Rao-Blackwellised Particle Smoothing by Resampling

Pete Bunch

April 19, 2012

With smoothing, we would like to estimate

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

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

The nonlinear portion of this can be expressed in the following form:

$$P(u_{1:T}|y_{1:T}) = P(u_T|y_{1:T}) \prod_{t=1}^{T-1} P(u_t|u_{t+1:T}, y_{1:T}) \tag{2}$$

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

$$\begin{aligned}
P(u_t|u_{t+1:T}, y_{1:T}) &= \int P(u_{1:t}, z_t|u_{t+1:T}, y_{1:T}) dz_t du_{1:t-1} \\
&= \int \frac{P(y_{t+1:T}|z_t, u_{1:T}, y_{1:t})P(u_{1:t}, z_t|u_{t+1:T}, y_{1:t})}{P(y_{t+1:T}|u_{t+1:T}, y_{1:t})} dz_t du_{1:t-1} \\
&= \int \frac{P(y_{t+1:T}|z_t, u_{t+1:T})P(u_{1:t}, z_t|u_{t+1}, y_{1:t})}{P(y_{t+1:T}|u_{t+1:T}, y_{1:t})} dz_t du_{1:t-1} \\
&= \int \frac{P(y_{t+1:T}|z_t, u_{t+1:T})P(u_{t+1}|u_{1:t}, z_t, y_{1:t})P(z_t, u_{1:t}|y_{1:t})}{P(y_{t+1:T}|u_{t+1:T}, y_{1:t})P(u_{t+1}|y_{1:t})} dz_t du_{1:t-1} \\
&= \int \frac{P(y_{t+1:T}|z_t, u_{t+1:T})P(u_{t+1}|u_t)P(z_t|u_{1:t}, y_{1:t})P(u_{1:t}|y_{1:t})}{P(y_{t+1:T}|u_{t+1:T}, y_{1:t})P(u_{t+1}|y_{1:t})} dz_t du_{1:t-1} \\
&= \frac{\int P(u_{1:t}|y_{1:t})P(u_{t+1}|u_t) \int P(z_t|u_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u_{t+1:T}) dz_t du_{1:t-1}}{P(y_{t+1:T}|u_{t+1:T}, y_{1:t})P(u_{t+1}|y_{1:t})}
\end{aligned} \tag{3}$$

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

$$P(u_{1:t}|y_{1:t}) \approx \sum_i w_t^{(i)} \delta_{u_{1:t}^{(i)}}(u_{1:t}) \tag{4}$$

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

$$
\begin{aligned}
P(u_t|u^{(j)}_{t+1:T}, y_{1:T}) &\propto \int P(u_{1:t}|y_{1:t})P(u^{(j)}_{t+1}|u_t) \int P(z_t|u_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T})dz_t du_{1:t-1} \\
&\approx \int \sum_i w^{(i)}_t \delta_{u^{(i)}_{1:t}}(u_{1:t})P(u^{(j)}_{t+1}|u_t) \int P(z_t|u_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T})dz_t du_{1:t-1} \\
&= \sum_i w^{(i)}_t P(u^{(j)}_{t+1}|u^{(i)}_t) \int P(z_t|u^{(i)}_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T})dz_t \, \delta_{u^{(i)}_t}(u_t) \\
&= \sum_i \rho^{(i)} \delta_{u^{(i)}_t}(u_t) \qquad (5)
\end{aligned}
$$

where

$$
\rho^{(i)} \propto w^{(i)}_t P(u^{(j)}_{t+1}|u^{(i)}_t) \int P(z_t|u^{(i)}_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T})dz_t \qquad (6)
$$

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

$$
P(z_t|u^{(i)}_{1:t}, y_{1:t}) = \mathcal{N}(z_t|\mu^{(i)}_t, \Sigma^{(i)}_t) \qquad (7)
$$

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

$$
P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T}) = \hat{Z}^{(j)}_t \mathcal{N}(z_t|\hat{m}^{(j)}_t, \hat{S}^{(j)}_t) \qquad (8)
$$

The integral is then given by

$$
\begin{aligned}
\int & P(z_t|u^{(i)}_{1:t}, y_{1:t})P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T})dz_t \\
&= \hat{Z}^{(j)}_t \int \mathcal{N}(z_t|\mu^{(i)}_t, \Sigma^{(i)}_t)\mathcal{N}(z_t|\hat{m}^{(j)}_t, \hat{S}^{(j)}_t)dz_t \\
&= \hat{Z}^{(j)}_t \mathcal{N}(\mu^{(i)}_t|\hat{m}^{(j)}_t, \Sigma^{(i)}_t + \hat{S}^{(j)}_t) \qquad (9)
\end{aligned}
$$

which uses the product of Gaussians formula. Thus the backward sampling weights are given by

$$
\rho^{(i)} \propto w^{(i)}_t P(u^{(j)}_{t+1}|u^{(i)}_t)\hat{Z}^{(j)}_t \mathcal{N}(\mu^{(i)}_t|\hat{m}^{(j)}_t, \Sigma^{(i)}_t + \hat{S}^{(j)}_t) \qquad (10)
$$

The normalisation factor $\hat{Z}^{(j)}_t$ is constant over the sampling index i, and may thus be ommited (comes out in the normalisation). It may be useful for reweighting-type smoothers.

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

$$
\begin{aligned}
P(y_{t:T}|z_t, u_{t:T}) &= Z_t \mathcal{N}(z_t|m_t, S_t) \qquad (11) \\
P(y_{t+1:T}|z_t, u_{t+1:T}) &= \hat{Z}_t \mathcal{N}(z_t|\hat{m}_t, \hat{S}_t) \qquad (12)
\end{aligned}
$$

The following use the Gaussian product formulas and the Gaussian change of variable formula.

Predict:

$$
\begin{aligned}
&P(y_{t+1:T}|z_t, u_{t+1:T}) \\
&\qquad = \int P(Y_{t+1}|z_{t+1}, u_{t+1:T})P(z_{t+1}|u_{t+1}, z_t)dz_{t+1} \\
&\qquad = Z_{t+1} \int \mathcal{N}(z_{t+1}|m_{t+1}, S_t)\mathcal{N}(z_{t+1}|A_t z_t, Q_{t+1}) \\
&\qquad = Z_{t+1}\mathcal{N}(m_{t+1}|A_t z_t, S_{t+1} + Q_{t+1}) \\
&\qquad = Z_{t+1} \left|A_t^{-1}\right| \mathcal{N}(z_t|A_t^{-1}m_{t+1}, A_t^{-1}(S_{t+1} + Q_{t+1})A_t^{-T})
\end{aligned} \tag{13}
$$

Thus:

$$\hat{m}_t = A_t^{-1}m_{t+1} \tag{14}$$

$$\hat{S}_t = A_t^{-1}(S_{t+1} + Q_{t+1})A_t^{-T} \tag{15}$$

$$\hat{Z}_t = Z_{t+1} \left|A_t\right|^{-1} \tag{16}$$

Update:

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

where:

$$P_t = C_t\hat{S}_t C_t^T + R_t \tag{18}$$

$$K_t = \hat{S}_t C_t^T P_t^{-1} \tag{19}$$

$$m_t = \hat{m}_t + K_t(y_t - C_t\hat{m}_t) \tag{20}$$

$$S_t = \hat{S}_t - K_t P_t K_t^T \tag{21}$$

$$Z_t = \hat{Z}_t\mathcal{N}(y_t|C_t\hat{m}_t, P_t) \tag{22}$$

Finally we need to consider the initialisation of the two filter smoother. At time T, the joint likelihood is given by

$$P(y_T|z_T, u_T) = \mathcal{N}(y_T|C_T z_T, R_T) \tag{23}$$

which cannot be inverted because C_t is generally not square and full rank. If we wait a few steps then the set of observations can be written as

$$\begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{t+1} \\ y_t \end{bmatrix} = \underbrace{\begin{bmatrix} C_T \prod_{k=t}^{T-1} A_k \\ C_{T-1} \prod_{k=t}^{T-2} A_k \\ \vdots \\ C_{t+1} A_t \\ C_t \end{bmatrix}}_{H_t} x_t + \underbrace{\begin{bmatrix} C_T & C_T A_{T-1} & \dots & C_T \prod_{k=t+1}^{T-1} A_k \\ 0 & C_{T-1} & \dots & C_{T-1} \prod_{k=t+1}^{T-2} A_k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{t+1} \\ 0 & 0 & \dots & 0 \end{bmatrix}}_{K_t} \begin{bmatrix} w_{T-1} \\ w_{T-2} \\ \vdots \\ w_{t+1} \end{bmatrix}$$

$$+ \begin{bmatrix} I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \\ 0 & 0 & \dots & 0 & I \end{bmatrix} \begin{bmatrix} v_T \\ v_{T-1} \\ \vdots \\ v_{t+1} \\ v_t \end{bmatrix} \tag{24}$$

Thus

$$P(y_{t:T} | Z_t, u_{t:T}) = \mathcal{N}(\underline{y}_t | H_t x_t, V_t) \tag{25}$$

where

$$V_t = K_t \begin{bmatrix} Q_T & 0 & \dots & 0 \\ 0 & Q_{T-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{t+1} \end{bmatrix} K_t^T + \begin{bmatrix} R_T & 0 & \dots & 0 & 0 \\ 0 & R_{T-1} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & R_{t+1} & 0 \\ 0 & 0 & \dots & 0 & R_t \end{bmatrix} \tag{26}$$

This can be inverted once $T - t$ becomes large enough.

$$P(y_{t:T} | Z_t, u_{t:T}) = \frac{|\Lambda_t|^{1/2}}{|V_t|^{1/2}} \exp\{-\frac{1}{2}[\underline{y}_t^T (V_t^{-1} - V_t^{-1} H_t \Lambda_t H_t^T V_t^{-1}) \underline{y}_t]\} \mathcal{N}(x_t | \nu_t, \Lambda_t) \tag{27}$$

$$= \frac{|\Lambda_t|^{1/2}}{|V_t|^{1/2}} \mathcal{N}(x_t | \nu_t, \Lambda_t) \tag{28}$$

with

$$\Lambda_t = [H_t^T V_t^{-1} H_t]^{-1} \tag{29}$$

$$\nu_t = \Lambda_t H_t^T V_t^{-1} \underline{y}_t \tag{30}$$

Before the backward filter recursion begins, the backwards sampling weights are given by

$$
\begin{aligned}
&\int P(z_t|u^{(i)}_{1:t}, y_{1:t}) P(y_{t+1:T}|z_t, u^{(j)}_{t+1:T}) dz_t \\
&\qquad = \int \mathcal{N}(z_t|\mu^{(i)}_t, \Sigma^{(i)}_t) \mathcal{N}(\underline{y}_{t+1}|\hat{H}_t x^{(j)}_t, \hat{V}_t) dz_t \\
&\qquad = \mathcal{N}(\underline{y}_{t+1}|\hat{H}_t \mu^{(i)}_t, \hat{H}_t \Sigma^{(i)}_t \hat{H}^T_t + \hat{V}_t)
\end{aligned} \tag{31}
$$

where

$$
\hat{H}_t = H^{(j)}_{t+1} A^{(i)}_t \tag{32}
$$

$$
\hat{V}_t = V^{(j)}_{t+1} + H^{(j)}_{t+1} Q^{(j)}_{t+1} H^{(j)T}_{t+1} \tag{33}
$$