t+1}L_t \end{aligned}$$

where $r_{n+1} = 0$ and $N_{n+1} = 0$.

Univariate Filter

When any H_t is non-diagonal, the observation equation is transformed by taking the LDL factorization of the H_t matrices:

$$y_t^* = Z_t^*\alpha_t + d_t^* + \epsilon_t^* \qquad \epsilon_t \sim N(0, H_t^*)$$

$$y_t^* = C_t^{-1}y_t \qquad Z_t^* = C_t^{-1}Z_t \qquad d_t^* = C_t^{-1}d_t \qquad \epsilon_t^* = C_t^{-1}\epsilon_t \qquad H_t = C_tH_t^*C_t'$$

To avoid taking the inverse of F_t in the above recursion, several quantities are computed using the univariate filter that can be used on this transformed system:

$$a_{t,i+1} = a_{t,i} + K_{t,i}v_{t,i}^* \qquad P_{t,i+1} = P_{t,i} - K_{t,i}F_{t,i}K_{t,i}'$$

$$v_{t,i}^* = y_{t,i}^* - Z_{t,i}^*a_{t,i} - d_{t,i}^* \qquad F_{t,i} = Z_{t,i}^*P_{t,i}Z_{t,i}^{*\prime} + H_{t,i}^* \qquad K_{t,i} = P_{t,i}Z_{t,i}^{*\prime}F_{t,i}^{-1}$$

$$a_{t+1,1} = T_{t+1}a_{t,p+1} + c_{t+1} \qquad P_{t+1,1} = T_{t+1}P_{t,p+1}T_{t+1}' + R_{t+1}Q_{t+1}R_{t+1}'$$

where $Z_{t,i}^*$ is the ith row of Z_t^*, $d_{t,i}^*$ is the ith element of d_t^*, and $H_{t,i}^*$ is the ith diagonal element of H_t^*. The filtered estimates of the state $a_t = a_{t,1}$ and $P_t = P_{t,1}$ are equivalent to those computed above. Similarly, the univariate smoother provides $r_t = r_{t,0}$ and $N_t = N_{t,0}$:

$$r_{t,i-1} = Z_{t,i}^{*\prime}F_{t,i}^{-1}v_{t,i}^* + L_{t,i}'r_{t,i} \qquad N_{t,i-1} = Z_{t,i}^{*\prime}F_{t,i}^{-1}Z_{t,i}^* + L_{t,i}'N_{t,i}L_{t,i} \qquad L_{t,i} = I_m - K_{t,i}Z_{t,i}^*F_{t,i}^{-1}$$

$$r_{t-1,p} = T_t'r_{t,0} \qquad N_{t-1,p} = T_{t-1}'N_{t,0}T_{t-1}$$

where $r_{t+1,p} = 0$ and $N_{t+1,p} = 0$.

Exact Initial Kalman Filter

When the state α_t is stationary, the initial values a_0 and P_0 can be computed as the unconditional mean and variance of the state given the system parameters:

$$\begin{aligned} a_0 &= (I_m - T_0)^{-1} c_0 \\ \text{vec}(P_0) &= (I_{m^2} - T_0 \otimes T_0)^{-1} \text{vec}(R_0 Q_0 R_0') \end{aligned}$$

In cases where the state α_t is nonstationary, the state is separated into stationary components and nonstationary components via the eigenvalues of the T_0 matrix – states associated with an eigenvalue less than one in absolute value are considered stationary while those with an eigenvalue greater than one are diffuse. The initial values are given by

$$\begin{aligned} a_0 &= a + R_0\eta_0 + A\delta \qquad \eta_0 \sim N(0, Q_0) \qquad \delta \sim N(0, \kappa I) \\ P_0 &= P_{*,0} + \kappa P_{\infty,0} \qquad P_{*,0} = R_0 Q_0 R_0' \qquad P_{\infty,0} = AA' \end{aligned}$$

The unconditional mean of the stationary states is computed as a_0 was above and placed in a with any elements of a associated with diffuse states set to 0. The unconditional variance of the stationary states, Q_0, is computed as P_0 was in the stationary case. The selection matrix R_0 is composed of columns of the identity matrix such that the initial shock η_0 is applied to the stationary states. The selection matrix A is composed of the columns of the identity matrix associated with the diffuse states such that taking the limit as $\kappa \to \infty$ allows for states with infinite initial variance.

The univariate filter recursions must be altered to separate the states with finite v. infinite variances:

$$\begin{aligned} F_{*,t,i} &= Z_{t,i}^* P_{*,t,i} Z_{t,i}^{*\prime} + H_{t,i}^* \qquad F_{\infty,t,i} = Z_{t,i}^* P_{\infty,t,i} Z_{t,i}^{*\prime} \\ K_{*,t,i} &= P_{*,t,i} Z_{t,i}^{*\prime} \qquad K_{\infty,t,i} = P_{\infty,t,i} Z_{t,i}^{*\prime} \end{aligned}$$

$$a_{t,i+1} = \begin{cases} a_{t,i} + K_{*,t,i} F_{*,t,i}^{-1} v_{t,i}^* & F_{\infty,t,i} = 0 \\ a_{t,i} + K_{\infty,t,i} F_{\infty,t,i}^{-1} v_{t,i}^* & F_{\infty,t,i} \neq 0 \end{cases}$$

$$P_{*,t,i+1} = \begin{cases} P_{*,t,i} - K_{*,t,i} K_{*,t,i}' F_{*,t,i}^{-1} & F_{\infty,t,i} = 0 \\ P_{*,t,i} + K_{\infty,t,i} K_{\infty,t,i}' F_{*,t,i} F_{\infty,t,i}^{-2} - (K_{*,t,i} K_{\infty,t,i}' + K_{\infty,t,i} K_{*,t,i}') F_{\infty,t,i}^{-1} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$P_{\infty,t,i+1} = \begin{cases} P_{\infty,t,i} & F_{\infty,t,i} = 0 \\ P_{\infty,t,i} - K_{\infty,t,i} K_{\infty,t,i}' F_{\infty,t,i}^{-1} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$\begin{aligned} a_{t+1,1} &= T_{t+1} a_{t,p+1} + c_{t+1} \\ P_{\infty,t+1,1} &= T_{t+1} P_{\infty,t,p+1} T_{t+1}' \qquad P_{*,t+1,1} = T_{t+1} P_{*,t,p+1} T_{t+1}' + R_{t+1} Q_{t+1} R_{t+1}' \end{aligned}$$

For any set of system parameters where the state can be identified, there exists some time d such that $F_{\infty,d,i} = 0$ for all i. For time $t > d$, the simpler Kalman filter recursion above can be employed.

The smoother must be similarly altered so that beginning at $t = d$, the computation of $r_{t,i}$ is expanded to account for the initialization:

$$L_{\infty,t,i} = I_m - K_{\infty,t,i} Z^*_{t,i} F^{-1}_{\infty,t,i} \qquad L_{*,t,i} = I_m - K_{*,t,i} Z^*_{t,i} F^{-1}_{*,t,i}$$

$$L^{(0)}_{t,i} = \left(K_{\infty,t,i} F_{*,t,i} F^{-1}_{\infty,t,i} - K_{*,t,i}\right) Z^*_{t,i} F^{-1}_{\infty,t,i}$$

$$r^{(0)}_{t,i-1} = \begin{cases} Z^{*\prime}_{t,i} F^{-1}_{*,t,i} v^*_{t,i} + L'_{*,t,i} r^{(0)}_{t,i} & F_{\infty,t,i} = 0 \\ L'_{\infty,t,i} r^{(0)}_{t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$r^{(1)}_{t,i-1} = \begin{cases} r^{(1)}_{t,i} & F_{\infty,t,i} = 0 \\ Z^{*\prime}_{t,i} F^{-1}_{\infty,t,i} v^*_{t,i} + L^{(0)\prime}_{t,i} r^{(0)}_{t,i} + L'_{\infty,t,i} r^{(1)}_{t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$r^{(0)}_{t-1,p} = T'_t r^{(0)}_{t,0} \qquad r^{(1)}_{t-1,p} = T'_t r^{(1)}_{t,0}$$

$$\hat{\alpha}_t = a_t + P_{*,t,1} r^{(0)}_{t,0} + P_{\infty,t,1} r^{(1)}_{t,0}$$

where $r^{(0)}_{d,p} = r_{d,p}$ and $r^{(1)}_{d,p} = 0$. Note that $L_{\infty,t,i}$ and $L^{(0)}_{t,i}$ only need to be computed when $F_{\infty,t,i} \neq 0$ and $L_{*,t,i}$ only needs to be computed when $F_{\infty,t,i} = 0$.

For the smoothed variance of the state,

$$V_t = P_{*,t,1} - P_{*,t,1} N^{(0)}_{t,0} P_{*,t,1} - \left(P_{\infty,t,1} N^{(1)}_{t,0} P_{*,t,1}\right)' - P_{\infty,t,1} N^{(1)}_{t,0} P_{*,t,1} - P_{\infty,t,1} N^{(2)}_{t,0} P_{\infty,t,1}$$

$$N^{(0)}_{t-1,p} = T'_t N^{(0)}_{t,0} T_t \qquad N^{(1)}_{t-1,p} = T'_t N^{(1)}_{t,0} T_t \qquad N^{(2)}_{t-1,p} = T'_t N^{(1)}_{t,0} T_t$$

where when $F_{\infty,t,i} = 0$,

$$N^{(0)}_{t,i-1} = Z^{*\prime}_{t,i} F^{-1}_{*,t,i} Z^*_{t,i} + L'_{*,t,i} N^{(0)}_{t,i} L_{*,t,i} \qquad N^{(1)}_{t,i-1} = N^{(1)}_{t,i} \qquad N^{(2)}_{t,i-1} = N^{(2)}_{t,i}$$

and when $F_{\infty,t,i} \neq 0$,

$$N^{(0)}_{t,i-1} = L'_{\infty,t,i} N^{(0)}_{t,i} L_{\infty,t,i}$$
$$N^{(1)}_{t,i-1} = Z^{*\prime}_{t,i} F^{-1}_{\infty,t,i} Z^*_{t,i} + L'_{\infty,t,i} N^{(0)}_{t,i} L^{(0)}_{t,i} + L'_{\infty,t,i} N^{(1)}_{t,i} L_{\infty,t,i}$$
$$N^{(2)}_{t,i-1} = Z^{*\prime}_{t,i} F^{-2}_{\infty,t,i} Z^*_{t,i} F_{*,t,i} + L^{(0)\prime}_{t,i} N^{(1)}_{t,i} L^{(0)}_{t,i} + L'_{\infty,t,i} N^{(1)}_{t,i} L^{(0)}_{t,i} + L^{(0)\prime}_{t,i} N^{(1)}_{t,i} L_{\infty,t,i} + L'_{\infty,t,i} N^{(2)}_{t,i} L_{\infty,t,i}$$

where $N^{(0)}_{d,p} = N_{d,p}$ and $N^{(1)}_{d,p} = N^{(2)}_{d,p} = 0$.

Likelihood Calculation

The likelihood of of data $y_1, \ldots, y_n$ is given by

$$\log L(Y_n) \equiv -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log|F_t| + v_t' F_t^{-1} v_t$$

Univariate Gradient

From Jungbacker, Koopman & van der Wel (2011), the gradient of the likelihood can be computed without taking the inverse of F_t provided the smoothed estimates are computed. The gradient of the likelihood L_n with respect to each parameter matrix is (note, wrong timing, corrected for naming differences):

$$\begin{aligned}
\frac{\partial L_n}{\partial Z_t} &= H_t^{-1}\left[(y_t - d_t)\hat{\alpha}_t' - Z_t M_{Z_t}\right] \\
\frac{\partial L_n}{\partial d_t} &= H_t^{-1}(y_t - Z_t\hat{\alpha}_t - d_t) \\
\frac{\partial L_n}{\partial H_t} &= H_t^{-1} M_{H_t} H_t^{-1} - \frac{1}{2}\text{diag}\left\{H_t^{-1} M_{H_t} H_t^{-1}\right\} \\
\frac{\partial L_n}{\partial T_t} &= \bar{R}_t' Q_t^{-1}\bar{R}_t(M_{T_t} - T_t M_{Z_t}) \\
\frac{\partial L_n}{\partial c_t} &= \bar{R}_t' Q_t^{-1}\bar{R}_t(\hat{\alpha}_{t+1} - T_t\hat{\alpha}_t - c_t) \\
\frac{\partial L_n}{\partial Q_t} &= Q_t^{-1} M_{Q_t} Q_t^{-1} - \frac{1}{2}\text{diag}\left\{Q_t^{-1} M_{Q_t} Q_t^{-1}\right\}
\end{aligned}$$

where

$$\begin{aligned}
M_{Z_t} &= \hat{\alpha}_t\hat{\alpha}_t' + V_t \\
M_{H_t} &= (y_t - Z_t\hat{\alpha}_t)(y_t - Z_t\hat{\alpha}_t)' + Z_t V_t Z_t' - H_t \\
M_{T_t} &= \hat{\alpha}_t\hat{\alpha}_t' + J_t \\
M_{Q_t} &= \mathbb{E}(\eta_t\eta_t'|y_1, \ldots, y_n) - Q_t
\end{aligned}$$

Jungbacker & Koopman (2015), also have a formulation of the gradient but do so with time-invariant matricies without mean adjustments. Starting from the likelihood function

$$\begin{aligned} L_n = c - \frac{n}{2}\log|H| - \frac{1}{2}\mathrm{tr}M_H - \frac{n-1}{2}\log|Q| - \frac{1}{2}\mathrm{tr}M_Q \\ -\frac{1}{2}\log|P| - \frac{1}{2}\mathrm{tr}\left(P^{-1}((\hat{\alpha}_1 - a)(\hat{\alpha}_1 - a)' + V_1)\right) \end{aligned}$$

they produce parameter gradients (with a fair amount of guessing on my part for the timing)

$$\begin{aligned} \frac{\partial L_n}{\partial Z_t} &= H_t^{-1}\left[y_t\hat{\alpha}_t' - Z_t M_{Z_t}\right] \\ \frac{\partial L_n}{\partial H_t} &= H_t^{-1}M_{H_t}H_t^{-1} - \frac{1}{2}\mathrm{diag}\left\{H_t^{-1}M_{H_t}H_t^{-1}\right\} \\ \frac{\partial L_n}{\partial T_t} &= Q_t^{-1}(M_{T_t} - T_t M_{Z_t}) \qquad \text{(or is this } M_{Z_{t-1}}\text{?)} \\ \frac{\partial L_n}{\partial Q_t} &= Q_t^{-1}M_{Q_t}Q_t^{-1} - \frac{1}{2}\mathrm{diag}\left\{Q_t^{-1}M_{Q_t}Q_t^{-1}\right\} \end{aligned}$$

where

$$\begin{aligned} M_{Z_t} &= \hat{\alpha}_t\hat{\alpha}_t' + V_t \\ M_{H_t} &= \hat{\epsilon}_t\hat{\epsilon}_t' + \mathrm{Var}(\epsilon_t|Y_n,\theta) \\ M_{T_t} &= \hat{\alpha}_t\hat{\alpha}_t' + J_t \\ M_{Q_t} &= \hat{\eta}_t\hat{\eta}_t' + \mathrm{Var}(\eta_t|Y_n,\theta) \end{aligned}$$

Multivariate Gradient

From Nagakura (working paper), we can get the gradient:

$$\begin{aligned} G_\theta(\ell_t) = &\, G_\theta(a_t) Z_t' w_t \\ &+ \frac{1}{2} G_\theta(P_t) \mathrm{vec}(Z_t' w_t w_t' Z_t - Z_t' F_t^{-1} Z_t) \\ &+ G_\theta(d_t) w_t \\ &+ G_\theta(Z_t) \mathrm{vec}(w_t a_t' + w_t v_t' M_t' - M_t') \\ &+ \frac{1}{2} G_\theta(H_t) \mathrm{vec}(w_t w_t' - F_t^{-1}) \end{aligned}$$

where

$$\begin{aligned} G_\theta(a_{t+1}) = &\, G_\theta(a_t) L_t' \\ &+ G_\theta(P_t)(Z_t' w_t \otimes L_t') \\ &+ G_\theta(c_{t+1}) \\ &- G_\theta(d_t) K_t' \\ &+ G_\theta(Z_t)[P_t L_t' \otimes w_t - (a_t + M_t v_t) \otimes K_t'] \\ &- G_\theta(H_t)(w_t \otimes K_t') \\ &+ G_\theta(T_{t+1})[(a_t + M_t v_t) \otimes I_m] \end{aligned}$$

$$\begin{aligned} G_\theta(P_{t+1}) = &\, G_\theta(P_t)(L_t' \otimes L_t') \\ &+ G_\theta(H_t)(K_t' \otimes K_t') \\ &+ G_\theta(Q_{t+1})(R_{t+1}' \otimes R_{t+1}') \\ &+ [G_\theta(T_{t+1})(P_t L_t' \otimes I_m) \\ &\quad - G_\theta(Z_t)(P_t L_t' \otimes K_t') \\ &\quad + G_\theta(R_{t+1})(Q_{t+1} R_{t+1}' \otimes I_m)] N_m \end{aligned}$$

The initial conditions for the recursion are given by a simplification of the expressions above (which can also be easily derived from the explicit expressions for a_1 and P_1):

$$\begin{aligned} G_\theta(a_1) = &\, G_\theta(a_0) T_1' \\ &+ G_\theta(c_1) \\ &+ G_\theta(T_1)[a_0 \otimes I_m] \end{aligned}$$

$$\begin{aligned} G_\theta(P_1) = &\, G_\theta(P_0)(T_t' \otimes T_t') \\ &+ G_\theta(Q_1)(R_1' \otimes R_1') \\ &+ [G_\theta(T_1)(P_0 T_1' \otimes I_m) \\ &\quad + G_\theta(R_1)(Q_1 R_1' \otimes I_m)] N_m \end{aligned}$$

To determine $G_\theta(a_0)$ and $G_\theta(P_0)$ when a_0 and P_0 are set as the unconditional mean and variance of the state (i.e., when they are not explicitly provided and the system is stationary) use the definitions of the unconditional state:

$$
\begin{aligned}
a_0 &= (I_m - T)^{-1} c \\
G_\theta(a_0) &= G_\theta([I_m - T]^{-1} c) \\
&= G_\theta([I_m - T]^{-1})(c \otimes I_m) + G_\theta(c)(I_1 \otimes [(I_m - T)^{-1})]' \\
&= -G_\theta(I_m - T)[(I_m - T)^{-1} \otimes (I_m - T)^{-1}](c \otimes I_m) + G_\theta(c)(I_m - T)'^{-1} \\
&= G_\theta(T)[(I_m - T)^{-1} \otimes (I_m - T)^{-1}](c \otimes I_m) + G_\theta(c)(I_m - T)'^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\text{vec}(P_0) &= (I_{m^2} - T \otimes T)^{-1} \text{vec}(RQR') \\
G_\theta(P_0) &= G_\theta(\text{vec}(P_0)) \\
&= G_\theta(S \ \text{vec}(RQR')) \\
&= G_\theta(S)[\text{vec}(RQR') \otimes I_{m^2}] + G_\theta(\text{vec}(RQR'))(I_1 \otimes S') \\
&= -G_\theta(I_{m^2} - T \otimes T)(S \otimes S')[\text{vec}(RQR') \otimes I_{m^2}] \\
&\qquad + [G_\theta(R)(QR' \otimes I_m)N_m + G_\theta(Q)(R' \otimes R')]S' \\
&= G_\theta(T \otimes T)(S \otimes S')[\text{vec}(RQR') \otimes I_{m^2}] \\
&\qquad + [G_\theta(R)(QR' \otimes I_m)N_m + G_\theta(Q)(R' \otimes R')]S'
\end{aligned}
$$

where $S = (I_{m^2} - T \otimes T)^{-1}$. Note that $G_\theta(T \otimes T)$ must be computed separately.