Rao-Blackwellised Particle Smoothing: Incorporating suggested improvements from Fredrik Lindsten and Gerlach et al.

Pete Bunch

25th September 2012

1 Basics

Our model is,

$$\begin{aligned} u_t &\sim p(u_t|u_{t-1}) \\ z_t &= A(u_t)z_{t-1} + q_t \\ y_t &= C(u_t)z_t + r_t \\ q_t &\sim \mathcal{N}(\cdot|0, Q(u_t)) \\ r_t &\sim \mathcal{N}(\cdot|0, R(u_t)) \end{aligned} \tag{1}$$

This is the more standard model, not the one used in our conference paper. With smoothing, we would like to estimate

$$p(z_{1:T}, u_{1:T}|y_{1:T}) = p(z_{1:T}|u_{1:T}, y_{1:T})p(u_{1:T}|y_{1:T}), \tag{2}$$

where $z_{1:T}$ and $u_{1:T}$ are the linear and nonlinear parts of the state respectively.

2 Backward Simulation

The nonlinear smoothing distribution can be factorised in the following way,

$$p(u_{1:T}|y_{1:T}) = p(u_T|y_{1:T}) \prod_{t=1}^{T-1} p(u_t|u_{t+1:T}, y_{1:T}), \tag{3}$$

which suggests a backwards sequential sampling approach. The factors may be expanded as,

$$
\begin{aligned}
p(u_t|u_{t+1:T}, y_{1:T}) &= \int p(u_{1:t}, z_t|u_{t+1:T}, y_{1:T}) dz_t du_{1:t-1} \\
&\propto \int p(y_{t+1:T}|z_t, u_{t+1:T}) p(u_{1:t}, z_t|u_{t+1:T}, y_{1:t}) dz_t du_{1:t-1} \\
&\propto \int p(y_{t+1:T}|z_t, u_{t+1:T}) p(u_{t+1:T}|u_t) p(z_t|u_{1:t}, y_{1:t}) p(u_{1:t}|y_{1:t}) dz_t du_{1:t-1} \\
&\propto \int \left[\int p(z_t|u_{1:t}, y_{1:t}) p(y_{t+1:T}|z_t, u_{t+1:T}) dz_t \right] p(u_{1:t}|y_{1:t}) p(u_{t+1}|u_t) du_{1:t-1}. \quad (4)
\end{aligned}
$$

Running a Rao-Blackwellised particle filter gives us an estimate of,

$$
p(u_{1:t}|y_{1:t}) \approx \sum_i w_t^{(i)} \delta_{u_{1:t}^{(i)}}(u_{1:t}). \tag{5}
$$

Substituting this in, particles may be drawn from the smoothing distribution by sequentially sampling,

$$
\begin{aligned}
p(u_t|u_{t+1:T}, y_{1:T}) &\approx \int \sum_i w_t^{(i)} \delta_{u_{1:t}^{(i)}}(u_{1:t}) p(u_{t+1}|u_t) \int p(z_t|u_{1:t}, y_{1:t}) p(y_{t+1:T}|z_t, u_{t+1:T}) dz_t du_{1:t-1} \\
&= \sum_i w_t^{(i)} p(u_{t+1}|u_t^{(i)}) \int p(z_t|u_{1:t}^{(i)}, y_{1:t}) p(y_{t+1:T}|z_t, u_{t+1:T}) dz_t \, \delta_{u_t^{(i)}}(u_t) \\
&= \sum_i \rho^{(i)} \delta_{u_t^{(i)}}(u_t), \qquad (6)
\end{aligned}
$$

where,

$$
\rho^{(i)} \propto w_t^{(i)} p(u_{t+1}|u_t^{(i)}) \int p(z_t|u_{1:t}^{(i)}, y_{1:t}) p(y_{t+1:T}|z_t, u_{t+1:T}) dz_t. \tag{7}
$$

3 The Linear State Integral

3.1 Original Formulation

The first term in the integral is given by the forward Kalman filter estimate from the forward stage

$$
p(z_t|u_{1:t}^{(i)}, y_{1:t}) = \mathcal{N}(z_t|m_t^{(i)}, P_t^{(i)}). \tag{8}
$$

The second term is given by a backward filter in the style of the two-filter smoother,

$$
p(y_{t+1:T}|z_t, u_{t+1:T}) = \hat{Z}_t \mathcal{N}(z_t|\hat{m}_t^b, \hat{P}_t^b). \tag{9}
$$

The integral is then given by,

$$
\begin{aligned}
&\int p(z_t|u_{1:t}^{(i)}, y_{1:t}) p(y_{t+1:T}|z_t, u_{t+1:T}) dz_t \\
&\qquad = \hat{Z}_t \int \mathcal{N}(z_t|m_t^{(i)}, P_t^{(i)}) \mathcal{N}(z_t|\hat{m}_t^b, \hat{P}_t^b) dz_t \\
&\qquad = \hat{Z}_t \mathcal{N}(m_t^{(i)}|\hat{m}_t^b, P_t^{(i)} + \hat{P}_t^b). \qquad (10)
\end{aligned}
$$

The normalisation factor $\hat{Z}_t$ is constant over the sampling index i, and may thus be forgotten about (comes out in the normalisation) as long as it is finite. This will be the case apart from during the initialisation and when $A(u_t)$ is not invertible.

It only remains to calculate the backward filter estimate. These are given by a recursion,

$$p(y_{t:T}|z_t, u_{t:T}) = Z_t \mathcal{N}(z_t|m_t^b, P_t^b) \tag{11}$$
$$p(y_{t+1:T}|z_t, u_{t+1:T}) = \hat{Z}_t \mathcal{N}(z_t|\hat{m}_t^b, \hat{P}_t^b). \tag{12}$$

These are standard. We use the shorthand $A_t = A(u_t)$, $Q_t = Q(u_t)$, etc.

Predict:

$$\begin{aligned}
&p(y_{t+1:T}|z_t, u_{t+1:T}) \\
&\qquad = \int p(y_{t+1:T}|z_{t+1}, u_{t+1:T}) p(z_{t+1}|u_{t+1}, z_t) dz_{t+1} \\
&\qquad = Z_{t+1} \int \mathcal{N}(z_{t+1}|m_{t+1}^b, P_{t+1}^b) \mathcal{N}(z_{t+1}|A_t z_t, Q_{t+1}) \\
&\qquad = \hat{Z}_t \mathcal{N}(z_t|\hat{m}_t^b, \hat{P}_t^b)
\end{aligned} \tag{13}$$

Thus:

$$\hat{m}_t^b = A_t^{-1} m_{t+1}^b \tag{14}$$
$$\hat{P}_t^b = A_t^{-1}(P_{t+1}^b + Q_{t+1}) A_t^{-T} \tag{15}$$
$$\hat{Z}_t = Z_{t+1} \, ||A_t||^{-1} \tag{16}$$

Update:

$$\begin{aligned}
&p(y_{t:T}|z_t, u_{t:T}) \\
&\qquad = p(y_{t+1:T}|z_t, u_{t+1:T}) p(y_t|z_t, u_t) \\
&\qquad = \hat{Z}_t \mathcal{N}(z_t|\hat{m}_t^b, \hat{P}_t^b) \mathcal{N}(y_t|C_t z_t, R_t) \\
&\qquad = Z_t \mathcal{N}(z_t|m_t^b, P_t^b)
\end{aligned} \tag{17}$$

where:

$$S_t = C_t \hat{P}_t^b C_t^T + R_t \tag{18}$$
$$K_t = \hat{P}_t^b C_t^T S_t^{-1} \tag{19}$$
$$m_t^b = \hat{m}_t^b + K_t(y_t - C_t \hat{m}_t^b) \tag{20}$$
$$P_t^b = \hat{S}_t - K_t S_t K_t^T \tag{21}$$
$$Z_t = \hat{Z}_t \mathcal{N}(y_t|C_t \hat{m}_t, S_t) \tag{22}$$

This formulation has the disadvantage that it fails when A_t is not invertible. Also the initialisation is messy.

3.2 Better Formulation

We use the same Kalman filter for the forward estimate

$$p(z_t|u^{(i)}_{1:t}, y_{1:t}) = \mathcal{N}(z_t|m^{(i)}_t, P^{(i)}_t). \tag{23}$$

For the backward estimate, assume the following form,

$$p(y_{t+1:T}|z_t, u_{t+1:T}) = \hat{Z}_t \exp\left\{-\frac{1}{2}\left[z_t^T\hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t\right]\right\}. \tag{24}$$

We will need the following identity (repeatedly),

$$\int \exp\left\{-\frac{1}{2}\left[z^T\Upsilon z - 2\psi^T z + c\right]\right\} dz = \sqrt{||\Upsilon^{-1}||}\exp\left\{-\frac{1}{2}\left[c - \psi^T\Upsilon^{-1}\psi\right]\right\} \tag{25}$$

This is a simpler form of Fredrik's Lemma 1.

The integral is given by,

$$\begin{aligned}
&\int p(z_t|u^{(i)}_{1:t}, y_{1:t})p(y_{t+1:T}|z_t, u_{t+1:T})dz_t \\
&\qquad = \hat{Z}_t \int \mathcal{N}(z_t|m^{(i)}_t, P^{(i)}_t)\exp\left\{-\frac{1}{2}\left[z_t^T\hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t\right]\right\} dz_t \\
&\qquad = \hat{Z}_t \frac{1}{\sqrt{\left|\left|2\pi P^{(i)}_t\right|\right|}}\int \exp\left\{-\frac{1}{2}\eta\right\} dz_t \qquad (26)\\
&\qquad = \hat{Z}_t \sqrt{\frac{\left|\left|\left(P^{(i)-1}_t + \hat{\Omega}_t\right)^{-1}\right|\right|}{\left|\left|P^{(i)}_t\right|\right|}}\exp\left\{-\frac{1}{2}\nu\right\}, \qquad (27)
\end{aligned}$$

where,

$$\begin{aligned}
\eta &= z_t^T\hat{\Omega}_t z_t - \hat{\lambda}_t^T z_t + (z_t - m^{(i)}_t)^T P^{(i)-1}_t (z_t - m^{(i)}_t) \\
&= z_t^T\left[\hat{\Omega}_t + P^{(i)-1}_t\right]z_t - 2\left[\hat{\lambda}_t^T + m^{(i)T}_t P^{(i)-1}_t\right]z_t + \left[m^{(i)T}_t P^{(i)-1}_t m^{(i)}_t\right], \qquad (28)
\end{aligned}$$

and

$$\begin{aligned}
\nu &= m^{(i)T}_t P^{(i)-1}_t m^{(i)}_t - \left[\hat{\lambda}_t^T + m^{(i)T}_t P^{(i)-1}_t\right]\left[\hat{\Omega}_t + P^{(i)-1}_t\right]^{-1}\left[\hat{\lambda}_t + P^{(i)-1}_t m^{(i)}_t\right] \\
&= m^{(i)T}_t\left[P^{(i)-1}_t - P^{(i)-1}_t\left(P^{(i)-1}_t + \hat{\Omega}_t\right)^{-1}P^{(i)-1}_t\right]m^{(i)}_t \\
&\quad - 2\left[\hat{\lambda}_t^T\left(P^{(i)-1}_t + \hat{\Omega}_t\right)^{-1}P^{(i)-1}_t\right]m^{(i)}_t \\
&\quad - \left[\hat{\lambda}_t^T\left(P^{(i)-1}_t + \hat{\Omega}_t\right)^{-1}\hat{\lambda}_t\right]. \qquad (29)
\end{aligned}$$

We could simplify this significantly using the Woodbury and matrix inverse identities if $\hat{\Omega}_t$ was invertible, but we don't want to assume that to be the case.

Now we need backwards recursions for the following,

$$p(y_{t:T}|z_t, u_{t:T}) = Z_t \exp\left\{-\frac{1}{2}\left[z_t^T \Omega_t z_t - 2\lambda_t^T z_t\right]\right\} \tag{30}$$

$$p(y_{t+1:T}|z_t, u_{t+1:T}) = \hat{Z}_t \exp\left\{-\frac{1}{2}\left[z_t^T \hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t\right]\right\}. \tag{31}$$

Assume we have estimates of Z_{t+1}, λ_{t+1} and Ω_{t+1}.
Predict:

$$\begin{aligned} p(y_{t+1:T}|z_t, u_{t+1:T}) &= \int p(y_{t+1:T}|z_{t+1}, u_{t+1:T}) p(z_{t+1}|z_t, u_{t+1}) dz_{t+1} \\ &= Z_{t+1} \frac{1}{\sqrt{||2\pi Q_{t+1}||}} \int \exp\left\{-\frac{1}{2}\zeta\right\} dz_{t+1} && (32) \\ &= Z_{t+1} \sqrt{\frac{\left|\left|\left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1}\right|\right|}{||Q_{t+1}||}} \exp\left\{-\frac{1}{2}\xi\right\} && (33) \end{aligned}$$

where

$$\begin{aligned} \zeta &= z_{t+1}^T \Omega_{t+1} z_{t+1} - 2\lambda_{t+1}^T z_{t+1} + (z_{t+1} - A_{t+1} z_t)^T Q_{t+1}^{-1} (z_{t+1} - A_{t+1} z_t) \\ &= z_{t+1}^T \left[\Omega_{t+1} + Q_{t+1}^{-1}\right] z_{t+1} - 2\left[\lambda_{t+1}^T + z_t^T A_{t+1}^T Q_{t+1}^{-1}\right] z_{t+1} + \left[z_t^T A_{t+1}^T Q_{t+1}^{-1} A_{t+1} z_t\right], && (34) \end{aligned}$$

and,

$$\begin{aligned} \xi &= z_t^T A_{t+1}^T Q_{t+1}^{-1} A_{t+1} z_t - \left[\lambda_{t+1}^T + z_t^T A_{t+1}^T Q_{t+1}^{-1}\right] \left[\Omega_{t+1} + Q_{t+1}^{-1}\right]^{-1} \left[\lambda_{t+1} + Q_{t+1}^{-1} A_{t+1} z_t\right] \\ &= z_t^T \left[A_{t+1}^T \left(Q_{t+1}^{-1} - Q_{t+1}^{-1} \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{t+1}^{-1}\right) A_{t+1}\right] z_t \\ &\quad - 2\left[\lambda_{t+1}^T \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{t+1}^{-1} A_{t+1}\right] z_t \\ &\quad - \left[\lambda_{t+1}^T \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1}\right]. && (35) \end{aligned}$$

Hence the prediction recursions are,

$$\hat{\Omega}_t = A_{t+1}^T \left(Q_{t+1}^{-1} - Q_{t+1}^{-1} \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{t+1}^{-1}\right) A_{t+1} \tag{36}$$

$$\hat{\lambda}_t = A_{t+1}^T Q_{t+1}^{-1} \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1} \tag{37}$$

$$\hat{Z}_t = Z_{t+1} \sqrt{\frac{\left|\left|\left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1}\right|\right|}{||Q_{t+1}||}} \exp\left\{-\frac{1}{2}\lambda_{t+1}^T \left(Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1}\right\} \tag{38}$$

Update:

$$\begin{aligned} p(y_{t:T}|z_t, u_{t:T}) &= p(y_{t+1:T}|z_t, u_{t+1:T}) p(y_t|z_t, u_t) \\ &= \hat{Z}_t \frac{1}{\sqrt{||2\pi R_t||}} \exp\left\{-\frac{1}{2}\omega\right\}, && (39) \end{aligned}$$

where,

$$\begin{aligned}\omega &= z_t^T \hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t + (y_t - C_t z_t)^T R_t^{-1} (y_t - C_t z_t) \\ &= z_t^T \left[\hat{\Omega}_t + C_t^T R_t^{-1} C_t\right] z_t - 2\left[\hat{\lambda}_t^T + y_t^T C_t^T R_t^{-1}\right] z_t + \left[y_t^T R_t^{-1} y_t\right].\end{aligned} \tag{40}$$

Hence the update recursions are,

$$\Omega_t = \hat{\Omega}_t + C_t^T R_t^{-1} C_t \tag{41}$$

$$\lambda_t = \hat{\lambda}_t + R_t^{-1} C_t y_t \tag{42}$$

$$Z_t = \hat{Z}_t \frac{1}{\sqrt{||2\pi R_t||}} \exp\left\{-\frac{1}{2} y_t^T R_t^{-1} y_t\right\}. \tag{43}$$

These recursions work even when A_t or Ω_{t+1} is not invertible, provided Q_{t+1} and R_t are full rank. Moreover they are easy to initialise,

$$\begin{aligned}p(y_T|z_T, u_T) &= \mathcal{N}(y_T|C_T z_T, R_T) \\ &= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2}(y_T - C_T z_T)^T R_T^{-1} (y_T - C_T z_T)\right\} \\ &= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2} y_T R_T^{-1} y_T\right\} \exp\left\{z_T^T C_T^T R_T^{-1} C_T z_T - 2(y_T^T R_T^{-1} C_T) z_T\right\}\end{aligned} \tag{44}$$

Hence,

$$\begin{aligned}\Omega_T &= C_T^T R_T^{-1} C_T \\ \lambda_T &= C_T^T R_T^{-1} y_T \\ Z_T &= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2} y_T R_T^{-1} y_T\right\}.\end{aligned} \tag{45}$$

Ω_T is not invertible, but this doesn't matter, unlike the original formulation.

Because the normalisation constants are always finite (for full rank Q_t and R_t) and depend only on the future nonlinear states, we need never calculate them.

4 Mixed Linear/Nonlinear Models

The extension to mixed models is straightforward. The new model is,

$$\begin{aligned}u_{t+1} &= A_u(u_t) z_t + q_{u,t} \\ z_{t+1} &= A_z(u_t) z_t + q_{z,t} \\ y_t &= C(u_t) z_t + r_t. \\ q_{u,t} &\sim \mathcal{N}(\cdot|0, Q_u(u_t)) \\ q_{z,t} &\sim \mathcal{N}(\cdot|0, Q_z(u_t)) \\ r_t &\sim \mathcal{N}(r_t|0, R(u_t)).\end{aligned} \tag{46}$$

We use equivalent short-hands for the matrices to those used previously. The thing to watch out for here is that u_{t+1} is not independent of the past observations when conditioned on u_t. Furthermore, in order to predict z_t from future observations, we will need u_t, unlike before. Finally, z_{t+1} will be less useful as an auxiliary variable, because it will not "block" the sequence of state variables (because u_{t+1} depends directly on z_t). The following derivations proceeds essentially as before.

Start with the backwards conditional distribution,

$$\begin{aligned}
p(u_t|u_{t+1:T}, y_{1:T}) &= \int p(u_{1:t}, z_t|u_{t+1:T}, y_{1:T}) dz_t du_{1:t-1} \\
&\propto \int p(y_{t+1:T}, u_{t+1:T}|z_t, u_t) p(u_{1:t}, z_t|y_{1:t}) dz_t du_{1:t-1} \\
&\propto \int \left[\int p(y_{t+1:T}, u_{t+1:T}|z_t, u_t) p(z_t|u_{1:t}, y_{1:t}) dz_t \right] p(u_{1:t}|y_{1:t}) du_{1:t-1} \quad (47)
\end{aligned}$$

This expression has the same form as the hierarchical model except for the exclusion of the nonlinear transition density and the modified backwards filter term. Hence, the weights have a very similar form to that derived in the previous section. We can use the particle filter and forward Kalman filter approximations as before. It only remains to estimate $p(y_{t+1:T}, u_{t+1:T}|z_t, u_t)$.

We need backwards recursions for the following,

$$p(y_{t:T}, u_{t+1:T}|z_t, u_t) = Z_t \exp\left\{ -\frac{1}{2} \left[z_t^T \Omega_t z_t - 2\lambda_t^T z_t \right] \right\} \quad (48)$$

$$p(y_{t+1:T}, u_{t+1:T}|z_t, u_t) = \hat{Z}_t \exp\left\{ -\frac{1}{2} \left[z_t^T \hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t \right] \right\}. \quad (49)$$

Assume we have estimates of Z_{t+1}, λ_{t+1} and Ω_{t+1}.

Predict:

$$\begin{aligned}
p(y_{t+1:T}, u_{t+1:T}|z_t, u_t) &= \int p(y_{t+1:T}, u_{t+2:T}|z_{t+1}, u_{t+1}) p(z_{t+1}, u_{t+1}|z_t, u_t) dz_{t+1} \\
&= \int p(y_{t+1:T}, u_{t+2:T}|z_{t+1}, u_{t+1}) p(z_{t+1}|z_t, u_t), p(u_{t+1}|z_t, u_t) dz_{t+1} \\
&= Z_{t+1} \frac{1}{\sqrt{||2\pi Q_{z,t+1}||\, ||2\pi Q_{u,t+1}||}} \int \exp\left\{ -\frac{1}{2}\zeta \right\} dz_{t+1} \quad (50) \\
&= Z_{t+1} \sqrt{\frac{\left|\left| \left(Q_{z,t+1}^{-1} + \Omega_{t+1} \right)^{-1} \right|\right|}{||2\pi Q_{z,t+1}||\, ||2\pi Q_{u,t+1}||}} \exp\left\{ -\frac{1}{2}\xi \right\}, \quad (51)
\end{aligned}$$

where,

$$\begin{aligned}
\zeta &= z_{t+1}^T \Omega_{t+1} z_{t+1} - 2\lambda_{t+1}^T z_{t+1} + (z_{t+1} - A_{z,t+1} z_t)^T Q_{z,t+1}^{-1} (z_{t+1} - A_{z,t+1} z_t) \\
&\quad + (u_{t+1} - A_{u,t+1} z_t)^T Q_{u,t+1}^{-1} (u_{t+1} - A_{u,t+1} z_t) \\
&= z_{t+1}^T \left[\Omega_{t+1} + Q_{z,t+1}^{-1} \right] z_{t+1} - 2 \left[\lambda_{t+1}^T + z_t^T A_{z,t+1}^T Q_{z,t+1}^{-1} \right] z_{t+1} \\
&\quad + \left[z_t^T A_{z,t+1}^T Q_{z,t+1}^{-1} A_{z,t+1} z_t + (u_{t+1} - A_{u,t+1} z_t)^T Q_{u,t+1}^{-1} (u_{t+1} - A_{u,t+1} z_t) \right] \quad (52)
\end{aligned}$$

and,

$$
\begin{aligned}
\xi &= z_t^T A_{z,t+1}^T Q_{z,t+1}^{-1} A_{z,t+1} z_t + (u_{t+1} - A_{u,t+1} z_t)^T Q_{u,t+1}^{-1} (u_{t+1} - A_{u,t+1} z_t) \\
&\quad - \left[\lambda_{t+1}^T + z_t^T A_{z,t+1}^T Q_{z,t+1}^{-1}\right] \left[\Omega_{t+1} + Q_{z,t+1}^{-1}\right]^{-1} \left[\lambda_{t+1} + Q_{z,t+1}^{-1} A_{z,t+1} z_t\right] \\
&= z_t^T \left[A_{z,t+1}^T \left(Q_{z,t+1}^{-1} - Q_{z,t+1}^{-1} \left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{z,t+1}^{-1}\right) A_{z,t+1} + A_{u,t+1}^T Q_{u,t+1}^{-1} A_{u,t+1}\right] z_t \\
&\quad - 2\left[\lambda_{t+1}^T \left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{z,t+1}^{-1} A_{z,t+1} + u_{t+1}^T Q_{u,t+1}^{-1} A_{u,t+1}\right] z_t \\
&\quad - \left[\lambda_{t+1}^T \left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1} + u_{t+1}^T Q_{u,t+1}^{-1} u_{t+1}\right]. \qquad (53)
\end{aligned}
$$

Hence the prediction recursions are,

$$
\begin{aligned}
\hat{\Omega}_t &= A_{z,t+1}^T \left(Q_{z,t+1}^{-1} - Q_{z,t+1}^{-1} \left(z, Q_{t+1}^{-1} + \Omega_{t+1}\right)^{-1} Q_{z,t+1}^{-1}\right) A_{z,t+1} + A_{u,t+1}^T Q_{u,t+1}^{-1} A_{u,t+1} && (54) \\
\hat{\lambda}_t &= A_{z,t+1}^T Q_{z,t+1}^{-1} \left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1} + A_{u,t+1}^T Q_{u,t+1}^{-1} u_{t+1} && (55) \\
\hat{Z}_t &= Z_{t+1} \sqrt{\frac{\left|\left|\left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1}\right|\right|}{||2\pi Q_{z,t+1}||\,||2\pi Q_{u,t+1}||}} \\
&\qquad \times \exp\left\{-\frac{1}{2}\left[\lambda_{t+1}^T \left(Q_{z,t+1}^{-1} + \Omega_{t+1}\right)^{-1} \lambda_{t+1} + u_{t+1}^T Q_{u,t+1}^{-1} u_{t+1}\right]\right\}. && (56)
\end{aligned}
$$

Note that unlike hierarchical model, these all depend on u_t through the matrices $A_{z,t}$, $A_{u,t}$, $Q_{z,t}$, $Q_{u,t}$, so we will need to run a backward filter for every particle in the filter distribution (or at least, every one which we propose if we use rejection or MH sampling).

Update (this is identical to the hierarchical model):

$$
\begin{aligned}
p(y_{t:T}, u_{t+1:T} | z_t, u_t) &= p(y_{t+1:T}, u_{t+1:T} | z_t, u_t) p(y_t | z_t, u_t) \\
&= \hat{Z}_t \frac{1}{\sqrt{||2\pi R_t||}} \exp\left\{-\frac{1}{2}\omega\right\}, && (57)
\end{aligned}
$$

where,

$$
\begin{aligned}
\omega &= z_t^T \hat{\Omega}_t z_t - 2\hat{\lambda}_t^T z_t + (y_t - C_t z_t)^T R_t^{-1} (y_t - C_t z_t) \\
&= z_t^T \left[\hat{\Omega}_t + C_t^T R_t^{-1} C_t\right] z_t - 2\left[\hat{\lambda}_t^T + y_t^T C_t^T R_t^{-1}\right] z_t + \left[y_t^T R_t^{-1} y_t\right]. && (58)
\end{aligned}
$$

Hence the update recursions are,

$$
\begin{aligned}
\Omega_t &= \hat{\Omega}_t + C_t^T R_t^{-1} C_t && (59) \\
\lambda_t &= \hat{\lambda}_t + R_t^{-1} C_t y_t && (60) \\
Z_t &= \hat{Z}_t \frac{1}{\sqrt{||2\pi R_t||}} \exp\left\{-\frac{1}{2} y_t^T R_t^{-1} y_t\right\}. && (61)
\end{aligned}
$$

Finally, the initialisation. This requires the observation that when $t = T$, $p(y_{t:T}, u_{t+1:T} | z_t, u_t) = p(y_T | z_T, u_T)$. But then it's identical to the hierarchical model.

$$
\begin{aligned}
p(y_T|z_T, u_T) &= \mathcal{N}(y_T|C_T z_T, R_T) \\
&= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2}(y_T - C_T z_T)^T R_T^{-1}(y_T - C_T z_T)\right\} \\
&= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2} y_T R_T^{-1} y_T\right\} \exp\left\{z_T^T C_T^T R_T^{-1} C_T z_T - 2(y_T^T R_T^{-1} C_T) z_T\right\}
\end{aligned} \tag{62}
$$

Hence,

$$
\begin{aligned}
\Omega_T &= C_T^T R_T^{-1} C_T \\
\lambda_T &= C_T^T R_T^{-1} y_T \\
Z_T &= \frac{1}{\sqrt{||2\pi R_T||}} \exp\left\{-\frac{1}{2} y_T R_T^{-1} y_T\right\}.
\end{aligned} \tag{63}
$$

All done.