Within this document nn = number of neurons
Negative Log-Posterior :

$$\mathcal{L} = logP\left(\bar{x}|y,\Theta\right) \quad = \quad +\frac{1}{2}log\left(|Q_0|\right) + \frac{(T-1)}{2}log\left(|Q|\right) - \sum_{t=1}^{T}\left(y_t^T\left(Cx_t+d\right) - \sum_{i=1}^{nn}exp\left[Cx_t+d\right]_i\right)$$
$$+\frac{1}{2}\left(x_1-m_0\right)^T Q_0^{-1}\left(x_1-m_0\right) + \frac{1}{2}\sum_{t=1}^{T-1}\left(x_{t+1}-Ax_t-Bu_t\right)^T Q^{-1}\left(x_{t+1}-Ax_t-Bu_t\right)$$

Negative Log-Posterior Derivative:

$$\frac{d\mathcal{L}}{dx_1} = -C^T y_1 + C^T \exp\left[Cx_1+d\right] + Q_0^{-1}\left(x_1-m_0\right) - A^T Q^{-1}\left(x_2-Ax_1-Bu_1\right)$$

$$\left.\frac{d\mathcal{L}}{dx_t}\right|_{1<t<T} = -C^T y_t + C^T\exp\left[Cx_t+d\right] - A^TQ^{-1}x_{t+1} + A^TQ^{-1}Ax_t + A^TQ^{-1}Bu_t + Q^{-1}x_t - Q^{-1}Ax_{t-1} - Q^{-1}Bu_{t-1}$$

$$\frac{d\mathcal{L}}{dx_T} = -C^Ty_T + C^T\exp\left[Cx_T+d\right] + Q^{-1}x_T - Q^{-1}Ax_{T-1} - Q^{-1}Bu_{T-1}$$

Negative Log-Posterior Hessian:

$$H_{1,1} = Q_0^{-1} + A^TQ^{-1}A + \sum_{i=1}^{nn}exp\left[C_ix_1+d_i\right]C_iC_i^T$$

$$H_{t,t}|_{1<t<T} = \sum_{i=1}^{nn}exp\left[C_ix_t+d_i\right]C_iC_i^T + A^TQ^{-1}A + Q^{-1}$$

$$H_{T,T}^T = Q^{-1} + \sum_{i=1}^{nn}exp\left[C_ix_T+d_i\right]C_iC_i^T$$

$$H_{t,t+1}|_{t<T} = -Q^{-1}A$$

$$H_{t+1,t}|_{t<T} = \left.H_{t,t+1}^T\right|_{t<T} = -A^TQ^{-1}$$

Joint Log Likelihood (Note that for training, terms below the first line RHS are irrelevant:

$$J = \sum_{t=1}^{T-1}\left[-y_t^TC\mu_t - y_t^Td + \sum_{i=1}^{nn}exp\left[C\mu_t+d+\frac{1}{2}diag\left(C^T\Sigma_{t,t}C\right)\right]_i\right.$$
$$+\frac{1}{2}\Big(\mu_{t+1}^TQ^{-1}\mu_{t+1} + Tr\left(Q^{-1}\Sigma_{t+1,t+1}\right) - \mu_{t+1}^TQ^{-1}A\mu_t - Tr\left(Q^{-1}A\Sigma_{t+1,t}\right) - \mu_{t+1}^TQ^{-1}Bu_t$$
$$-\mu_t^TA^TQ^{-1}\mu_{t+1} - Tr\left(A^TQ^{-1}\Sigma_{t,t+1}\right) + \mu_t^TA^TQ^{-1}A\mu_t + Tr\left(A^TQ^{-1}A\Sigma_{t,t}\right) + \mu_t^TA^TQ^{-1}Bu_t$$
$$\left.-u_{t+1}^TB^TQ^{-1}\mu_{t+1} + u_{t+1}^TB^TQ^{-1}A\mu_t + u_{t+1}^TB^TQ^{-1}Bu_t\Big)\right]$$
$$+\frac{1}{2}\left(log(|Q_0|) + (T-1)\,log(|Q|) + Tr(Q_0^{-1}\Sigma_{1,1}) + (\mu_1-x_0)^TQ_0^{-1}(\mu_1-x_0)\right)$$

Joint Log Likelihood dC Derivative:

$$\frac{dJ}{dd_i} = \sum_{t=1}^{T-1}\left(-y_{t,i} + exp(C_i\mu_t+d_i+\frac{1}{2}C_i^T\Sigma_{t,t}C_i))\right)$$

$$\frac{dJ}{dC_i} = \sum_{t=1}^{T-1}\left[-y_{t,i}\mu_t + exp\left[C_i\mu_t+d_i+\frac{1}{2}C_i^T\Sigma_{t,t}C_i\right]\left(\mu_t+\Sigma_{t,t}C_i\right)\right]$$

Joint Log Likelihood dC Hessian:

$$H_{d_i,d_i} = \sum_{t=1}^{T-1} exp\left[C_i\mu_t + d_i + \frac{1}{2}C_i^T\Sigma_{t,t}C_i\right]$$

$$H_{d_i,C_i} = \sum_{t=1}^{T-1} exp\left[C_i\mu_t + d_i + \frac{1}{2}C_i\Sigma_{t,t}C_i\right](\mu_t + \Sigma_{t,t}C_i)$$

$$H_{C_i,C_i} = \sum_{t=1}^{T-1} exp\left[C_i\mu_t + d_i + \frac{1}{2}C_i\Sigma_{t,t}C_i\right]\left(\Sigma_{t,t} + (\mu_t + \Sigma_{t,t}C_i)(\mu_t + \Sigma_{t,t}C_i)^T\right)$$

Variable Updates:

$$A = \left(\sum_{t=1}^{T-1}\left(\Sigma_{t+1,t} + \mu_{t+1}\mu_t^T - Bu_t\mu_t^T\right)\right)\left(\sum_{t=1}^{T-1}\left(\Sigma_{t,t} + \mu_t\mu_t^T\right)\right)^{-1}$$

$$B = \left(\sum_{t=1}^{T-1}\mu_{t+1}u_t^T - A\mu_t u_t^T\right)\left(\sum_{t=1}^{T-1}u_t u_t^T\right)^{-1}$$

$$\begin{aligned} Q \quad = \quad & \frac{1}{T-1}\sum_{t=1}^{T-1}\Big[\Sigma_{t+1,t+1} + \mu_{t+1}\mu_{t+1}^T - (\Sigma_{t+1,t} + \mu_{t+1}\mu_t^T)A^T - \mu_{t+1}u_t^T B^T - A(\Sigma_{t,t+1} + \mu_t\mu_{t+1}^T) \\ & + A(\Sigma_{t,t} + \mu_t\mu_t^T)A^T + A\mu_t u_t^T B^T - Bu_t\mu_{t+1}^T + Bu_t\mu_t^T A^T + Bu_t u_t^T B^T)\Big] \end{aligned}$$

Calculating A,B,Q:

Considering only JLL terms with Q (more conveniently Q^{-1}):

$$\frac{\partial}{\partial Q^{-1}}\left(\frac{1-T}{2}ln\left(\left|Q^{-1}\right|\right) + \frac{1}{2}\mathbb{E}\left(\sum_{t=1}^{T-1}\left(x_{t+1} - Ax_t - Bu_t\right)^T Q^{-1}\left(x_{t+1} - Ax_t - Bu_t\right)\right)\right) \stackrel{!}{=} 0$$

$$\frac{1-T}{2}Q + \frac{1}{2}\mathbb{E}\left(\sum_{t=1}^{T-1}\left(x_{t+1} - Ax_t - Bu_t\right)\left(x_{t+1} - Ax_t - Bu_t\right)^T\right) = 0$$

$$\begin{aligned} Q \quad = \quad & \frac{1}{T-1}\sum_{t=1}^{T-1}\Big(\mathbb{E}\left(x_{t+1}x_{t+1}^T\right)\text{-}\mathbb{E}\left(x_{t+1}x_t^T\right)A^T\text{-}\mathbb{E}\left(x_{t+1}u_t^T\right)B^T \\ & -A\mathbb{E}\left(x_t x_{t+1}^T\right) + A\mathbb{E}\left(x_t x_t^T\right)A^T + A\mathbb{E}\left(x_t u_t^T\right)B^T \\ & -B\mathbb{E}\left(u_t x_{t+1}^T\right) + B\mathbb{E}\left(u_t x_t^T\right)A^T + B\mathbb{E}\left(u_t u_t^T\right)B^T\Big) \end{aligned}$$

which yields the expression for Q above.

$$\frac{\partial}{\partial A_{ij}}\left(\frac{1}{2}\mathbb{E}\left(\sum_{t=1}^{T-1}\left(x_{t+1} - Ax_t - Bu_t\right)^T Q^{-1}\left(x_{t+1} - Ax_t - Bu_t\right)\right)\right) \quad \stackrel{!}{=} \quad 0$$

let $v = x_{t+1} - Ax_t - Bu_t$,

$$\frac{\partial v_k}{\partial A_{ij}}\frac{\partial}{\partial v_k}\left(\frac{1}{2}\mathbb{E}\left(\sum_{t=1}^{T-1}v^T Q^{-1}v\right)\right) \quad = \quad \mathbb{E}\left(\sum_{t=1}^{T-1}Q^{-1}vx_t^T\right)$$

so that

$$\begin{aligned} Q^{-1}\mathbb{E}\left(\sum_{t=1}^{T-1}\left(x_{t+1} - Ax_t - Bu_t\right)x_t^T\right) \quad &= \quad 0 \\ \sum_{t=1}^{T-1}\mathbb{E}\left(x_{t+1}x_t^T\right) - A\sum_{t=1}^{T-1}\mathbb{E}\left(x_t x_t^T\right) - B\sum_{t=1}^{T-1}\mathbb{E}\left(u_t x_t^T\right) \quad &= \quad 0 \end{aligned}$$

$$A = \left(\sum_{t=1}^{T-1} \mathbb{E}\left(x_t x_t^T\right) \right)^{-1} \left(\sum_{t=1}^{T-1} \left(\mathbb{E}\left(x_{t+1} x_t^T\right) - B\mathbb{E}\left(u_t x_t^T\right) \right) \right)$$

similarly for B:

$$B = \left(\sum_{t=1}^{T-1} \mathbb{E}\left(u_t u_t^T\right) \right)^{-1} \left(\sum_{t=1}^{T-1} \left(\mathbb{E}\left(x_{t+1} u_t^T\right) - A\mathbb{E}\left(x_t u_t^T\right) \right) \right)$$