Switching Hierarchical Gaussian Filter

İsmail Şenöz, Albert Podusenko, Semih Akbayrak
Eindhoven University of Technology
Eindhoven, the Netherlands
{i.senoz,a.podusenko,s.akbayrak}@tue.nl

Christoph Mathys
Aarhus University
Aarhus, Denmark
chmathys@cas.au.dk

Bert de Vries
TU Eindhoven & GN Hearing
Eindhoven, the Netherlands
bert.de.vries@tue.nl

I. Extra notational convention

Throughout the derivations we deal with unnormalized distributions. We note that the messages are not required to be normalized. Although all marginals are required to be normalized.

II. Marginal Computations

A. Message update rules for $s_t^{(i)}$

In order to compute the marginal for $s_t^{(i)}$ the following computations are required.

$$q\left(s_t^{(i)}\right) \propto \overrightarrow{\nu}\left(s_t^{(i)}\right) \nu\!\uparrow\!\left(s_t^{(i)}\right) \overleftarrow{\nu}\left(s_t^{(i)}\right) \tag{1a}$$

$$\overrightarrow{\nu}\left(s_t^{(i)}\right) \propto \sum_{s_{t-1}^{(i)}} \exp\left(\mathbb{E}_{\setminus q\left(s_t^{(i)}, s_{t-1}^{(i)}\right)}\left[\log p\left(s_t^{(i)} | s_{t-1}^{(i)}, \boldsymbol{A}^{(i)}\right)\right]\right) \overrightarrow{\nu}\left(s_{t-1}^{(i)}\right) \tag{1b}$$

$$\overleftarrow{\nu}\left(s_t^{(i)}\right) \propto \sum_{s_{t+1}^{(i)}} \exp\left(\mathbb{E}_{\setminus q\left(s_{t+1}^{(i)}, s_t^{(i)}\right)}\left[\log p\left(s_{t+1}^{(i)} | s_t^{(i)}, \boldsymbol{A}^{(i)}\right)\right]\right) \overleftarrow{\nu}\left(s_{t+1}^{(i)}\right) \tag{1c}$$

$$\nu\!\uparrow\!\left(s_t^{(i)}\right) \propto \exp\left(\mathbb{E}_{\setminus q\left(s_t^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_t^{(i)}=m\right]}\right]\right). \tag{1d}$$

function $[x = y]$ evaluates to 1 if x is equal to y and 0 otherwise.

Firstly we address the computation of forward message (1a). We assume that the messages from the previous and future time step $\overrightarrow{\nu}\left(s_{t-1}^{(i)}\right)$ and $\overleftarrow{\nu}\left(s_{t+1}^{(i)}\right)$ are Categorical and the marginal of the transition matrix $q_t\left(\boldsymbol{A}^{(i)}\right)$ follows Dirichlet distribution, i.e.

$$\overrightarrow{\nu}\left(s_{t-1}^{(i)}\right) \propto \prod_{m=1}^{M_i} \left(\overrightarrow{\pi}_{t-1,m}^{(i)}\right)^{\left[s_{t-1}^{(i)}=m\right]} \tag{2a}$$

$$\overleftarrow{\nu}\left(s_{t+1}^{(i)}\right) \propto \prod_{m=1}^{M_i} \left(\overleftarrow{\pi}_{t+1,m}^{(i)}\right)^{\left[s_{t+1}^{(i)}=m\right]} \tag{2b}$$

$$q_t\left(\boldsymbol{A}^{(i)}\right) \propto \prod_{m=1}^{M_i} \prod_{k=1}^{M_i} \left(\alpha_{km}^{(i)}\right)^{\alpha_{t,km}^{(i)}-1}. \tag{2c}$$

The choices of (2a), (2b) and (2c) are motivated by conjugacy properties. To compute the forward message we determine an intermediate expectation form

$$\mathbb{E}_{\setminus q\left(s_t^{(i)}, s_{t-1}^{(i)}\right)}\left[\log p\left(s_t^{(i)} | s_{t-1}^{(i)}, \boldsymbol{A}^{(i)}\right)\right] = \mathbb{E}_{q_t\left(\boldsymbol{A}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \prod_{k=1}^{M_i} \left(\alpha_{km}^{(i)}\right)^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]}\right] \tag{3a}$$

$$= \sum_m \sum_k \mathbb{E}_{q_t\left(\boldsymbol{A}^{(i)}\right)}\left[\log \left(\alpha_{km}^{(i)}\right)^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]}\right] \tag{3b}$$

$$= \sum_m \sum_k \mathbb{E}_{q_t\left(\boldsymbol{A}^{(i)}\right)}\left[\log \left(\alpha_{km}^{(i)}\right)\right]^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]} \tag{3c}$$

$$= \sum_m \sum_k \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]}. \tag{3d}$$

where ψ denotes digamma function [1].

Substituting (3d) into (1b) we obtain

$$\overrightarrow{\nu}\left(s_t^{(i)}\right) \propto \sum_{s_{t-1}^{(i)}} \exp\left(\sum_m \sum_k \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]}\right) \overrightarrow{\nu}\left(s_{t-1}^{(i)}\right) \tag{4a}$$

$$\propto \sum_{s_{t-1}^{(i)}} \exp\left(\sum_m \sum_k \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)^{\left[s_t^{(i)}=k, s_{t-1}^{(i)}=m\right]}\right) \exp\left(\log\left(\overrightarrow{\pi}_{t-1,m}^{(i)}\right)^{\left[s_{t-1}^{(i)}=m\right]}\right) \tag{4b}$$

$$\propto \exp\left(\sum_k \left(\sum_m \log \overrightarrow{\pi}_{t-1,m}^{(i)} \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)\right)^{\left[s_t^{(i)}=k\right]}\right) \tag{4c}$$

$$\propto \prod_k \exp\left(\sum_m \log \overrightarrow{\pi}_{t-1,m}^{(i)} \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)\right)^{\left[s_t^{(i)}=k\right]} \tag{4d}$$

$$\propto \prod_k \left(\overrightarrow{\pi}_{t,k}^{(i)}\right)^{\left[s_t^{(i)}=k\right]} \tag{4e}$$

$$\overrightarrow{\pi}_{t,k}^{(i)} \triangleq \exp\left(\sum_m \log \overrightarrow{\pi}_{t-1,m}^{(i)} \left(\psi\left(\alpha_{t,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t,jm}^{(i)}\right)\right)\right) . \tag{4f}$$

Equation (4d) means that the forward message has a functional form proportional to a Categorical distribution. Following the steps in the derivation of the forward message (1c), we obtain the backward message proportional to a Categorical distribution.

$$\overleftarrow{\nu}\left(s_t^{(i)}\right) \propto \sum_{s_{t+1}^{(i)}} \exp\left(\sum_m \sum_k \left(\psi\left(\alpha_{t+1,km}^{(i)}\right) - \psi\left(\sum_j \alpha_{t+1,jm}^{(i)}\right)\right)^{\left[s_{t+1}^{(i)}=m, s_t^{(i)}=k\right]}\right) \overleftarrow{\nu}\left(s_{t+1}^{(i)}\right) \tag{5a}$$

$$\propto \prod_k \left(\overleftarrow{\pi}_{t,k}^{(i)}\right)^{\left[s_t^{(i)}=k\right]} \tag{5b}$$

$$\overleftarrow{\pi}_{t,k}^{(i)} \triangleq \exp\left(\sum_m \log \overleftarrow{\pi}_{t+1,m}^{(i)} \left(\psi\left(\alpha_{t+1,mk}^{(i)}\right) - \psi\left(\sum_j \alpha_{t+1,kj}^{(i)}\right)\right)\right) . \tag{5c}$$

Lastly we compute (1d). We assume that the following marginals are available

$$q\left(x_t^{(i)}, x_{t-1}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{x}_{t,t-1}^{(i)} | \boldsymbol{m}_{t,t-1}^{(i)}, \boldsymbol{\Sigma}_{t,t-1}^{(i)}\right) \tag{6a}$$

$$q\left(x_t^{(i+1)}\right) \propto \mathcal{N}\left(x_t^{(i+1)} | m_t^{(i+1)}, v_t^{(i+1)}\right) \tag{6b}$$

$$q_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{\kappa}^{(i)} | \boldsymbol{\mu}_t^{(i)}, \boldsymbol{\Omega}_t^{(i)}\right) \tag{6c}$$

$$q_t\left(\boldsymbol{\omega}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{\omega}^{(i)} | \boldsymbol{\vartheta}_t^{(i)}, \boldsymbol{\Xi}_t^{(i)}\right) . \tag{6d}$$

Computation of the message requires the following expectation quantities

$$\eta_{t,j}^{(i)} \triangleq \mathbb{E}\left[\kappa_j^{(i)} x_t^{(i+1)} + \omega_j^{(i)}\right] = \left(\boldsymbol{\mu}_t^{(i)}\right)_j m_t^{(i+1)} + \left(\boldsymbol{\vartheta}_t^{(i)}\right)_j \tag{7a}$$

$$\zeta_t^{(i)} \triangleq \mathbb{E}\left[\left(x_t^{(i)} - x_{t-1}^{(i)}\right)^2\right] = \left(\left(\boldsymbol{m}_{t,t-1}^{(i)}\right)_1 - \left(\boldsymbol{m}_{t,t-1}^{(i)}\right)_2\right)^2 + \left(\boldsymbol{\Sigma}_{t,t-1}^{(i)}\right)_{11} + \left(\boldsymbol{\Sigma}_{t,t-1}^{(i)}\right)_{22} - \left(\boldsymbol{\Sigma}_{t,t-1}^{(i)}\right)_{12} - \left(\boldsymbol{\Sigma}_{t,t-1}^{(i)}\right)_{21} \tag{7b}$$

$$\mathbb{E}\left[\exp\left(\kappa_j^{(i)} x_t^{(i+1)} + \omega_j^{(i)}\right)^{-1}\right] \approx \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right) \tag{7c}$$

$$\gamma_{t,j}^{(i)} \triangleq -\left(\boldsymbol{\mu}_t^{(i)}\right)_j m_t^{(i+1)} + 0.5\left(\left(\boldsymbol{\mu}_t^{(i)}\right)_j^2 v_t^{(i+1)} + \left(\boldsymbol{\Omega}_t^{(i)}\right)_{jj}\left(m_t^{(i+1)}\right)^2 + v_t^{(i+1)}\left(\boldsymbol{\Omega}_t^{(i)}\right)_{jj}^2\right) \tag{7d}$$

$$\beta_{t,j}^{(i)} \triangleq -\left(\boldsymbol{\vartheta}_t^{(i)}\right)_j + 0.5\left(\boldsymbol{\Xi}_t^{(i)}\right)_{jj} \tag{7e}$$

Using the results from (7), the message (1d) is determined as Categorical. Eq. (7c) uses results from [2, Section 2] showing that multiplication of two Gaussian random variables can be approximated by a Gaussian.

$$\nu\!\uparrow\!\left(s_t^{(i)}\right) \propto \mathbb{E}_{\setminus q\left(s_t^{(i)}\right)}\left[\log \prod_{j=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_j^{(i)}, \omega_j^{(i)}\right)\right)^{\left[s_t^{(i)}=j\right]}\right] \tag{8a}$$

$$= \prod_j \exp\left(-0.5\left(\eta_{t,j}^{(i)} + \zeta_t^{(i)} \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right)\right)\right)^{\left[s_t^{(i)}=j\right]}. \tag{8b}$$

Due to exponential family being closed under multiplication, we know that the marginal distribution computed by the multiplication of 3 un-normalized Categorical messages (1a), after normalization, will be a Categorical distribution (9).

$$q\left(s_t^{(i)}\right) = \frac{\overrightarrow{\nu}\left(s_t^{(i)}\right)\nu\!\uparrow\!\left(s_t^{(i)}\right)\overleftarrow{\nu}\left(s_t^{(i)}\right)}{\sum_{s_t^{(i)}} \overrightarrow{\nu}\left(s_t^{(i)}\right)\nu\!\uparrow\!\left(s_t^{(i)}\right)\overleftarrow{\nu}\left(s_t^{(i)}\right)} \tag{9a}$$

$$= \frac{\prod_{j=1}^{M_i}\left(\overrightarrow{\pi}_{t,j}^{(i)}\overleftarrow{\pi}_{t,j}^{(i)} \exp\left(-0.5\left(\eta_{t,j}^{(i)} + \zeta_t^{(i)} \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right)\right)\right)\right)^{\left[s_t^{(i)}=j\right]}}{\sum_{j=1}^{M_i} \overrightarrow{\pi}_{t,j}^{(i)}\overleftarrow{\pi}_{t,j}^{(i)} \exp\left(-0.5\left(\eta_{t,j}^{(i)} + \zeta_t^{(i)} \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right)\right)\right)} \tag{9b}$$

$$= \prod_j \left(\pi_{t,j}^{(i)}\right)^{\left[s_t^{(i)}=j\right]} \tag{9c}$$

$$\pi_{t,j}^{(i)} \triangleq \frac{\left(\overrightarrow{\pi}_{t,j}^{(i)}\overleftarrow{\pi}_{t,j}^{(i)} \exp\left(-0.5\left(\eta_{t,j}^{(i)} + \zeta_t^{(i)} \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right)\right)\right)\right)}{\sum_{j=1}^{M_i} \overrightarrow{\pi}_{t,j}^{(i)}\overleftarrow{\pi}_{t,j}^{(i)} \exp\left(-0.5\left(\eta_{t,j}^{(i)} + \zeta_t^{(i)} \exp\left(\gamma_{t,j}^{(i)} + \beta_{t,j}^{(i)}\right)\right)\right)}. \tag{9d}$$

This concludes the marginal computation for the discrete states $s_t^{(i)}$.

B. Message update rules for $x_t^{(i)}$

Required message computations for the continuous hierarchical states $x_t^{(i)}$

$$q\left(x_t^{(i)}\right) \propto \overrightarrow{\nu}\left(x_t^{(i)}\right)\nu\!\uparrow\!\left(x_t^{(i)}\right)\overleftarrow{\nu}\left(x_t^{(i)}\right) \tag{10a}$$

$$\overrightarrow{\nu}\left(x_t^{(i)}\right) \propto \int \exp\left(\mathbb{E}_{\setminus q\left(x_t^{(i)}, x_{t-1}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_t^{(i)}=m\right]}\right]\right) \overrightarrow{\nu}\left(x_{t-1}^{(i)}\right) \mathrm{d}x_{t-1}^{(i)} \tag{10b}$$

$$\overleftarrow{\nu}\left(x_t^{(i)}\right) \propto \int \exp\left(\mathbb{E}_{\setminus q\left(x_{t+1}^{(i)}, x_t^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_{t+1}^{(i)} | x_t^{(i)}, g_t^{(i)}\left(x_{t+1}^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_{t+1}^{(i)}=m\right]}\right]\right) \overleftarrow{\nu}\left(x_{t-1}^{(i)}\right) \mathrm{d}x_{t-1}^{(i)} \tag{10c}$$

$$\nu\!\uparrow\!\left(x_t^{(i)}\right) \propto \exp\left(\mathbb{E}_{\setminus q\left(x_t^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i-1)} | x_{t-1}^{(i-1)}, g_t^{(i-1)}\left(x_t^{(i)}, \kappa_m^{(i-1)}, \omega_m^{(i-1)}\right)\right)^{\left[s_t^{(i-1)}=m\right]}\right]\right). \tag{10d}$$

In order to compute the marginal, we need to make assumptions on the distributions and messages that are need to compute (10). We assume that

$$q\left(x_t^{(i-1)}, x_{t-1}^{(i-1)}\right) \propto \mathcal{N}\left(\boldsymbol{x}_{t,t-1}^{(i+1)} | \boldsymbol{m}_{t,t-1}^{(i+1)}, \boldsymbol{\Sigma}_{t,t-1}^{(i+1)}\right) \tag{11a}$$

$$q\left(x_t^{(i+1)}\right) \propto \mathcal{N}\left(x_t^{(i+1)} | m_t^{(i+1)}, v_t^{(i+1)}\right) \tag{11b}$$

$$q_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{\kappa}^{(i)} | \boldsymbol{\mu}_t^{(i)}, \boldsymbol{\Omega}_t^{(i)}\right) \tag{11c}$$

$$q_t\left(\boldsymbol{\omega}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{\omega}^{(i)} | \boldsymbol{\vartheta}_t^{(i)}, \boldsymbol{\Xi}_t^{(i)}\right) \tag{11d}$$

$$q\left(s_t^{(i)}\right) = \prod_{k=1}^{M_i} \left(\pi_{t,k}^{(i)}\right)^{\left[s_t^{(i)}=k\right]} \tag{11e}$$

$$\overrightarrow{\nu}\left(s_{t-1}^{(i)}\right) \propto \mathcal{N}\left(s_{t-1}^{(i)} | \overrightarrow{m}_{t-1}^{(i)}, \overrightarrow{v}_{t-1}^{(i)}\right) \tag{11f}$$

$$\overleftarrow{\nu}\left(s_t^{(i)}\right) \propto \mathcal{N}\left(s_t^{(i)} | \overleftarrow{m}_t^{(i)}, \overleftarrow{v}_t^{(i)}\right). \tag{11g}$$

Firstly, we compute the forward message (10b). In order to obtain the message we need determine the following expectation

$$\begin{aligned}
&\mathbb{E}_{\backslash q\left(x_t^{(i)}, x_{t-1}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_t^{(i)}=m\right]}\right] \\
&= \sum_{m=1}^{M_i} \pi_{t,m}^{(i)} \mathbb{E}\left[\log \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)\right] &(12a) \\
&\approx \sum_m \pi_{t,m}^{(i)} \log \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, \exp\left(\gamma_{t,m}^{(i)} + \beta_{t,m}^{(i)}\right)^{-1}\right) &(12b)
\end{aligned}$$

Plugging (12b) into (10b) we obtain

$$\begin{aligned}
\overrightarrow{\nu}\left(x_t^{(i)}\right) &\propto \int \exp\left(\sum_m \pi_{t,m}^{(i)} \log \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, \exp\left(\gamma_{t,m}^{(i)} + \beta_{t,m}^{(i)}\right)^{-1}\right)\right) \overrightarrow{\nu}\left(x_{t-1}^{(i)}\right) \mathrm{d}x_{t-1}^{(i)} &(13a) \\
&\propto \mathcal{N}\left(x_t^{(i)} | \overrightarrow{m}_{t-1}^{(i)}, \overrightarrow{v}_{t-1}^{(i)} + \sum_{k=1}^{M_i} \pi_{t,k}^{(i)} \exp\left(\gamma_{t,k}^{(i)} + \beta_{t,k}^{(i)}\right)^{-1}\right). &(13b)
\end{aligned}$$

Backwards message (10c) has the identical form with (10b) and it is evaluated as

$$\begin{aligned}
\overleftarrow{\nu}\left(x_t^{(i)}\right) &\propto \int \exp\left(\sum_m \pi_{t,m}^{(i)} \log \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, \exp\left(\gamma_{t,m}^{(i)} + \beta_{t,m}^{(i)}\right)^{-1}\right)\right) \overleftarrow{\nu}\left(x_{t+1}^{(i)}\right) \mathrm{d}x_{t+1}^{(i)} &(14a) \\
&\propto \mathcal{N}\left(x_t^{(i)} | \overleftarrow{m}_{t+1}^{(i)}, \overleftarrow{v}_{t+1}^{(i)} + \sum_{k=1}^{M_i} \pi_{t,k}^{(i)} \exp\left(\gamma_{t,k}^{(i)} + \beta_{t,k}^{(i)}\right)^{-1}\right). &(14b)
\end{aligned}$$

Forward (13b) and backward (14b) messages have variances that are composed of mixture terms. These mixture terms are scaled by the categorical probabilities. In order to enforce ”pure” switching behaviour, we can modify the VMP updates for the backward and forward message by constraining the categorical variable $s_t^{(i)}$ to have a Kronecker delta distribution centered around the mode of $q\left(s_t^{(i)}\right)$. We first determine the mode of the approximate posterior distribution for the categorical variable

$$\tilde{s}_t^{(i)} = \arg\max_{s_t^{(i)}} q\left(s_t^{(i)}\right) \tag{15a}$$

$$\tilde{q}\left(s_t^{(i)}\right) = \delta\left(s_t^{(i)} - \tilde{s}_t^{(i)}\right). \tag{15b}$$

Then we use the point mass distribution to evaluate (12b) as

$$\mathbb{E}_{\backslash q\left(x_t^{(i)}, x_{t-1}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_t^{(i)}=m\right]}\right] \approx \log \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, \exp\left(\gamma_{t,\tilde{s}_t^{(i)}}^{(i)} + \beta_{t,\tilde{s}_t^{(i)}}^{(i)}\right)^{-1}\right). \tag{16}$$

Using (16) we can modify the forward and backward message such that the variance information corresponds to selecting the parameters that come from the most probable switch. This means that, instead of mixture of components in the variance terms of (13b) and (14b) there is only one parameter regime influencing the message.

$$\overrightarrow{\nu}\left(x_t^{(i)}\right) \propto \mathcal{N}\left(x_t^{(i)}|\overrightarrow{m}_{t-1}^{(i)}, \overrightarrow{v}_{t-1}^{(i)} + \exp\left(\gamma_{t,\tilde{s}_t^{(i)}}^{(i)} + \beta_{t,\tilde{s}_t^{(i)}}^{(i)}\right)^{-1}\right) \tag{17a}$$

$$\overleftarrow{\nu}\left(x_t^{(i)}\right) \propto \mathcal{N}\left(x_t^{(i)}|\overleftarrow{m}_{t-1}^{(i)}, \overleftarrow{v}_{t-1}^{(i)} + \exp\left(\gamma_{t,\tilde{s}_t^{(i)}}^{(i)} + \beta_{t,\tilde{s}_t^{(i)}}^{(i)}\right)^{-1}\right). \tag{17b}$$

Depending on the desired behaviour , message updates for the continuous valued hierarchical states can be changed and used accordingly. Next we compute the message towards continuous valued hierarchical states propagated from the lower layers.

Firstly, we evaluate the following expectation

$$\mathbb{E}_{\backslash q\left(x_t^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i-1)}|x_{t-1}^{(i-1)}, g_t^{(i-1)}\left(x_t^{(i)}, \kappa_m^{(i-1)}, \omega_m^{(i-1)}\right)\right)^{\left[s_t^{(i-1)}=m\right]}\right] = \tag{18a}$$

$$\sum_m \pi_{t,m}^{(i)} \mathbb{E}\left[\log \mathcal{N}\left(x_t^{(i-1)}|x_{t-1}^{(i-1)}, g_t^{(i-1)}\left(x_t^{(i)}, \kappa_m^{(i-1)}, \omega_m^{(i-1)}\right)\right)\right]. \tag{18b}$$

Evaluation requires the following intermediate result

$$\mathbb{E}_{\backslash q\left(x_t^{(i)}\right)}\left[\exp\left(\kappa_j^{(i-1)} x_t^{(i)} + \omega_j^{(i-1)}\right)^{-1}\right] \approx \exp\left(-\left(\boldsymbol{\mu}_t^{(i-1)}\right)_j x_t^{(i)} + 0.5\left(x_t^{(i)}\right)^2\left(\boldsymbol{\Omega}_t^{(i-1)}\right)_{jj}\right) \tag{19}$$

Using this result we can write

$$\nu\uparrow\left(x_t^{(i)}\right) \propto \exp\left(\mathbb{E}_{\backslash q\left(x_t^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i-1)}|x_{t-1}^{(i-1)}, g_t^{(i-1)}\left(x_t^{(i)}, \kappa_m^{(i-1)}, \omega_m^{(i-1)}\right)\right)^{\left[s_t^{(i-1)}=m\right]}\right]\right) \tag{20a}$$

$$\propto \exp\left(-0.5\sum_j \pi_{t,j}^{(i-1)}\left(\left(\boldsymbol{\mu}_t^{(i-1)}\right)_j x_t^{(i)} + \zeta_t^{(i)} \exp\left(-\left(\boldsymbol{\mu}_t^{(i-1)}\right)_j x_t^{(i)} + 0.5\left(x_t^{(i)}\right)^2\left(\boldsymbol{\Omega}_t^{(i-1)}\right)_{jj}\right)\right)\right). \tag{20b}$$

The functional form of (20b) means that $\nu\uparrow\left(x_t^{(i)}\right)$ is not in the exponential family and is composed as a mixture. Again, by constraining the categorical variable to have a point mass distribution centered at the MAP estimate of the variational distribution (15a) and (15b), it is possible to induce a switching behaviour as opposed to a mixture behaviour. As (20b) is not in the exponential family, the further approximations are needed to obtain the marginal (10a). We know that the forward and backward messages are Gaussian (14b) and (13b). This means that we can utilize these messages to approximate the marginal of $x_t^{(i)}$ with a Gaussian. This is achieved by employing Gauss-Hermite quadrature to evaluate the non-conjugate multiplication by moment matching. The reason we want to approximate with Gaussian can be motivated from different perspectives. First of all, keeping the functional form and trying to propagate the functional form is computationally challenging as with every functional form additional rules are needed.

C. Message update rules for $\boldsymbol{\omega}^{(i)}$

Required messages for computing the marginal of $\boldsymbol{\omega}^{(i)}$ are

$$q_t\left(\boldsymbol{\omega}^{(i)}\right) \propto \overrightarrow{\nu}\left(\boldsymbol{\omega}^{(i)}\right) \prod_{t'=1}^{t} \overleftarrow{\nu}_{t'}\left(\boldsymbol{\omega}^{(i)}\right) \tag{21a}$$

$$\overrightarrow{\nu}\left(\boldsymbol{\omega}^{(i)}\right) \propto p\left(\boldsymbol{\omega}^{(i)}\right) \tag{21b}$$

$$\overleftarrow{\nu}_t\left(\boldsymbol{\omega}^{(i)}\right) \propto \exp\left(\mathbb{E}_{\backslash q\left(\boldsymbol{\omega}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)}|x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{\left[s_t^{(i)}=m\right]}\right]\right). \tag{21c}$$

The forward message is the prior hence it does not require computation. We compute the only required message $\overleftarrow{\nu}_t\left(\boldsymbol{\omega}^{(i)}\right)$.

In order to compute the marginal, we need to make assumptions on the distributions and messages. We assume that

$$q\left(x_t^{(i)}, x_{t-1}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{x}_{t,t-1}^{(i)} | \boldsymbol{m}_{t,t-1}^{(i)}, \boldsymbol{\Sigma}_{t,t-1}^{(i)}\right) \tag{22a}$$

$$q\left(x_t^{(i+1)}\right) \propto \mathcal{N}\left(x_t^{(i+1)} | m_t^{(i+1)}, v_t^{(i+1)}\right) \tag{22b}$$

$$q_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \mathcal{N}\left(\boldsymbol{\kappa}^{(i)} | \boldsymbol{\mu}_t^{(i)}, \boldsymbol{\Omega}_t^{(i)}\right) \tag{22c}$$

$$q\left(s_t^{(i)}\right) = \prod_{k=1}^{M_i} \left(\pi_{t,k}^{(i)}\right)^{\left[s_t^{(i)}=k\right]} \tag{22d}$$

$$\tag{22e}$$

Using the above distributions we determine the backward message as

$$\overleftarrow{\nu}_t\left(\boldsymbol{\omega}^{(i)}\right) \propto \exp\left(-0.5 \sum_k \pi_{t,k}^{(i)} \left(\omega_k^{(i)} + \zeta_t^{(i)} \exp\left(-\omega_k^{(i)}\right)\right)\right) \tag{23}$$

D. Message update rules for $\boldsymbol{\kappa}^{(i)}$

Required messages for $\boldsymbol{\kappa}^{(i)}$

$$q_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \overrightarrow{\nu}\left(\boldsymbol{\kappa}^{(i)}\right) \prod_{t'=1}^{t} \overleftarrow{\nu}_{t'}\left(\boldsymbol{\kappa}^{(i)}\right) \tag{24a}$$

$$\overrightarrow{\nu}\left(\boldsymbol{\kappa}^{(i)}\right) \propto p\left(\boldsymbol{\kappa}^{(i)}\right) \tag{24b}$$

$$\overleftarrow{\nu}_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \exp\left(\mathbb{E}_{\backslash q\left(\boldsymbol{\kappa}^{(i)}\right)}\left[\log \prod_{m=1}^{M_i} \mathcal{N}\left(x_t^{(i)} | x_{t-1}^{(i)}, g_t^{(i)}\left(x_t^{(i+1)}, \kappa_m^{(i)}, \omega_m^{(i)}\right)\right)^{[s_t^{(i)}=m]}\right]\right). \tag{24c}$$

We evaluate the backward message

$$\overleftarrow{\nu}_t\left(\boldsymbol{\kappa}^{(i)}\right) \propto \exp\left(-0.5 \sum_k \pi_{t,k}^{(i)}\left(\kappa_k^{(i)} m_t^{(i+1)} + \zeta_t^{(i)} \exp\left(-m_t^{(i+1)}\kappa_k^{(i)} + 0.5 v_t^{(i+1)}\left(\kappa_k^{(i)}\right)^2\right)\right)\right) \tag{25}$$

E. Message update rules for $\boldsymbol{A}^{(i)}$

Required messages for $\boldsymbol{A}^{(i)}$ are given by

$$q_t\left(\boldsymbol{A}^{(i)}\right) \propto \overrightarrow{\nu}\left(\boldsymbol{A}^{(i)}\right) \prod_{t'=1}^{t} \overleftarrow{\nu}_{t'}\left(\boldsymbol{A}^{(i)}\right) \tag{26a}$$

$$\overrightarrow{\nu}\left(\boldsymbol{A}^{(i)}\right) \propto p\left(\boldsymbol{A}^{(i)}\right) \tag{26b}$$

$$\overleftarrow{\nu}_t\left(\boldsymbol{A}^{(i)}\right) \propto \exp\left(\mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[\log p\left(s_t^{(i)} | s_{t-1}^{(i)}, \boldsymbol{A}^{(i)}\right)\right]\right). \tag{26c}$$

For a sake of brevity let us rewrite the message $\overleftarrow{\nu}_t\left(\boldsymbol{A}^{(i)}\right)$ in a log domain.

$$\log \overleftarrow{\nu}_t\left(\boldsymbol{A}^{(i)}\right) = \mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[\log p\left(s_t^{(i)} | s_{t-1}^{(i)}, \boldsymbol{A}^{(i)}\right)\right] + const \tag{27}$$

$$= \mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[\log \prod_{k=1}^{M_i}\prod_{m=1}^{M_i}\left(\alpha_{km}^{(i)}\right)^{[s_t^{(i)}=k, s_{t-1}^{(i)}=m]}\right] + const = \sum_{k=1}^{M_i}\sum_{m=1}^{M_i} \mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[\log\left(\alpha_{km}^{(i)}\right)^{[s_t^{(i)}=k, s_{t-1}^{(i)}=m]}\right] + const \tag{28}$$

$$= \sum_{k=1}^{M_i}\sum_{m=1}^{M_i} \mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[\log\left(\alpha_{km}^{(i)}\right)\right]^{[s_t^{(i)}=k, s_{t-1}^{(i)}=m]} + const = \sum_{k=1}^{M_i}\sum_{m=1}^{M_i} \mathbb{E}_{\backslash q\left(\boldsymbol{A}^{(i)}\right)}\left[[s_t^{(i)}=k, s_{t-1}^{(i)}=m]\log\left(\alpha_{km}^{(i)}\right)\right] + const \tag{29}$$

$$= \sum_{k=1}^{M_i}\sum_{m=1}^{M_i} \rho_{k,m} \log\left(\alpha_{km}^{(i)}\right) + const \tag{30}$$

where $\rho_{k,m}$ are entries of contingency matrix $\mathbf{P}$ of, i.e. $q(s_t^{(i)}, s_{t-1}^{(i)}) \propto \mathcal{C}on\left(s_t^{(i)}, s_{t-1}^{(i)}|\mathbf{P}\right)$ Hence,

$$\overleftarrow{\nu}_t\left(\boldsymbol{A}^{(i)}\right) = \prod_{m=1}^{M_i} \frac{\Gamma\left(\sum_{k=1}^{M_i}(\rho_{k,m}^{(i)}+1)\right)}{\prod_{k=1}^{M_i}\Gamma\left(\rho_{k,m}^{(i)}+1\right)} \left(\prod_{k=1}^{M_i}\alpha_{km}^{(i)}\right)^{\rho_{k,m}^{(i)}} \tag{31}$$

or defining $\hat{\rho} = \rho + 1$ we can rewrite (32) as

$$\overleftarrow{\nu}_t\left(\boldsymbol{A}^{(i)}\right) = \prod_{m=1}^{M_i} \frac{\Gamma\left(\sum_{k=1}^{M_i}(\hat{\rho}_{k,m}^{(i)})\right)}{\prod_{k=1}^{M_i}\Gamma\left(\hat{\rho}_{k,m}^{(i)}\right)} \left(\prod_{k=1}^{M_i}\alpha_{km}^{(i)}\right)^{\hat{\rho}_{k,m}^{(i)}-1} \tag{32}$$

III. Non-conjugate multiplication with Gauss-Hermite Quadrature

In this section we explain multiplication of two non-conjugate messages, where one of the messages is a Gaussian (as in (10a)). Speaking generally, given a Gaussian message $f_1(x)$ and a non-conjugate message $f_2(x)$ we want to determine the normalized distribution corresponding to the multiplication,i.e.

$$p(x) = \frac{f_1(x)f_2(x)}{\underbrace{\int f_1(x)f_2(x)\mathrm{d}x}_{Z}} \tag{33}$$

Since we have access to the functional forms, the only quantity needed to be computed is the normalization constant. To do this we note that the underlying integral for the normalization constant is a Gaussian integral. Meaning that we can employ sigma-point integration methods to find a numerical answer to the normalization constant. We write the integral Z as:

$$Z \approx \frac{1}{\sqrt{\pi}} \sum_i \mathcal{W}^{(i)} f_1\left(m_x^{[f_2(x)]} + \phi^{(i)}\sqrt{2v_x^{[f_2(x)]}}\right), \tag{34}$$

where i indexes the quadrature weights $\mathcal{W}^{(i)}$ and abscissas $\phi^{(i)}$, while $m_x^{[f_2(x)]}$ and $v_x^{[f_2(x)]}$ denote the first and second moments of $x \sim f_2(x)$. To evaluate nth moment of x $p(x)$ we use quadrature:

$$\mathbb{E}_{\tilde{q}(x)}\left[x^n\right] \approx \frac{1}{Z\sqrt{\pi}} \sum_i \mathcal{W}^{(i)} \left(m_x^{[f_2(x)]} + \phi^{(i)}\sqrt{2v_x^{[f_2(x)]}}\right)^n f_1\left(m_x^{[f_2(x)]} + \phi^{(i)}\sqrt{2v_x^{[f_2(x)]}}\right) \tag{35}$$

Finally, we approximate the marginals $\tilde{q}(x)$ by a Gaussian distribution. We compute the mean and variance parameters using (35). As a result, using a quadrature approximation, we can keep propagating Gaussian messages even after receiving a non-conjugate message from a layer below.

References

[1] M. Tables, M. Abramowitz, I. Stegun, and A. G. Greenhill, "Handbook of mathematical functions with formulas, graphs,," 1971.

[2] L. A. Aroian, "The probability function of the product of two normally distributed variables," *The Annals of Mathematical Statistics*, vol. 18, no. 2, pp. 265–271, 1947.