Switch-Reset Models : Exact and Approximate Inference

Chris Bracegirdle **David Barber**
University College London, Gower Street, London UK
{c.bracegirdle, d.barber}@cs.ucl.ac.uk

A SUPPLEMENTARY MATERIAL

This appendix contains results referred to throughout: in particular, the full derivation of β backward recursions (general and reset-system case), the marginal inference derivations in respect of the switch-reset LDS, and the full bracket smoother recursions. Finally, we give two further examples of the models in practice.

A.1 β Smoother Update Rules: Full Derivation

In this section, we first derive the update rules for the β backward recursion.

Standard canonical form smoothing in respect of visible variables y, used for the updates in β messages, can be derived as follows.

$$\begin{aligned}
\beta(h_t) &= p(y_{t+1:T}|h_t) \\
&= \int_{h_{t+1}} p(y_{t+1}|h_{t+1}, \cancel{h_t}, \cancel{y_{t+2:T}}) p(h_{t+1}, y_{t+2:T}|h_t) \\
&= \int_{h_{t+1}} p(y_{t+1}|h_{t+1}) \underbrace{p(y_{t+2:T}|h_{t+1}, \cancel{h_t})}_{=\beta(h_{t+1})} p(h_{t+1}|h_t)
\end{aligned}$$

In the case of a linear dynamical system, we assume squared-exponential messages in canonical form, given by $\beta(\mathbf{h}_{t+1}) = k_{t+1} \exp -\frac{1}{2}\left(\mathbf{h}_{t+1}^\top \mathbf{G}_{t+1}\mathbf{h}_{t+1} - 2\mathbf{h}_{t+1}^\top \mathbf{g}_{t+1}\right)$. Then

$$\begin{aligned}
\beta(\mathbf{h}_t) &= \int_{\mathbf{h}_{t+1}} \frac{1}{\sqrt{|2\pi\mathbf{R}|}} \exp -\frac{1}{2}\left(\mathbf{y}_{t+1} - \mathbf{B}\mathbf{h}_{t+1} - \bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1} - \mathbf{B}\mathbf{h}_{t+1} - \bar{\mathbf{y}}\right) \\
&\quad \times \frac{1}{\sqrt{|2\pi\mathbf{Q}|}} \exp -\frac{1}{2}\left(\mathbf{h}_{t+1} - \mathbf{A}\mathbf{h}_t - \bar{\mathbf{h}}\right)^\top \mathbf{Q}^{-1}\left(\mathbf{h}_{t+1} - \mathbf{A}\mathbf{h}_t - \bar{\mathbf{h}}\right) \\
&\quad \times k_{t+1} \exp -\frac{1}{2}\left(\mathbf{h}_{t+1}^\top \mathbf{G}_{t+1}\mathbf{h}_{t+1} - 2\mathbf{h}_{t+1}^\top \mathbf{g}_{t+1}\right) \\
&= \frac{k_{t+1}}{\sqrt{|2\pi\mathbf{R}|\,|2\pi\mathbf{Q}|}} \exp -\frac{1}{2}\left\{\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right) + \left(\mathbf{A}\mathbf{h}_t + \bar{\mathbf{h}}\right)^\top \mathbf{Q}^{-1}\left(\mathbf{A}\mathbf{h}_t + \bar{\mathbf{h}}\right)\right\} \\
&\quad \times \int_{\mathbf{h}_{t+1}} \exp -\frac{1}{2}\left\{\begin{array}{c} \mathbf{h}_{t+1}^\top\left[\mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B} + \mathbf{Q}^{-1} + \mathbf{G}_{t+1}\right]\mathbf{h}_{t+1} \\ -2\mathbf{h}_{t+1}^\top\left[\mathbf{B}^\top\mathbf{R}^{-1}\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right) + \mathbf{Q}^{-1}\left(\mathbf{A}\mathbf{h}_t + \bar{\mathbf{h}}\right) + \mathbf{g}_{t+1}\right] \end{array}\right\}
\end{aligned}$$

Set $\mathbf{M} = \mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B} + \mathbf{Q}^{-1} + \mathbf{G}_{t+1}$ and $\mathbf{b} = \mathbf{B}^\top\mathbf{R}^{-1}\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right) + \mathbf{Q}^{-1}\bar{\mathbf{h}} + \mathbf{g}_{t+1}$. Then

$$\begin{aligned}
\beta(\mathbf{h}_t) &= \frac{k_{t+1}}{\sqrt{|2\pi\mathbf{R}|\,|2\pi\mathbf{Q}|}} \exp -\frac{1}{2}\left\{\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}\right) + \left(\mathbf{A}\mathbf{h}_t + \bar{\mathbf{h}}\right)^\top \mathbf{Q}^{-1}\left(\mathbf{A}\mathbf{h}_t + \bar{\mathbf{h}}\right)\right\} \\
&\quad \times \int_{\mathbf{h}_{t+1}} \exp -\frac{1}{2}\left(\mathbf{h}_{t+1} - \mathbf{M}^{-1}\left[\mathbf{b} + \mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]\right)^\top \mathbf{M}\left(\mathbf{h}_{t+1} - \mathbf{M}^{-1}\left[\mathbf{b} + \mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]\right) \\
&\quad \times \exp \frac{1}{2}\left[\mathbf{b} + \mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]^\top \mathbf{M}^{-1}\left[\mathbf{b} + \mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]
\end{aligned}$$

Since $\mathbf{M}$ is symmetric,

$$\beta(\mathbf{h}_t) = \frac{k_{t+1}\sqrt{|2\pi\mathbf{M}^{-1}|}}{\sqrt{|2\pi\mathbf{R}|\,|2\pi\mathbf{Q}|}} \exp -\tfrac{1}{2}\left\{\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right) + \left(\mathbf{A}\mathbf{h}_t+\bar{\mathbf{h}}\right)^\top \mathbf{Q}^{-1}\left(\mathbf{A}\mathbf{h}_t+\bar{\mathbf{h}}\right)\right\}$$
$$\times \exp \tfrac{1}{2}\left[\mathbf{b}+\mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]^\top \mathbf{M}^{-1}\left[\mathbf{b}+\mathbf{Q}^{-1}\mathbf{A}\mathbf{h}_t\right]$$

Aim to write $\beta\left(\mathbf{h}_t\right) = k_t \exp -\frac{1}{2}\left(\mathbf{h}_t^\top \mathbf{G}_t \mathbf{h}_t - 2\mathbf{h}_t^\top \mathbf{g}_t\right)$. Then, noting $\dim \mathbf{M} = \dim \mathbf{Q}$

$$\begin{aligned}
\mathbf{G}_t &= \mathbf{A}^\top\left[\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{M}^{-1}\mathbf{Q}^{-1}\right]\mathbf{A} \\
\mathbf{g}_t &= \mathbf{A}^\top\mathbf{Q}^{-1}\left[\mathbf{M}^{-1}\mathbf{b} - \bar{\mathbf{h}}\right] \\
k_t &= \frac{k_{t+1}}{\sqrt{|2\pi\mathbf{R}|\,|\mathbf{Q}\mathbf{M}|}} \exp -\tfrac{1}{2}\left[\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right) + \bar{\mathbf{h}}^\top\mathbf{Q}^{-1}\bar{\mathbf{h}}\right] \times \exp \tfrac{1}{2}\mathbf{b}^\top\mathbf{M}^{-1}\mathbf{b}
\end{aligned}$$

where $\mathbf{M} = \mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B} + \mathbf{Q}^{-1} + \mathbf{G}_{t+1}$ and $\mathbf{b} = \mathbf{B}^\top\mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right) + \mathbf{Q}^{-1}\bar{\mathbf{h}} + \mathbf{g}_{t+1}$.

Now in the reset LDS case,

$$p(\mathbf{h}_t, c_t|\mathbf{y}_{1:T}) = \frac{p(\mathbf{h}_t, c_t, \mathbf{y}_{t+1:T}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1:T}|\mathbf{y}_{1:t})} = \frac{p(\mathbf{y}_{t+1:T}|\mathbf{h}_t, c_t, \cancel{\mathbf{y}_{1:t}})p(\mathbf{h}_t, c_t|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1:T}|\mathbf{y}_{1:t})}$$

we use $p(\mathbf{h}_t, c_t|\mathbf{y}_{1:t}) \propto \alpha(\mathbf{h}_t, c_t)$ and $p(\mathbf{y}_{t+1:T}|\mathbf{h}_t, c_t) = \beta(\mathbf{h}_t, c_t)$. We have

$$\begin{aligned}
\beta(\mathbf{h}_t, c_t) &= \sum_{c_{t+1}}\int_{\mathbf{h}_{t+1}} p(\mathbf{y}_{t+1}|\mathbf{h}_{t+1}, c_{t+1})p(\mathbf{h}_{t+1}|\mathbf{h}_t, c_{t+1})p(c_{t+1}|c_t)\beta(\mathbf{h}_{t+1}, c_{t+1}) \\
&= p(c_{t+1}=0|c_t)\underbrace{\int_{\mathbf{h}_{t+1}} p^0(\mathbf{y}_{t+1}|\mathbf{h}_{t+1})p^0(\mathbf{h}_{t+1}|\mathbf{h}_t)\beta(\mathbf{h}_{t+1}, c_{t+1}=0)}_{\beta^0(\mathbf{h}_t)} \\
&\quad + p(c_{t+1}=1|c_t)\underbrace{\int_{\mathbf{h}_{t+1}} p^1(\mathbf{y}_{t+1}|\mathbf{h}_{t+1})p^1(\mathbf{h}_{t+1})\beta(\mathbf{h}_{t+1}, c_{t+1}=1)}_{\beta_t^1}
\end{aligned}$$

where we have written $\beta(\mathbf{h}_t, c_t) = p(c_{t+1}=0|c_t)\beta^0(\mathbf{h}_t) + p(c_{t+1}=1|c_t)\beta_t^1$. We then have

$$\beta^0(\mathbf{h}_t) = \int_{\mathbf{h}_{t+1}} p^0(\mathbf{y}_{t+1}|\mathbf{h}_{t+1})p^0(\mathbf{h}_{t+1}|\mathbf{h}_t)\left[p(c_{t+2}=0|c_{t+1}=0)\beta^0(\mathbf{h}_{t+1}) + p(c_{t+2}=1|c_{t+1}=0)\beta_{t+1}^1\right]$$

and

$$\beta_t^1 = \int_{\mathbf{h}_{t+1}} p^1(\mathbf{y}_{t+1}|\mathbf{h}_{t+1})p^1(\mathbf{h}_{t+1})\left[p(c_{t+2}=0|c_{t+1}=1)\beta^0(\mathbf{h}_{t+1}) + p(c_{t+2}=1|c_{t+1}=1)\beta_{t+1}^1\right]$$

We can calculate β^0 using the canonical update rule given above. For each component from the previous iteration,

$$\begin{aligned}
\beta_t^1 &= \int_{\mathbf{h}_{t+1}} p^1(\mathbf{y}_{t+1}|\mathbf{h}_{t+1})p^1(\mathbf{h}_{t+1})\beta(\mathbf{h}_{t+1}) \\
&= \int_{\mathbf{h}_{t+1}} \frac{1}{\sqrt{|2\pi\mathbf{R}|}} \exp -\tfrac{1}{2}\left(\mathbf{y}_{t+1}-\mathbf{B}\mathbf{h}_{t+1}-\bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\mathbf{B}\mathbf{h}_{t+1}-\bar{\mathbf{y}}\right) \\
&\quad \times \frac{1}{\sqrt{|2\pi\mathbf{Q}|}} \exp -\tfrac{1}{2}\left(\mathbf{h}_{t+1}-\bar{\mathbf{h}}\right)^\top \mathbf{Q}^{-1}\left(\mathbf{h}_{t+1}-\bar{\mathbf{h}}\right) \times k_{t+1}\exp -\tfrac{1}{2}\left(\mathbf{h}_{t+1}^\top\mathbf{G}_{t+1}\mathbf{h}_{t+1} - 2\mathbf{h}_{t+1}^\top\mathbf{g}_{t+1}\right) \\
&= \frac{k_{t+1}}{\sqrt{|2\pi\mathbf{R}|\,|2\pi\mathbf{Q}|}} \exp -\frac{1}{2}\left\{\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right) + \bar{\mathbf{h}}^\top\mathbf{Q}^{-1}\bar{\mathbf{h}}\right\} \\
&\quad \times \int_{\mathbf{h}_{t+1}} \exp -\tfrac{1}{2}\left\{\mathbf{h}_{t+1}^\top\left[\mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B} + \mathbf{Q}^{-1} + \mathbf{G}_{t+1}\right]\mathbf{h}_{t+1} - 2\mathbf{h}_{t+1}^\top\left[\mathbf{B}^\top\mathbf{R}^{-1}\left(\mathbf{y}_{t+1}-\bar{\mathbf{y}}\right) + \mathbf{Q}^{-1}\bar{\mathbf{h}} + \mathbf{g}_{t+1}\right]\right\}
\end{aligned}$$

As before, set $\mathbf{M} = \mathbf{B}^\top \mathbf{R}^{-1}\mathbf{B} + \mathbf{Q}^{-1} + \mathbf{G}_{t+1}$ and $\mathbf{b} = \mathbf{B}^\top \mathbf{R}^{-1}(\mathbf{y}_{t+1} - \bar{\mathbf{y}}) + \mathbf{Q}^{-1}\bar{\mathbf{h}} + \mathbf{g}_{t+1}$. Then

$$\begin{aligned}\beta_t^1 &= \frac{k_{t+1}}{\sqrt{|2\pi\mathbf{R}|\,|\mathbf{QM}|}} \exp -\tfrac{1}{2}\left\{(\mathbf{y}_{t+1} - \bar{\mathbf{y}})^\top \mathbf{R}^{-1}(\mathbf{y}_{t+1} - \bar{\mathbf{y}}) + \bar{\mathbf{h}}^\top \mathbf{Q}^{-1}\bar{\mathbf{h}}\right\} \times \exp \tfrac{1}{2}\mathbf{b}^\top \mathbf{M}^{-1}\mathbf{b} \\ &= k_t\end{aligned}$$

Now to combine the forward and backward messages[8]

$$\begin{aligned}p(\mathbf{h}_t, c_t|\mathbf{y}_{1:T}) &\propto \alpha(\mathbf{h}_t, c_t)\beta(\mathbf{h}_t, c_t) = \alpha(\mathbf{h}_t, c_t)\left[p(c_{t+1} = 0|c_t)\beta^0(\mathbf{h}_t) + p(c_{t+1} = 1|c_t)\beta_t^1\right] \\ &\propto \left[\sum_{i(c_t)} w_i \mathcal{N}(\mathbf{f}_i, \mathbf{F}_i)\right]\left\{p(c_{t+1} = 0|c_t)\left[\sum_j k_j \mathcal{CAN}(\mathbf{g}_j, \mathbf{G}_j)\right] + p(c_{t+1} = 1|c_t)\beta_t^1\right\}\end{aligned}$$

Consider each term

$$\begin{aligned}w_i\mathcal{N}(\mathbf{f}_i, \mathbf{F}_i)\, k_j \mathcal{CAN}(\mathbf{g}_j, \mathbf{G}_j) &= \frac{w_i}{\sqrt{|2\pi\mathbf{F}_i|}} \exp -\tfrac{1}{2}(\mathbf{h}_t - \mathbf{f}_i)^\top \mathbf{F}_i^{-1}(\mathbf{h}_t - \mathbf{f}_i) \times k_j \exp -\tfrac{1}{2}\left(\mathbf{h}_t^\top \mathbf{G}_j \mathbf{h}_t - 2\mathbf{h}_t^\top \mathbf{g}_j\right) \\ &= \frac{w_i k_j}{\sqrt{|2\pi\mathbf{F}_i|}} \exp -\tfrac{1}{2}\left\{\mathbf{h}_t^\top\left[\mathbf{F}_i^{-1} + \mathbf{G}_j\right]\mathbf{h}_t - 2\mathbf{h}_t^\top\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right] + \mathbf{f}_i^\top \mathbf{F}_i^{-1}\mathbf{f}_i\right\}\end{aligned}$$

Set $\mathbf{M}' = \mathbf{F}_i^{-1} + \mathbf{G}_j$. Then the posterior component is given by

$$\begin{aligned}&w_i\mathcal{N}(\mathbf{f}_i, \mathbf{F}_i)\, k_j \mathcal{CAN}(\mathbf{g}_j, \mathbf{G}_j) \\ &= \frac{w_i k_j}{\sqrt{|2\pi\mathbf{F}_i|}} \exp -\tfrac{1}{2}\left\{\begin{array}{c}\left(\mathbf{h}_t - \mathbf{M}'^{-1}\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right]\right)^\top \mathbf{M}'\left(\mathbf{h}_t - \mathbf{M}'^{-1}\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right]\right) \\ -\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right]^\top \mathbf{M}'^{-1}\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right] + \mathbf{f}_i^\top \mathbf{F}_i^{-1}\mathbf{f}_i\end{array}\right\} \\ &= \frac{w_i k_j}{\sqrt{|\mathbf{F}_i\mathbf{M}'|}} \exp -\tfrac{1}{2}\left\{\mathbf{f}_i^\top \mathbf{F}_i^{-1}\mathbf{f}_i - \left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right]^\top \mathbf{M}'^{-1}\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right]\right\}\mathcal{N}\left(\mathbf{M}'^{-1}\left[\mathbf{F}_i^{-1}\mathbf{f}_i + \mathbf{g}_j\right], \mathbf{M}'^{-1}\right)\end{aligned}$$

A.2 Switch-Reset Models: Inference

Marginal inference is straightforward based on extending the variable $h_t \to (h_t, s_t)$ and using the recursions in section(2). Briefly, we first define the equivalent run-length model, with transition $p(\rho_t|s_t, s_{t-1}, \rho_{t-1})$. Then the filtered posterior is $p(h_t, s_t, \rho_t, y_{1:t}) = \tilde{\alpha}(h_t|s_t, \rho_t)\tilde{\alpha}(s_t, \rho_t)$. The discrete component updates according to

$$\tilde{\alpha}(s_t, \rho_t) = p(y_t|s_t, \rho_t) \sum_{s_{t-1}, \rho_{t-1}} p(\rho_t|s_t, s_{t-1}, \rho_{t-1})p(s_t|s_{t-1}, \rho_{t-1})\tilde{\alpha}(s_{t-1}, \rho_{t-1})$$

where we note $\rho_t \neq 0 \Rightarrow (\rho_{t-1} = \rho_t - 1) \wedge (s_{t-1} = s_t)$. This shows that discrete filtered distribution scales as $O\left(S^2T^2\right)$. The continuous component is calculated using standard forward propagation, conditioned on $\rho_{1:t}, s_{1:t}$. For smoothing, we may apply either the $\tilde{\alpha} - \tilde{\gamma}$ approach, or use bracketing. For bracketing, in the no-reset case,

$$p(s_t, \rho_t, \varsigma_t|y_{1:T}) = p(s_{t+1} = s_t, \rho_{t+1} = \rho_t + 1, \varsigma_{t+1} = \varsigma_t - 1|y_{1:T})$$

and for the reset case $\varsigma_t = 1 \Leftrightarrow s_{t+1} \neq s_t \Leftrightarrow \rho_{t+1} = 0$, so $p(s_t, \rho_t, \varsigma_t = 1|y_{1:T})$ is given by

$$p(s_t, \rho_t|y_{1:t}) \sum_{s_{t+1} \neq s_t} \frac{p(s_{t+1}|s_t, \rho_t)}{p(s_{t+1}, \rho_{t+1} = 0|y_{1:t})} \sum_{\varsigma_{t+1}} p(s_{t+1}, \rho_{t+1} = 0, \varsigma_{t+1}|y_{1:T})$$

with (for $s_{t+1} \neq s_t$)

$$p(s_{t+1}, \rho_{t+1} = 0|y_{1:t}) = \sum_{s_t \neq s_{t+1}, \rho_t} p(s_{t+1}|s_t, \rho_t)p(s_t, \rho_t|y_{1:t})$$

This gives an overall $O\left(S^2T^3\right)$ complexity for smoothing. The continuous component of the smoothed posterior $p(h_t|s_t, \rho_t, \varsigma_t, y_{1:T})$ is calculated by standard smoothing on the bracket.

[8] For brevity, we here use $\mathcal{CAN}(\mathbf{g}, \mathbf{G})$ to refer to a squared exponential component of the form $\exp -\frac{1}{2}\left(\mathbf{h}_t^\top \mathbf{G}\mathbf{h}_t - 2\mathbf{h}_t^\top \mathbf{g}\right)$.

A.3 LDS Algorithms

We give the full algorithms for the LDS bracket smoother, along with Kalman filter and correction smoother update routines from Barber (2011). Note that, for the general model, LDSForward and LDSBackward can be replaced with any update routine calculating sufficient statistics and corresponding likelihood.

Algorithm 1 RLDS Filtering for a model with parameters θ^0 (no reset) and θ^1 (reset).

```
{f_1(ρ=0), F_1(ρ=0), p_1} ← LDSForward(0, 0, y_1; θ^1)                                   ▷ Initial reset case
w_1(ρ=0) ← p_1 × p(c_1 = 1)
{f_1(ρ=1), F_1(ρ=1), p_1} ← LDSForward(0, 0, y_1; θ^0)                                   ▷ Initial non-reset case
w_1(ρ=1) ← p_1 × p(c_1 = 0)
l_1 ← Σ w_1, w_1 ← w_1/Σ w_1                                                             ▷ Likelihood, Normalise
for t ← 2, T do
    {f_t(ρ=0), F_t(ρ=0), p_t} ← LDSForward(0, 0, y_t; θ^1)                                ▷ Reset case
    w_t(ρ=0) ← p_t × [p(c_{t+1}=1|c_t=1) w_{t-1}(ρ_{t-1}=0) + p(c_{t+1}=1|c_t=0) Σ_{ρ_{t-1}=1}^{t-1} w_{t-1}(ρ_{t-1})]
    for ρ ← 1, t do
        {f_t(ρ), F_t(ρ), p_t} ← LDSForward(f_{t-1}(ρ-1), F_{t-1}(ρ-1), y_t; θ^0)          ▷ Non-reset cases
        w_t(ρ) ← p_t × p(c_{t+1}=0|c_t=𝕀(ρ=1)) w_{t-1}(ρ_{t-1}=ρ-1)
    end for
    l_t ← l_{t-1} × Σ w_t, w_t ← w_t/Σ w_t                                                ▷ Likelihood, Normalise
end for
```

Algorithm 2 LDS standard Kalman filter, with parameters θ.

```
function LDSForward(f, F, y; θ)
    μ_h ← Af + h̄, μ_y ← Bμ_h + ȳ                                                          ▷ Mean of p(h_t, y_t|y_{1:t-1})
    Σ_hh ← AFA^⊤ + Q, Σ_yy ← BΣ_hh B^⊤ + R, Σ_yh ← BΣ_hh                                   ▷ Covariance of p(h_t, y_t|y_{1:t-1})
    f' ← μ_h + Σ_yh^⊤ Σ_yy^{-1}(y - μ_y), F' ← Σ_hh - Σ_yh^⊤ Σ_yy^{-1} Σ_yh                ▷ Find p(h_t|y_{1:t}) by conditioning
    p' ← exp(-½(y - μ_y)^⊤ Σ_yy^{-1}(y - μ_y)) / sqrt(det(2πΣ_yy))                          ▷ Compute likelihood
    return f', F', p'
end function
```

Algorithm 3 RLDS Bracket Correction Smoothing, with parameters θ^0 (no reset) and θ^1 (reset).

```
x_T ← w_T, g_T ← f_T, G_T ← F_T                                                          ▷ Initialise to filtered posterior
for t ← T-1, 1 do
    x_t(0:t, 2:T-t+1) ← x_{t+1}(1:t+1, 1:T-t)                                             ▷ Non-reset cases
    for ρ ← 0, t do
        x_t(ρ, 1) ← p(c_{t+1}=1|c_t=𝕀(ρ=0)) × w_{t+1}(ρ)                                  ▷ Reset cases
    end for
    x_t(:, 1) ← x_t(:, 1) × Σ x_{t+1}(0, :)/Σ x_t(:, 1)                                   ▷ Normalise
    g_t(:, 1) ← f_t, G_t(:, 1) ← F_t                                                      ▷ Copy filtered moments
    for ρ ← 0, t; ς ← 2, T-t+1 do                                                         ▷ Calculate moments
        {g_t(ρ, ς), G_t(ρ, ς)} ← LDSBackward(g_{t+1}(ρ+1, ς-1), G_{t+1}(ρ+1, ς-1), f_t(ρ), F_t(ρ); θ^0)
    end for
end for
```

Algorithm 4 LDS standard RTS correction update, with parameters θ.

```
function LDSBackward(g, G, f, F; θ)
    μ_h ← Af + h̄, Σ_h'h' ← AFA^⊤ + Q, Σ_h'h ← AF                                          ▷ Statistics of p(h_t, h_{t+1}|y_{1:t})
    ←Σ ← F - Σ_h'h^⊤ Σ_h'h'^{-1} Σ_h'h, ←A ← Σ_h'h^⊤ Σ_h'h'^{-1}, ←m ← f - ←A μ_h          ▷ Dynamics reversal p(h_t|h_{t+1}, y_{1:t})
    g' ← ←A g + ←m, G' ← ←A G ←A^⊤ + ←Σ                                                   ▷ Backward propagation
    return g', G'
end function
```

A.4 Piecewise-Constant Model: Well-Log Example

A piecewise reset model as described in section(5), widely known as a change-point model, is implemented by specifying the reset-case latent prior $p^1(h_t)$, emission $p(y_t|h_t, c_t)$, and deriving the forward updates by appealing to equation (5.1) in the no-reset case and equation (2.4) in the reset case.

Change-point models have been widely applied to the well-log data of Ó Ruanaidh and Fitzgerald (1996)[9], in the form of a noisy step function. Adams and MacKay (2007) used a Gaussian prior distribution over the piecewise-constant mean of the Gaussian-distributed data for filtered inference. We implemented such model in our smoothing approximation framework ($h_t = \mu_t$, $p^0(\mu_t|\mu_{t-1}) = \delta(\mu_t - \mu_{t-1})$), using the same parameters: $p^1(\mu_t) = \mathcal{N}\left(\mu_t|1.15 \times 10^5, 10^8\right)$ and change-point probability $p(c_t = 1) = \frac{1}{250}$. Results are shown in fig(6).

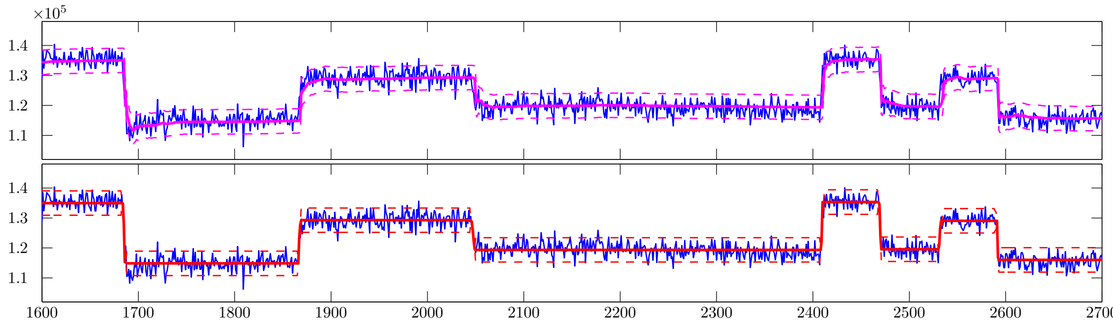

Figure 6: Window of the 4050-datum well-log data set, as shown by Adams and MacKay (2007). We show (i) the observed data overlaid with the filtered mean and *a posteriori* observation standard deviation; (ii) the smoothed equivalent. This example was run with $N = 10$ components comprising each approximate message.

A.5 Switch-Reset LDS: Generated Example

We give a generated example of the switch-reset LDS, as an experiment for which the truth of the state mass is known. The results are shown in fig(7), in which we observe that the approximate posterior tends to the exact posterior as the number of components increases. As we see, good results can be obtained based on using a very limited number of message components (10) compared to the number required to perform exact smoothing (10, 100). Intuitively, the reason is that in the exact case, information is kept for filtering and smoothing time t from the whole sequence before and after t.

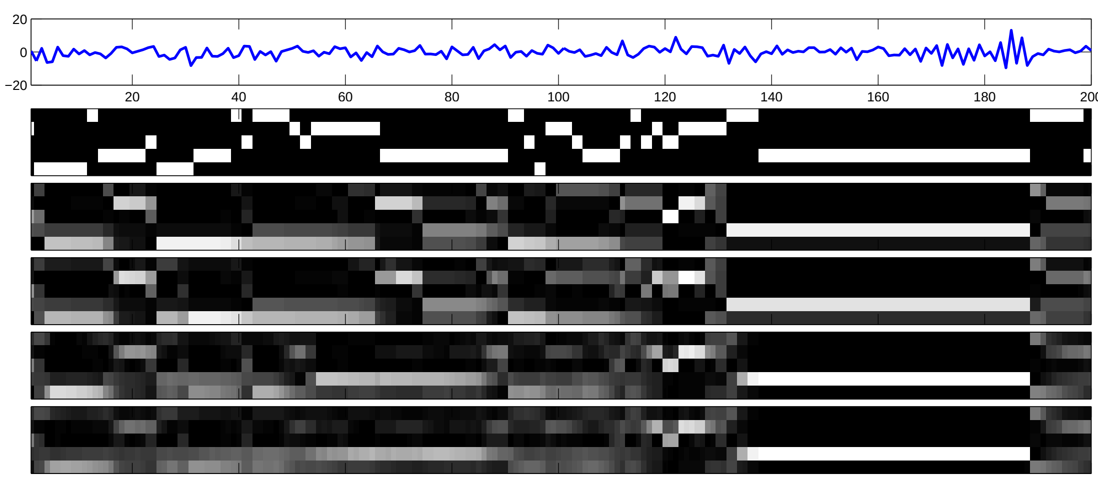

Figure 7: Switch-Reset LDS example. We generated a single-dimensional timeseries, with $S = 5$ states, $T = 200$, and two-dimensional latent dynamics using random parameters. From top to bottom, we show (i) the generated signal; (ii) the generated state mass; (iii)-(v) the mass of the approximate smoothed posterior of each state using 1, 2 and 10 components; and (vi) the exact case, which contains a maximum of 10, 100 components.

[9] Obtained from Fearnhead and Clifford (2003).