Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models

Technical note

This note concerns variational Bayesian inversion of nonlinear state-space models of the form specified in the companion paper [Daunizeau et al., *Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models*, Physical D (2009)] :

$$
\begin{aligned}
& p\left(y_t \middle| x_t,\phi\right) = N\left(g\left(x_t,\phi\right),\sigma^{-1}Q_y\right) \\
& p\left(x_t \middle| x_{t-1},\theta\right) = N\left(f\left(x_t,\theta\right),\alpha^{-1}Q_x\right) \\
& p\left(x_0\right) = N\left(\mu_{00},\Sigma_{00}\right) \\
& p\left(\theta\right) = N\left(\mu_\theta,Q_{\theta 0}\right) \\
& p\left(\phi\right) = N\left(\mu_\phi,Q_{\phi 0}\right) \\
& p\left(\sigma\right) = Ga\left(a_\sigma,b_\sigma\right) \\
& p\left(\alpha\right) = Ga\left(a_\alpha,b_\alpha\right)
\end{aligned}
,
$$

where we dropped the dependency on known inputs u to the system.

This class of models is general and subsumes static nonlinear models (no time dependency for the hidden states x)[1].

This document explicitly focuses on the mathematical stages of VB calculus which lead to the algorithm as described in the paper.

[1] Alternatively, it may be argued that static models actually grand-father state-space models, in the sense that the set of (Markov) conditional independencies assumed in state-space models are *additional* assumption.

1. Observation parameters

$$\log q(\phi) = -\frac{\sigma}{2}\sum_{t=1}^{T}\left(y_t^{\mathrm{T}} Q_y^{-1} y_t - 2 y_t^{\mathrm{T}} Q_y^{-1}\left\langle g\left(x_t,\phi\right)\right\rangle + \left\langle g\left(x_t,\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\phi\right)\right\rangle\right) - \frac{1}{2}\left(\phi-\mu_{\phi 0}\right)^{\mathrm{T}} Q_{\phi 0}^{-1}\left(\phi-\mu_{\phi 0}\right)$$

Taylor-expand $g(x_t,\varphi)$ around μ_t:

$$g(x_t,\varphi) = g(\mu_t,\varphi) + \frac{\partial g}{\partial x}^{\mathrm{T}}(x_t-\mu_t) + O^{(2)}.$$

Then:

$$\left\langle g\left(x_t,\phi\right)\right\rangle = g\left(\mu_t,\phi\right)$$

$$\left\langle g\left(x_t,\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\phi\right)\right\rangle = g\left(\mu_t,\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\phi\right) + \underbrace{tr\left[\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \Psi_{t,t}\right]}_{j_x(\phi)}.$$

Taylor-expand $g(\mu_t,\varphi)$ around μ_φ:

$$g(\mu_t,\varphi) = g(\mu_t,\mu_\varphi) + \frac{\partial g}{\partial \varphi}^{\mathrm{T}}(\varphi-\mu_\varphi) + O^{(2)}$$

And Taylor-expand $j_x(\varphi)$ around μ_φ:

$$j_x(\varphi) = j_x(\mu_\varphi) + \frac{\partial j_x}{\partial \varphi}^{\mathrm{T}}(\varphi-\mu_\varphi) + \frac{1}{2}(\varphi-\mu_\varphi)^{\mathrm{T}}\frac{\partial^2 j_x}{\partial \varphi^2}(\varphi-\mu_\varphi) + O^{(3)}$$

Then:

$$\left\langle g\left(x_t,\phi\right)\right\rangle = g\left(\mu_t,\mu_\phi\right) + \frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right)$$

$$\begin{aligned}\left\langle g\left(x_t,\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\phi\right)\right\rangle &= g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) + 2\, g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) + \left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\ &+ j_x\left(\mu_\phi\right) + \frac{\partial j_x}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) + \frac{1}{2}\left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial^2 j_x}{\partial \phi^2}\left(\phi-\mu_\phi\right)\end{aligned}$$

Then the log variational posterior pdf writes:

$$\begin{aligned}\log q\left(\phi\right) = &-\frac{\sigma}{2}\sum_{t=1}^{T}\Bigg(y_t^{\mathrm{T}} Q_y^{-1} y_t - 2\, y_t^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) - 2\, y_t^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\ &+ g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) + 2\, g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) + \left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\ &+ j_x\left(\mu_\phi\right) + \frac{\partial j_x}{\partial \phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) + \frac{1}{2}\left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial^2 j_x}{\partial \phi^2}\left(\phi-\mu_\phi\right)\Bigg) \\ &-\frac{1}{2}\left(\phi-\mu_{\phi 0}\right)^{\mathrm{T}} Q_{\phi 0}^{-1}\left(\phi-\mu_{\phi 0}\right) \\ =& -\frac{1}{2}\phi^{\mathrm{T}}\left[\sigma\sum_{t=1}^{T}\left(\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_x}{\partial \phi^2}\right) + Q_{\phi 0}^{-1}\right]\phi \\ &-\frac{1}{2}\left[\sigma\sum_{t=1}^{T}\left(-2\, y_t^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + 2\, g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{\partial j_x}{\partial \phi}^{\mathrm{T}} - 2\mu_\phi^{\mathrm{T}}\left(\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_x}{\partial \phi^2}\right)\right) - 2\mu_{\phi 0}^{\mathrm{T}} Q_{\phi 0}^{-1}\right]\phi \\ =& -\frac{1}{2}\phi^{\mathrm{T}} B\phi - A^{\mathrm{T}}\phi\end{aligned}$$

Then the update rule for observation parameters writes:

$$\begin{aligned}\mu_\phi^{(k+1)} &= \Sigma_\phi^{(k+1)}\left[Q_{\phi 0}^{-1}\mu_{\phi 0} + \sigma\sum_{t=1}^{T}\left(\frac{\partial g}{\partial \phi} Q_y^{-1}\left(y_t - g\left(\mu_t,\mu_\phi\right)\right) - \frac{1}{2}\frac{\partial j_x}{\partial \phi} + \left(\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_x}{\partial \phi^2}\right)\mu_\phi^{(k)}\right)\right] \\ \Sigma_\phi^{(k+1)} &= \left[\sigma\sum_{t=1}^{T}\left(\frac{\partial g}{\partial \phi} Q_y^{-1}\frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_x}{\partial \phi^2}\right) + Q_{\phi 0}^{-1}\right]^{-1}\end{aligned}$$

We then have to disambiguate the terms that come from the mean-field approximation, which takes into account the posterior uncertainty of the hidden-states, i.e. $j_x\left(\varphi\right)$, defined as:

$$
\begin{aligned}
j_x(\phi) &= tr\left[\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \Psi_{t,t}\right] \\
&= tr\left[\frac{\partial g}{\partial x}^{\mathrm{T}} \Psi_{t,t} \frac{\partial g}{\partial x} Q_y^{-1}\right] \\
&= vec\left(\frac{\partial g}{\partial x}\right)^{\mathrm{T}} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] vec\left(\frac{\partial g}{\partial x}\right) \equiv g_x(\phi)^{\mathrm{T}} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] g_x(\phi)
\end{aligned},
$$

where $g_x(\varphi)$ is the vectorized gradient of the observation function $g(x_t,\varphi)$ w.r.t to the hidden-states x, evaluated at μ_t but still depending on φ.

To get a close form for the evaluation of the update rules of the variational pdf, we still have to resort to a further approximation of the gradient and Hessian of $j_x(\varphi)$. This is done by Taylor-expanding the function $g_x(\varphi)$, which means that we consider a bilinear approximation to the observation function:

$$
g_x(\varphi) = g_x(\mu_\varphi) + \frac{\partial g_x}{\partial \varphi}^{\mathrm{T}} (\varphi - \mu_\varphi) + O^{(2)}.
$$

Then, the trace term $j_x(\varphi)$ becomes:

$$
j_x(\phi) = g_x(\mu_\phi)^{\mathrm{T}} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] g_x(\mu_\phi) + 2 g_x(\mu_\phi)^{\mathrm{T}} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] \frac{\partial g_x}{\partial \phi}^{\mathrm{T}} (\phi - \mu_\phi) + (\phi - \mu_\phi) \frac{\partial g_x}{\partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] \frac{\partial g_x}{\partial \phi}^{\mathrm{T}} (\phi - \mu_\phi)
$$

This yields the expressions of its gradient and Hessian, evaluated at μ_φ:

$$
\begin{aligned}
\left.\frac{\partial j_x}{\partial \phi}\right|_{\phi=\mu_\phi} &= 2 \frac{\partial g_x}{\partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] g_x(\mu_\phi) \\
\left.\frac{\partial^2 j_x}{\partial \phi^2}\right|_{\phi=\mu_\phi} &= 2 \frac{\partial g_x}{\partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t}\right] \frac{\partial g_x}{\partial \phi}^{\mathrm{T}}
\end{aligned}
$$

Finally, the VB update for the observation parameters writes:

$$\mu_\phi^{(k+1)} - \mu_\phi^{(k)} = \Sigma_\phi^{(k+1)} \left[Q_{\phi 0}^{-1} \left(\mu_{\phi 0} - \mu_\phi^{(k)} \right) + \sigma \sum_{t=1}^{T} \left(\frac{\partial g}{\partial \phi} Q_y^{-1} \left(y_t - g\left(\mu_t, \mu_\phi \right) \right) - \frac{\partial^2 \tilde{g}}{\partial x \partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t} \right] \frac{\partial \tilde{g}}{\partial x} \right) \right]$$

$$\Sigma_\phi^{(k+1)} = \left[\sigma \sum_{t=1}^{T} \left(\frac{\partial g}{\partial \phi} Q_y^{-1} \frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{\partial^2 \tilde{g}}{\partial x \partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t} \right] \frac{\partial^2 \tilde{g}}{\partial x \partial \phi}^{\mathrm{T}} \right) + Q_{\phi 0}^{-1} \right]^{-1}$$

where the use of the notation $\tilde{g}$ refers to the hidden vectorization:

$$\frac{\partial \tilde{g}}{\partial x} = vec\left(\frac{\partial g}{\partial x} \right)$$

$$\frac{\partial^2 \tilde{g}}{\partial x \partial \varphi} = \frac{\partial}{\partial \varphi} \left[vec\left(\frac{\partial g}{\partial x} \right) \right]$$

2. Evolution parameters

$$\log q(\theta) = -\frac{1}{2}\alpha\sum_{t=0}^{T-1}\left(\left\langle x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}x_{t+1}\right\rangle - 2\left\langle x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}f\left(x_t,\theta\right)\right\rangle + \left\langle f\left(x_t,\theta\right)^{\mathrm{T}}Q_x{}^{-1}f\left(x_t,\theta\right)\right\rangle\right)$$
$$-\frac{1}{2}\left(\theta-\mu_{\theta 0}\right)^{\mathrm{T}}Q_{\theta 0}{}^{-1}\left(\theta-\mu_{\theta 0}\right)$$

Taylor-expand $f(x_t,\theta)$ around μ_t:

$$f(x_t,\theta) = f(\mu_t,\theta) + \frac{\partial f}{\partial x}^{\mathrm{T}}(x_t-\mu_t) + O^{(2)}.$$

Then:

$$\begin{aligned}\left\langle x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}f\left(x_t,\theta\right)\right\rangle &= \left\langle x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_t,\theta\right) + x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}x_t - x_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_t\right\rangle \\ &= \mu_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_t,\theta\right) + \mu_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_t + \underbrace{tr\left[Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t+1}\right]}_{h^{(1)}{}_x(\theta)} - \mu_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_t\,, \\ &= \mu_{t+1}{}^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_t,\theta\right) + h_x^{(1)}(\theta)\end{aligned}$$

knowing that, for the initial condition ($t=0$), $h_x^{(1)}(\theta)=0$, due to the mean-field approximation between x_0 and the time series $x_{1:n_t}$ (i.e. $\Psi_{1,0} := 0$).

And also:

$$\left\langle f\left(x_t,\theta\right)^{\mathrm{T}}Q_x{}^{-1}f\left(x_t,\theta\right)\right\rangle = f\left(\mu_t,\theta\right)^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_t,\theta\right) + \underbrace{tr\left[\frac{\partial f}{\partial x}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t}\right]}_{h^{(2)}{}_x(\theta)}.$$

Taylor-expand $f(\mu_t,\theta)$ around μ_θ:

$$f(\mu_t,\theta)=f(\mu_t,\mu_\theta)+\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+O^{(2)}$$

And Taylor-expand $h_x^{(1)}(\theta)$ and $h_x^{(2)}(\theta)$ around μ_θ:

$$\begin{aligned}
h_x^{(1)}(\theta)&=h_x^{(1)}(\mu_\theta)+\frac{\partial h_x^{(1)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+\frac{1}{2}(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(1)}}{\partial\theta^2}(\theta-\mu_\theta)+O^{(3)}\\
h_x^{(2)}(\theta)&=h_x^{(2)}(\mu_\theta)+\frac{\partial h_x^{(2)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+\frac{1}{2}(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}(\theta-\mu_\theta)+O^{(3)}
\end{aligned}$$

Then:

$$\begin{aligned}
\left\langle x_{t+1}{}^{\mathrm{T}}Q_x^{-1}f(x_t,\varphi)\right\rangle=\mu_{t+1}{}^{\mathrm{T}}Q_x^{-1}f(\mu_t,\mu_\theta)+\mu_{t+1}{}^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)\\
+h_x^{(1)}(\mu_\theta)+\frac{\partial h_x^{(1)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+\frac{1}{2}(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(1)}}{\partial\theta^2}(\theta-\mu_\theta)
\end{aligned}$$

And also:

$$\begin{aligned}
\left\langle f(x_t,\theta)^{\mathrm{T}}Q_x^{-1}f(x_t,\theta)\right\rangle=f(\mu_t,\mu_\theta)^{\mathrm{T}}Q_x^{-1}f(\mu_t,\mu_\theta)+2\,f(\mu_t,\mu_\theta)^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial f}{\partial\theta}Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)\\
+h_x^{(2)}(\mu_\theta)+\frac{\partial h_x^{(2)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta)+\frac{1}{2}(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}(\theta-\mu_\theta)
\end{aligned}$$

.

Then the log variational posterior pdf writes:

$$
\begin{aligned}
\log q(\theta) = & -\frac{\alpha}{2}\sum_{t=0}^{T-1}\Bigg(-2\mu_{t+1}{}^{\mathrm{T}} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) - 2\frac{\partial h_x^{(1)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) - (\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(1)}}{\partial\theta^2}(\theta-\mu_\theta) \\
& + 2 f(\mu_t,\mu_\theta)^{\mathrm{T}} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) + (\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial f}{\partial\theta} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) \\
& + h_x^{(2)}(\mu_\theta) + \frac{\partial h_x^{(2)}}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) + \frac{1}{2}(\theta-\mu_\theta)^{\mathrm{T}}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}(\theta-\mu_\theta)\Bigg) \\
& -\frac{1}{2}(\theta-\mu_{\theta 0})^{\mathrm{T}} Q_{\theta 0}{}^{-1}(\theta-\mu_{\theta 0}) \\
= & -\frac{1}{2}\theta^{\mathrm{T}}\left[\alpha\sum_{t=0}^{T-1}\left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2} - \frac{\partial^2 h_x^{(1)}}{\partial\theta^2}\right) + Q_{\theta 0}{}^{-1}\right]\theta \\
& -\frac{1}{2}\Bigg[\alpha\sum_{t=0}^{T-1}\Bigg(2\left(f(\mu_t,\mu_\theta)^{\mathrm{T}} - \mu_{t+1}{}^{\mathrm{T}} - \mu_\theta{}^{\mathrm{T}}\frac{\partial f}{\partial\theta}\right) Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} - 2\frac{\partial h_x^{(1)}}{\partial\theta}^{\mathrm{T}} + 2\mu_\theta{}^{\mathrm{T}}\left(\frac{\partial^2 h_x^{(1)}}{\partial\theta^2} - \frac{1}{2}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}\right) + \frac{\partial h_x^{(2)}}{\partial\theta}^{\mathrm{T}}\Bigg) \\
& - 2\mu_{\theta 0}{}^{\mathrm{T}} Q_{\theta 0}{}^{-1}\Bigg]\theta \\
= & -\frac{1}{2}\theta^{\mathrm{T}} B\theta - A^{\mathrm{T}}\theta
\end{aligned}
$$

Using the quadratic form (QF) transform, the update rules write:

$$
\begin{aligned}
\mu_\theta{}^{(k+1)} &= \Sigma_\theta{}^{(k+1)}\left[Q_{\theta 0}{}^{-1}\mu_{\theta 0} + \alpha\sum_{t=0}^{T-1}\left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\left(\mu_{t+1} - f(\mu_t,\mu_\theta)^{\mathrm{T}}\right) + \left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} - \frac{\partial^2 h_x^{(1)}}{\partial\theta^2} + \frac{1}{2}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}\right)\mu_\theta{}^{(k)} + \frac{\partial h_x^{(1)}}{\partial\theta} - \frac{1}{2}\frac{\partial h_x^{(2)}}{\partial\theta}\right)\right] \\
\Sigma_\theta{}^{(k+1)} &= \left[\alpha\sum_{t=1}^{T}\left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} - \frac{\partial^2 h_x^{(1)}}{\partial\theta^2} + \frac{1}{2}\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}\right) + Q_{\theta 0}{}^{-1}\right]^{-1}
\end{aligned}
$$

The first line may be rewritten as a formal update rule the following way:

$$
\begin{aligned}
\mu_\theta{}^{(k+1)} &= \Sigma_\theta{}^{(k+1)}\left[\Sigma_\theta{}^{(k+1)-1}\mu_\theta{}^{(k)} + Q_{\theta 0}{}^{-1}\left(\mu_{\theta 0} - \mu_\theta{}^{(k)}\right) + \alpha\sum_{t=0}^{T-1}\left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\left(\mu_{t+1} - f(\mu_t,\mu_\theta)^{\mathrm{T}}\right) + \frac{\partial h_x^{(1)}}{\partial\theta} - \frac{1}{2}\frac{\partial h_x^{(2)}}{\partial\theta}\right)\right] \\
&= \mu_\theta{}^{(k)} + \Sigma_\theta{}^{(k+1)}\left[Q_{\theta 0}{}^{-1}\left(\mu_{\theta 0} - \mu_\theta{}^{(k)}\right) + \alpha\sum_{t=0}^{T-1}\left(\frac{\partial f}{\partial\theta} Q_x{}^{-1}\left(\mu_{t+1} - f(\mu_t,\mu_\theta)^{\mathrm{T}}\right) + \frac{\partial h_x^{(1)}}{\partial\theta} - \frac{1}{2}\frac{\partial h_x^{(2)}}{\partial\theta}\right)\right]
\end{aligned}
$$

We still have to provide the expressions of the trace terms $h_x^{(1)}(\theta)$ and $h_x^{(2)}(\theta)$, which come from the mean-field approximation. Here again, these are provided by a similar bilinear approximation than that detailed in section 1:

$$\begin{aligned} h_x^{(1)}(\theta) &= tr\left[Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t+1}\right] = vec\left(\Psi_{t,t+1}Q_x^{-1}\right)^{\mathrm{T}} vec\left(\frac{\partial f}{\partial x}\right) \equiv vec\left(\Psi_{t,t+1}Q_x^{-1}\right)^{\mathrm{T}} f_x(\theta) \\ h_x^{(2)}(\theta) &= tr\left[\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t}\right] = vec\left(\frac{\partial f}{\partial x}\right)^{\mathrm{T}}\left[Q_x^{-1}\otimes\Psi_{t,t}\right] vec\left(\frac{\partial f}{\partial x}\right) \equiv f_x(\theta)^{\mathrm{T}}\left[Q_x^{-1}\otimes\Psi_{t,t}\right] f_x(\theta) \end{aligned},$$

where $f_x(\theta)$ is the vectorized gradient of the evolution function $f(x_t,\theta)$ w.r.t to the hidden-states x, evaluated at μ_t but still depending on θ : $f_x(\theta) := vec\left(\frac{\partial f}{\partial x}\right)$.

We then Taylor-expand $f_x(\theta)$ (bilinear approximation to the evolution function):

$$f_x(\theta) = f_x(\mu_\theta) + \frac{\partial f_x}{\partial \theta}^{\mathrm{T}}(\theta - \mu_\theta) + O^{(2)}.$$

Then, the trace terms $h_x^{(1)}(\theta)$ and $h_x^{(2)}(\theta)$, become:

$$\begin{aligned} h_x^{(1)}(\theta) &= vec\left(\Psi_{t,t+1}Q_x^{-1}\right)^{\mathrm{T}} f_x(\mu_\theta) + vec\left(\Psi_{t,t+1}Q_x^{-1}\right)^{\mathrm{T}}\frac{\partial f_x}{\partial \theta}^{\mathrm{T}}(\theta-\mu_\theta) \\ h_x^{(2)}(\theta) &= f_x(\mu_\theta)^{\mathrm{T}}\left[Q_x^{-1}\otimes\Psi_{t,t}\right] f_x(\mu_\theta) + 2 f_x(\mu_\theta)^{\mathrm{T}}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial f_x}{\partial \theta}^{\mathrm{T}}(\theta-\mu_\theta) + (\theta-\mu_\theta)\frac{\partial f_x}{\partial \theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial f_x}{\partial \theta}^{\mathrm{T}}(\theta-\mu_\theta) \end{aligned}$$

This yields the expressions of their gradients and Hessians, evaluated at μ_θ :

$$\left.\frac{\partial h_x^{(1)}}{\partial\theta}\right|_{\theta=\mu_\theta} = \frac{\partial f_x}{\partial\theta} vec\left(\Psi_{t,t+1} Q_x^{-1}\right) \quad \forall t\neq 0 \quad (=0 \text{ otherwise})$$

$$\left.\frac{\partial^2 h_x^{(1)}}{\partial\theta^2}\right|_{\theta=\mu_\theta} = 0$$

$$\left.\frac{\partial h_x^{(2)}}{\partial\theta}\right|_{\theta=\mu_\theta} = 2\frac{\partial f_x}{\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right] f_x\left(\mu_\theta\right)$$

$$\left.\frac{\partial^2 h_x^{(2)}}{\partial\theta^2}\right|_{\theta=\mu_\theta} = 2\frac{\partial f_x}{\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial f_x}{\partial\theta}^{\mathrm{T}}$$

Which finally yields:

$$\mu_\theta^{(k+1)} - \mu_\theta^{(k)} = \Sigma_\theta^{(k+1)}\left[Q_{\theta 0}^{-1}\left(\mu_{\theta 0} - \mu_\theta^{(k)}\right) + \alpha\sum_{t=0}^{T-1}\left(\frac{\partial f}{\partial\theta} Q_x^{-1}\left(\mu_{t+1} - f\left(\mu_t,\mu_\theta\right)^{\mathrm{T}}\right) + \frac{\partial^2 \tilde{f}}{\partial x\partial\theta}\left(vec\left(\Psi_{t,t+1}Q_x^{-1}\right) - \left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial\tilde{f}}{\partial x}\right)\right)\right]$$

$$\Sigma_\theta^{(k+1)} = \left[\alpha\sum_{t=1}^{T}\left(\frac{\partial f}{\partial\theta}Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} + \frac{\partial^2\tilde{f}}{\partial x\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{f}}{\partial x\partial\theta}^{\mathrm{T}}\right) + Q_{\theta 0}^{-1}\right]^{-1}$$

3. Hidden states

We use a variational generalization of the standard Kalman-Rauch-Striebel smoother, which is derived using the four following steps: prediction, update, β-message passing scheme, and full posterior derivation.

3.1 Forward pass: the ***predictive density***

$$\alpha_t^*(x_t) \propto \int \underbrace{\alpha_{t-1}(x_{t-1}) \exp\left\langle \log p\left(x_t \middle| x_{t-1},...\right)\right\rangle}_{\exp h(x_t, x_{t-1})} dx_{t-1} ,$$

where $\alpha_{t-1}(x_{t-1})$ is the previous updated hidden states density.

Laplace approximation on the updated density (see section 3.2) yields :

$$h(x_t, x_{t-1}) = -\frac{1}{2}(x_{t-1} - m_{t-1})^{\mathrm{T}} R_{t-1|t-1}{}^{-1}(x_{t-1} - m_{t-1}) - \frac{\alpha}{2}\left(x_t{}^{\mathrm{T}} Q_x{}^{-1} x_t - 2\, x_t{}^{\mathrm{T}} Q_x{}^{-1}\left\langle f\left(x_{t-1},\theta\right)\right\rangle + \left\langle f\left(x_{t-1},\theta\right)^{\mathrm{T}} Q_x{}^{-1} f\left(x_{t-1},\theta\right)\right\rangle\right)$$

Taylor-expand $f(x_{t-1},\theta)$ around μ_θ :

$$f\left(x_{t-1},\theta\right) = f\left(x_{t-1},\mu_\theta\right) + \frac{\partial f}{\partial \theta}^{\mathrm{T}}\left(\theta - \mu_\theta\right) + O^{(2)}$$

Then:

$$\left\langle f\left(x_{t-1},\theta\right)\right\rangle = f\left(x_{t-1},\mu_\theta\right)$$

$$\left\langle f\left(x_{t-1},\theta\right)^{\mathrm{T}} Q_x{}^{-1} f\left(x_{t-1},\theta\right)\right\rangle = f\left(x_{t-1},\mu_\theta\right)^{\mathrm{T}} Q_x{}^{-1} f\left(x_{t-1},\mu_\theta\right) + \underbrace{tr\left[\frac{\partial f}{\partial \theta} Q_x{}^{-1} \frac{\partial f}{\partial \theta}^{\mathrm{T}} \Sigma_\theta\right]}_{j_\theta(x)}$$

Taylor-expand $f\left(x_{t-1},\theta\right)$ and $j_\theta\left(x\right)$ around μ_{t-1}:

$$f\left(x_{t-1},\mu_\theta\right) = f\left(\mu_{t-1},\mu_\theta\right) + \frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + O^{(2)}$$

$$j_\theta\left(x\right) = j_\theta\left(\mu_{t-1}\right) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + \frac{1}{2}\left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\left(x_{t-1}-\mu_{t-1}\right) + O^{(3)}$$

Then the required first- and second-order expectations write:

$$\left\langle f\left(x_{t-1},\theta\right)\right\rangle = f\left(\mu_{t-1},\mu_\theta\right) + \frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + O^{(2)}$$

$$\begin{aligned}\left\langle f\left(x_{t-1},\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(x_{t-1},\theta\right)\right\rangle = {} & f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(\mu_{t-1},\mu_\theta\right) \\ & + 2\, f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + \left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) \\ & + j_\theta\left(\mu_{t-1}\right) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + \frac{1}{2}\left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\left(x_{t-1}-\mu_{t-1}\right)\end{aligned}$$

Then $h\left(x_t, x_{t-1}\right)$ writes:

$$
\begin{aligned}
h\left(x_t, x_{t-1}\right) = & -\frac{1}{2}\left(x_{t-1}{}^{\mathrm{T}} R_{t-1|t-1}{}^{-1} x_{t-1} - 2 m_{t-1}{}^{\mathrm{T}} R_{t-1|t-1}{}^{-1} x_{t-1} + m_{t-1}{}^{\mathrm{T}} R_{t-1|t-1}{}^{-1} m_{t-1}\right) \\
& -\frac{\alpha}{2}\Bigg(x_t{}^{\mathrm{T}} Q_x{}^{-1} x_t - 2\, x_t{}^{\mathrm{T}} Q_x{}^{-1} f\left(\mu_{t-1}, \mu_\theta\right) - 2\, x_t{}^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_{t-1} - \mu_{t-1}\right) \\
& \qquad + f\left(\mu_{t-1}, \mu_\theta\right)^{\mathrm{T}} Q_x{}^{-1} f\left(\mu_{t-1}, \mu_\theta\right) \\
& \qquad + 2\, f\left(\mu_{t-1}, \mu_\theta\right)^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_{t-1} - \mu_{t-1}\right) + \left(x_{t-1} - \mu_{t-1}\right)^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_{t-1} - \mu_{t-1}\right) \\
& \qquad + j_\theta\left(\mu_{t-1}\right) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} \left(x_{t-1} - \mu_{t-1}\right) + \frac{1}{2}\left(x_{t-1} - \mu_{t-1}\right)^{\mathrm{T}} \frac{\partial^2 j_\theta}{\partial x^2}\left(x_{t-1} - \mu_{t-1}\right)\Bigg) \\
= & -\frac{1}{2} x_{t-1}{}^{\mathrm{T}} \left[R_{t-1|t-1}{}^{-1} + \alpha\left(\frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right)\right] x_{t-1} \\
& -\frac{1}{2}\Bigg[-2 m_{t-1}{}^{\mathrm{T}} R_{t-1|t-1}{}^{-1} \\
& \quad + \alpha\left(-2\, x_t{}^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + 2\, f\left(\mu_{t-1}, \mu_\theta\right)^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} - 2 \mu_{t-1}{}^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} - \mu_{t-1}{}^{\mathrm{T}} \frac{\partial^2 j_\theta}{\partial x^2}\right)\Bigg] x_{t-1} \\
& -\frac{\alpha}{2}\left[x_t{}^{\mathrm{T}} Q_x{}^{-1} x_t - 2\, x_t{}^{\mathrm{T}} Q_x{}^{-1} f\left(\mu_{t-1}, \mu_\theta\right) + 2\, x_t{}^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}\right] \\
& + K^0 \\
= & -\frac{1}{2} x_{t-1}{}^{\mathrm{T}} B\, x_{t-1} - A^{\mathrm{T}} x_{t-1} + K\left(x_t\right) + K^0
\end{aligned}
$$

Using the QF transform:

$$
h\left(x_t, x_{t-1}\right) = -\frac{1}{2}\left(x_{t-1} - \nu\right)^{\mathrm{T}} B\left(x_{t-1} - \nu\right) + \frac{1}{2} A^{\mathrm{T}} B^{-1} A + K\left(x_t\right) + K^0
$$

where:

$$
\begin{aligned}
\nu &= -B^{-1} A \\
B &= R_{t-1|t-1}{}^{-1} + \alpha\left(\frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right) \\
A &= -R_{t-1|t-1}{}^{-1} m_{t-1} + \alpha\left(-\frac{\partial f}{\partial x} Q_x{}^{-1} x_t + \frac{\partial f}{\partial x} Q_x{}^{-1}\left(f\left(\mu_{t-1}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}\right) + \frac{1}{2}\left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}\right)\right)
\end{aligned}
$$

Then the predictive density $\alpha_t^*(x_t)$ becomes:

$$\begin{aligned}\alpha_t^*(x_t) &\propto \int \exp\left[-\frac{1}{2}(x_{t-1}-\nu)^{\mathrm{T}} B(x_{t-1}-\nu)+\frac{1}{2}A^{\mathrm{T}}B^{-1}A+K(x_t)+K\right]dx_{t-1} \\ &\propto \exp\left[\frac{1}{2}A^{\mathrm{T}}B^{-1}A+K(x_t)+K\right]\end{aligned},$$

which is a quadratic form in x_t:

$$\begin{aligned}\log\alpha_t^*(x_t) &= -\frac{\alpha}{2}\left[x_t^{\mathrm{T}}Q_x^{-1}x_t-2x_t^{\mathrm{T}}Q_x^{-1}f(\mu_{t-1},\mu_\theta)+2x_t^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right] \\ &\quad+\frac{1}{2}\alpha^2x_t^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\frac{\partial f}{\partial x}Q_x^{-1}x_t \\ &\quad-\alpha x_t^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\left[-R_{t-1|t-1}^{-1}m_{t-1}+\alpha\left(\frac{\partial f}{\partial x}Q_x^{-1}\left(f(\mu_{t-1},\mu_\theta)-\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right)+\frac{1}{2}\left(\frac{\partial j_\theta}{\partial x}-\frac{\partial^2 j_\theta}{\partial x^2}\mu_{t-1}\right)\right)\right] \\ &= -\frac{\alpha}{2}x_t^{\mathrm{T}}\left[I-\alpha Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\frac{\partial f}{\partial x}\right]Q_x^{-1}x_t \\ &\quad-\alpha x_t^{\mathrm{T}}Q_x^{-1}\left[-f(\mu_{t-1},\mu_\theta)+\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right. \\ &\qquad\left.+\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\left[-R_{t-1|t-1}^{-1}m_{t-1}+\alpha\left(\frac{\partial f}{\partial x}Q_x^{-1}\left(f(\mu_{t-1},\mu_\theta)-\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right)+\frac{1}{2}\left(\frac{\partial j_\theta}{\partial x}-\frac{\partial^2 j_\theta}{\partial x^2}\mu_{t-1}\right)\right)\right]\right] \\ &= -\frac{1}{2}x_t^{\mathrm{T}}Cx_t+x_t^{\mathrm{T}}C\left(f(\mu_{t-1},\mu_\theta)-\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right) \\ &\quad-\alpha x_t^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\left[-R_{t-1|t-1}^{-1}m_{t-1}+\frac{\alpha}{2}\left(\frac{\partial j_\theta}{\partial x}-\frac{\partial^2 j_\theta}{\partial x^2}\mu_{t-1}\right)\right] \\ &= -\frac{1}{2}x_t^{\mathrm{T}}Cx_t-x_t^{\mathrm{T}}D\end{aligned}$$

Then, using the QF allows us to derive the update rule for the predictive density:

$$m_t^* = -C^{-1}\left[-C\left(f(\mu_{t-1},\mu_\theta)-\frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}\right)+\alpha Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}B^{-1}\left[\frac{\alpha}{2}\left(\frac{\partial j_\theta}{\partial x}-\frac{\partial^2 j_\theta}{\partial x^2}\mu_{t-1}\right)-R_{t-1|t-1}^{-1}m_{t-1}\right]\right],$$

From which we write:

$$
\begin{aligned}
m_t^* &= f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + \alpha R_{t|t-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} \left[R_{t-1|t-1}{}^{-1} m_{t-1} - \frac{\alpha}{2} \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
R_{t|t-1} &= \frac{1}{\alpha} Q_x \left[I - \alpha Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} \frac{\partial f}{\partial x} \right]^{-1} \\
B &= R_{t-1|t-1}{}^{-1} + \alpha \left(\frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\theta}{\partial x^2} \right)
\end{aligned}
$$

Then, we make use of some algebra to get a simpler expression for (5):

$$
\begin{aligned}
m_t^* &= f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + Q_x \left[I - \alpha Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} \frac{\partial f}{\partial x} \right]^{-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} \left[R_{t-1|t-1}{}^{-1} m_{t-1} - \frac{1}{2} \alpha \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
&= f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + \left(-\frac{T_e}{\alpha} \right) Q_x \left[I + Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(-\alpha B^{-1} \right) \frac{\partial f}{\partial x} \right]^{-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(-\alpha B^{-1} \right) \left[R_{t-1|t-1}{}^{-1} m_{t-1} \right. \\
&\qquad \left. - \frac{\alpha}{2} \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
&== f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + \left(-\frac{1}{\alpha} \right) \left[Q_x^{-1} + Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(-\alpha B^{-1} \right) \frac{\partial f}{\partial x} Q_x^{-1} \right]^{-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(-\alpha B^{-1} \right) \left[R_{t-1|t-1}{}^{-1} m_{t-1} \right. \\
&\qquad \left. - \frac{\alpha}{2} \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
&= f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + \left(-\frac{1}{\alpha} \right) \frac{\partial f}{\partial x}^{\mathrm{T}} \left[\frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} - \frac{1}{\alpha} B \right]^{-1} \left[R_{t-1|t-1}{}^{-1} m_{t-1} - \frac{\alpha}{2} \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
&= f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1}^{(k)} + \frac{\partial f}{\partial x}^{\mathrm{T}} \left[\frac{1}{\alpha} R_{t-1|t-1}{}^{-1} + \frac{1}{2} \frac{\partial^2 j_\theta}{\partial x^2} \right]^{-1} \left[\frac{1}{\alpha} R_{t-1|t-1}{}^{-1} m_{t-1} - \frac{1}{2} \left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2} \mu_{t-1}^{(k)} \right) \right] \\
&= \underbrace{f\left(\mu_{t-1}^{(k)}, \mu_\theta\right) + \frac{\partial f}{\partial x}^{\mathrm{T}} \left(m_{t-1} - \mu_{t-1}^{(k)} \right)}_{\text{standard Gauss-Newton EKF prediction}} + \underbrace{\frac{1}{2} \underbrace{\frac{\partial f}{\partial x}^{\mathrm{T}} \left[\frac{1_e}{\alpha} R_{t-1|t-1}{}^{-1} + \frac{1}{2} \frac{\partial^2 j_\theta}{\partial x^2} \right]^{-1}}_{\frac{1}{\alpha} Q_x \left[I - \alpha Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} \frac{\partial f}{\partial x} \right]^{-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1} = \alpha^2 R_{t|t-1} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} B^{-1}} \left[\frac{\partial^2 j_\theta}{\partial x^2} \left(\mu_{t-1}^{(k)} - m_{t-1} \right) - \frac{\partial j_\theta}{\partial x} \right]}_{\text{mean-field perturbation term}}
\end{aligned}
$$

NB: the 2nd to 3rd line passage makes use of the Woodbury matrix inversion lemna (given in the appendix). The next passage eventually uses the expression of the B matrix (given in (5)). The last line is obtained by adding and removing the mean-field trace term to the previous updated hidden states mode in the right-hand term.

The first term is similar to that of the standard the Gauss-Newton version of the extended Kalman filter (see [Johnston, 2001]). The second one is related to the mean-field approximation. It is a perturbation term that accounts for the uncertainty related to the posterior belief upon the evolution parameters θ.

Here again, we use the bilinear approximation of the evolution function (see sections 1 and 2) to provide a numerical estimation of the mean-field trace term $j_\theta(x)$:

$$j_\theta(x) = tr\left[\frac{\partial f}{\partial\theta} Q_x^{-1} \frac{\partial f}{\partial\theta}^{\mathrm{T}} \Sigma_\theta\right] = vec\left(\frac{\partial f}{\partial\theta}\right)^{\mathrm{T}} \left[Q_x^{-1} \otimes \Sigma_\theta\right] vec\left(\frac{\partial f}{\partial\theta}\right) \equiv f_\theta(x)^{\mathrm{T}} \left[Q_x^{-1} \otimes \Sigma_\theta\right] f_\theta(x)$$

where $f_\theta(x)$ is the vectorized gradient of the evolution function $f(x_t,\theta)$ w.r.t to the evolution parameters θ, evaluated at μ_θ but still depending on x:

$$f_\theta(x) = vec\begin{pmatrix} \frac{\partial f_1}{\partial\theta_1} & \cdots & \frac{\partial f_n}{\partial\theta_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial\theta_{n_\theta}} & \cdots & \frac{\partial f_n}{\partial\theta_{n_\theta}} \end{pmatrix} = \left[\frac{\partial f_1}{\partial\theta_1} \cdots \frac{\partial f_1}{\partial\theta_{n_\theta}} \quad \cdots \quad \frac{\partial f_n}{\partial\theta_1} \cdots \frac{\partial f_n}{\partial\theta_{n_\theta}}\right]^{\mathrm{T}} \equiv \frac{\partial \tilde{f}}{\partial\theta}.$$

We then Taylor-expand $f_\theta(x)$ around μ_{t-1}:

$$f_\theta(x) = f_\theta(\mu_{t-1}) + \frac{\partial f_\theta}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) + O^{(2)}.$$

Then, the trace term $j_\theta(x)$ becomes:

$$\begin{aligned} j_\theta(x) = {} & f_\theta(\mu_{t-1})^{\mathrm{T}} \left[Q_x^{-1} \otimes \Sigma_\theta\right] f_\theta(\mu_{t-1}) + 2 f_\theta(\mu_{t-1})^{\mathrm{T}} \left[Q_x^{-1} \otimes \Sigma_\theta\right] \frac{\partial f_\theta}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) \\ & + (x_{t-1} - \mu_{t-1})^{\mathrm{T}} \frac{\partial f_\theta}{\partial x} \left[Q_x^{-1} \otimes \Sigma_\theta\right] \frac{\partial f_\theta}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) \end{aligned}$$

This yields the expressions of its gradient and Hessian, evaluated at μ_θ:

$$\left.\frac{\partial j_\theta}{\partial x}\right|_{x=\mu_{t-1}} = 2\frac{\partial f_\theta}{\partial x}\left[Q_x^{-1}\otimes\Sigma_\theta\right] f_\theta\left(\mu_{t-1}\right)$$
$$\left.\frac{\partial^2 j_\theta}{\partial x^2}\right|_{x=\mu_{t-1}} = 2\frac{\partial f_\theta}{\partial x}\left[Q_x^{-1}\otimes\Sigma_\theta\right]\frac{\partial f_\theta}{\partial x}^{\mathrm{T}} ,$$

where the mixed derivative $\frac{\partial f_\theta}{\partial x}$ is calculated as follows:

$$\frac{\partial f_\theta}{\partial x}=\begin{bmatrix}\frac{\partial f_{\theta 1}}{\partial x_1} & \cdots & \frac{\partial f_{\theta nn_\theta}}{\partial x_1}\\ \vdots & & \vdots\\ \frac{\partial f_{\theta 1}}{\partial x_n} & \cdots & \frac{\partial f_{\theta nn_\theta}}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial f_1}{\partial\theta_1\partial x_1} & \cdots & \frac{\partial f_1}{\partial\theta_{n_\theta}\partial x_1} & \cdots & \frac{\partial f_n}{\partial\theta_1\partial x_1} & \cdots & \frac{\partial f_n}{\partial\theta_{n_\theta}\partial x_1}\\ \vdots & & \vdots & & \vdots & & \vdots\\ \frac{\partial f_1}{\partial\theta_1\partial x_n} & \cdots & \frac{\partial f_1}{\partial\theta_{n_\theta}\partial x_n} & \cdots & \frac{\partial f_n}{\partial\theta_1\partial x_n} & \cdots & \frac{\partial f_1}{\partial\theta_{n_\theta}\partial x_n}\end{bmatrix}\equiv\frac{\partial^2\tilde{f}}{\partial x\partial\theta}.$$

3.2 Forward pass: the *updated density*

$$\alpha_t\left(x_t\right)\propto\alpha_t^*\left(x_t\right)\exp\left\langle\log p\left(y_t\middle|x_t,...\right)\right\rangle,$$

where $\alpha_t^*(x_t)$ is the current predictive hidden states density.

Then:

$$\log\alpha_t\left(x_t\right)=-\frac{1}{2}\left(x_t-m_t^*\right)^{\mathrm{T}}R_{t|t-1}{}^{-1}\left(x_t-m_t^*\right)-\frac{1}{2}\sigma\left(y_t{}^{\mathrm{T}}Q_y{}^{-1}y_t-2\,y_t{}^{\mathrm{T}}Q_y{}^{-1}\left\langle g\left(x_t,\varphi\right)\right\rangle+\left\langle g\left(x_t,\varphi\right)^{\mathrm{T}}Q_y{}^{-1}g\left(x_t,\varphi\right)\right\rangle\right)$$

Taylor-expand $g(x_t,\varphi)$ around μ_φ:

$$g\left(x_t,\varphi\right)=g\left(x_t,\mu_\varphi\right)+\frac{\partial g}{\partial\varphi}^{\mathrm{T}}\left(\varphi-\mu_\varphi\right)+O^{(2)}$$

Then:

$$\left\langle g\left(x_t,\varphi\right)\right\rangle = g\left(x_t,\mu_\varphi\right)$$
$$\left\langle g\left(x_t,\varphi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\varphi\right)\right\rangle = g\left(x_t,\mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\mu_\varphi\right) + \underbrace{tr\left[\frac{\partial g}{\partial\varphi} Q_y^{-1} \frac{\partial g}{\partial\varphi}^{\mathrm{T}} \Sigma_\varphi\right]}_{j_\varphi(x)}$$

Taylor-expand $g\left(x_t,\mu_\varphi\right)$ and $j_\varphi\left(x\right)$ around μ_t:

$$g\left(x_t,\mu_\varphi\right) = g\left(\mu_t,\mu_\varphi\right) + \frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + O^{(2)}$$
$$j_\varphi\left(x\right) = j_\varphi\left(\mu_t\right) + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + \frac{1}{2}\left(x_t-\mu_t\right)^{\mathrm{T}} \frac{\partial^2 j_\varphi}{\partial x^2}\left(x_t-\mu_t\right)$$

Then the required first- and second-order expectations write:

$$\left\langle g\left(x_t,\varphi\right)\right\rangle = g\left(\mu_t,\mu_\varphi\right) + \frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right)$$
$$\begin{aligned}\left\langle g\left(x_t,\varphi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_t,\varphi\right)\right\rangle &= g\left(\mu_t,\mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\varphi\right)\\ &+ 2\,g\left(\mu_t,\mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + \left(x_t-\mu_t\right)^{\mathrm{T}} \frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right)\\ &+ j_\varphi\left(\mu_t\right) + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + \frac{1}{2}\left(x_t-\mu_t\right)^{\mathrm{T}} \frac{\partial^2 j_\varphi}{\partial x^2}\left(x_t-\mu_t\right)\end{aligned}$$

Then, we rewrite the log- updated density:

$$
\begin{aligned}
\log \alpha_t\left(x_t\right) = & -\frac{1}{2}\left[x_t^{\mathrm{T}} R_{t|t-1}{}^{-1} x_t - 2 m_t^{*\mathrm{T}} R_{t|t-1}{}^{-1} x_t + m_t^{*\mathrm{T}} R_{t|t-1}{}^{-1} m_t^*\right] \\
& -\frac{\sigma}{2}\left[y_t^{\mathrm{T}} Q_y^{-1} y_t - 2 y_t^{\mathrm{T}} Q_y^{-1} g\left(\mu_t, \mu_\varphi\right) - 2 y_t^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} x_t + 2 y_t^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \mu_t \right.\\
& \qquad + g\left(\mu_t, \mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t, \mu_\varphi\right) \\
& \qquad + 2 g\left(\mu_t, \mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) + \left(x_t - \mu_t\right)^{\mathrm{T}} \frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) \\
& \qquad \left. + j_\varphi\left(\mu_t\right) + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) + \frac{1}{2}\left(x_t - \mu_t\right)^{\mathrm{T}} \frac{\partial^2 j_\varphi}{\partial x^2}\left(x_t - \mu_t\right) \right] \\
= & -\frac{1}{2} x_t^{\mathrm{T}} \left[R_{t|t-1}{}^{-1} + \left(\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\varphi}{\partial x^2} \right) \right] x_t \\
& -\frac{1}{2}\left[-2 m_t^{*\mathrm{T}} R_{t|t-1}{}^{-1} + \left(-2 y_t^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + 2 g\left(\mu_t, \mu_\varphi\right)^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x} - 2 \mu_t^{\mathrm{T}} \frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}} - \mu_t^{\mathrm{T}} \frac{\partial^2 j_\varphi}{\partial x^2} \right) \right] x_t \\
& + K^0 \\
= & -\frac{1}{2} x_t^{\mathrm{T}} D x_t - A^{\mathrm{T}} x_t + K^0
\end{aligned}
$$

Then, the use of QF yields the update equations for the updated density:

$$
\begin{aligned}
m_t &= R_{t|t}\left[R_{t|t-1}{}^{-1} m_t^* + \sigma \left[\frac{\partial g}{\partial x} Q_y^{-1} \left(y_t - g\left(\mu_t^{(k)}, \mu_\varphi\right) \right) + \left(\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\varphi}{\partial x^2} \right) \mu_t^{(k)} - \frac{1}{2} \frac{\partial j_\varphi}{\partial x} \right] \right] \\
R_{t|t} &= \left[R_{t|t-1}{}^{-1} + \sigma \left(\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\varphi}{\partial x^2} \right) \right]^{-1}
\end{aligned}
$$

Again, we use some algebra to simplify expressions (7):

$$\begin{aligned}
m_t &= R_{t|t}\left[R_{t|t-1}{}^{-1}m_t^* + \sigma\left[\frac{\partial g}{\partial x}Q_y{}^{-1}\left(y_t - g\left(\mu_t^{(k)},\mu_\varphi\right)\right) + \left(\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\varphi}{\partial x^2}\right)\mu_t^{(k)} - \frac{1}{2}\frac{\partial j_\varphi}{\partial x}\right]\right] \\
&= m_t^* + R_{t|t}\left[-\sigma\left(\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\varphi}{\partial x^2}\right)m_t^* + \sigma\left[\frac{\partial g}{\partial x}Q_y{}^{-1}\left(y_t - g\left(\mu_t^{(k)},\mu_\varphi\right)\right) + \left(\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\varphi}{\partial x^2}\right)\mu_t^{(k)} - \frac{1}{2}\frac{\partial j_\varphi}{\partial x}\right]\right] \\
&= m_t^* + \sigma R_{t|t}\left[\frac{\partial g}{\partial x}Q_y{}^{-1}\left(y_t - g\left(\mu_t^{(k)},\mu_\varphi\right)\right) + \left(\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\varphi}{\partial x^2}\right)\left(\mu_t^{(k)} - m_t^*\right) - \frac{1}{2}\frac{\partial j_\varphi}{\partial x}\right] \\
&= \underbrace{m_t^* + \sigma R_{t|t}\frac{\partial g}{\partial x}Q_y{}^{-1}\left[y_t - g\left(\mu_t^{(k)},\mu_\varphi\right) + \frac{\partial g}{\partial x}^{\mathrm{T}}\left(\mu_t^{(k)} - m_t^*\right)\right]}_{\text{standard Gauss-Newton EKF update}} + \underbrace{\frac{1}{2}\sigma R_{t|t}\left[\frac{\partial^2 j_\varphi}{\partial x^2}\left(\mu_t^{(k)} - m_t^*\right) - \frac{\partial j_\varphi}{\partial x}\right]}_{\text{mean-field perturbation term}}
\end{aligned}$$

As usual, we use the bilinear approximation of the evolution function to provide a numerical estimation of the mean-field trace term $j_\varphi(x)$:

$$j_\varphi(x) = tr\left[\frac{\partial g}{\partial \varphi}Q_y{}^{-1}\frac{\partial g}{\partial \varphi}^{\mathrm{T}}\Sigma_\varphi\right] = vec\left(\frac{\partial g}{\partial \varphi}\right)^{\mathrm{T}}\left[Q_y{}^{-1}\otimes\Sigma_\varphi\right]vec\left(\frac{\partial g}{\partial \varphi}\right) \equiv g_\varphi(x)^{\mathrm{T}}\left[Q_y{}^{-1}\otimes\Sigma_\varphi\right]g_\varphi(x),$$

where $g_\varphi(x)$ is the vectorized gradient of the observation function $g(x_t,\varphi)$ w.r.t to the observation parameters φ, evaluated at μ_φ but still depending on x.

We then Taylor-expand $g_\varphi(x)$ around μ_t:

$$g_\varphi(x) = g_\varphi(\mu_t) + \frac{\partial g_\varphi}{\partial x}^{\mathrm{T}}(x_t - \mu_t) + O^{(2)}.$$

Then, the trace term $j_\phi(x)$ becomes:

$$\begin{aligned}
j_\varphi(x) = g_\varphi(\mu_t)^{\mathrm{T}}\left[Q_y{}^{-1}\otimes\Sigma_\varphi\right]g_\varphi(\mu_t) + 2g_\varphi(\mu_t)^{\mathrm{T}}\left[Q_y{}^{-1}\otimes\Sigma_\varphi\right]\frac{\partial g_\varphi}{\partial x}^{\mathrm{T}}(x_t - \mu_t) \\
+ (x_t - \mu_t)^{\mathrm{T}}\frac{\partial g_\varphi}{\partial x}\left[Q_y{}^{-1}\otimes\Sigma_\varphi\right]\frac{\partial g_\varphi}{\partial x}^{\mathrm{T}}(x_t - \mu_t)
\end{aligned}$$

This yields the expressions of its gradient and Hessian, evaluated at μ_φ:

$$\left.\frac{\partial j_\varphi}{\partial x}\right|_{x=\mu_t} = 2\frac{\partial g_\varphi}{\partial x}\left[Q_y^{-1}\otimes\Sigma_\varphi\right]g_\varphi\left(\mu_t\right) = 2\frac{\partial^2 g}{\partial x\partial\varphi}\left[Q_y^{-1}\otimes\Sigma_\varphi\right]\frac{\partial g}{\partial\varphi}$$
$$\left.\frac{\partial^2 j_\varphi}{\partial x^2}\right|_{x=\mu_t} = 2\frac{\partial g_\varphi}{\partial x}\left[Q_y^{-1}\otimes\Sigma_\varphi\right]\frac{\partial g_\varphi}{\partial x}^{\mathrm{T}} = 2\frac{\partial^2 g}{\partial x\partial\varphi}\left[Q_y^{-1}\otimes\Sigma_\varphi\right]\frac{\partial^2 g}{\partial x\partial\varphi}^{\mathrm{T}}$$

3.3 Backward pass: the *β-message* passing scheme

The standard sequential implementation, using the so-called *γ-message*, is not doable in our variational framework. Therefore, we make use of the parallel recursion, which produces the β-messages, defined as:

$$\beta_t\left(x_t\right) \equiv p\left(y_{t+1:T}\left|x_t\right.\right).$$

These are obtained through a recursion analog to the forward pass:

$$\beta_{t-1}\left(x_{t-1}\right) \propto \int \underbrace{\beta_t\left(x_t\right)\exp\left\langle \log p\left(y_t\left|x_t,\ldots\right.\right) + \log p\left(x_t\left|x_{t-1},\ldots\right.\right)\right\rangle}_{\exp h\left(x_t, x_{t-1}\right)} dx_t\ ,$$

with the terminal condition $\beta_T\left(x_T\right)=1$.

We still will make use of a quadratic approximation to the log β-messages $\beta_t\left(x_t\right)$, of the (implicit) form:

$$\log\beta_t\left(x_t\right) = -\frac{1}{2}\left(x_t - n_t\right)^{\mathrm{T}}\Omega_t^{-1}\left(x_t - n_t\right).$$

Then the integrand writes:

$$
\begin{aligned}
h(x_t, x_{t-1}) = & -\frac{1}{2}(x_t - n_t)^{\mathrm{T}} \Omega_t^{-1} (x_t - n_t) \\
& -\frac{\alpha}{2}\left[x_t^{\mathrm{T}} Q_x^{-1} x_t - 2 x_t^{\mathrm{T}} Q_x^{-1} \left\langle f(x_{t-1}, \theta) \right\rangle + \left\langle f(x_{t-1}, \theta)^{\mathrm{T}} Q_x^{-1} f(x_{t-1}, \theta) \right\rangle \right] \\
& -\frac{\sigma}{2}\left[y_t^{\mathrm{T}} Q_y^{-1} y_t - 2 y_t^{\mathrm{T}} Q_y^{-1} \left\langle g(x_t, \varphi) \right\rangle + \left\langle g(x_t, \varphi)^{\mathrm{T}} Q_y^{-1} g(x_t, \varphi) \right\rangle \right]
\end{aligned}
$$

We then Taylor-expand the observation and evolution functions as in 3.1 and 3.2, i.e:

$$
\begin{aligned}
\left\langle g(x_t, \varphi) \right\rangle &= g(\mu_t, \mu_\varphi) + \frac{\partial g}{\partial x}^{\mathrm{T}} (x_t - \mu_t) \\
\left\langle g(x_t, \varphi)^{\mathrm{T}} Q_y^{-1} g(x_t, \varphi) \right\rangle &= g(\mu_t, \mu_\varphi)^{\mathrm{T}} Q_y^{-1} g(\mu_t, \mu_\varphi) \\
& \quad + 2\, g(\mu_t, \mu_\varphi)^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} (x_t - \mu_t) + (x_t - \mu_t)^{\mathrm{T}} \frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} (x_t - \mu_t) \\
& \quad + j_\varphi(\mu_t) + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}} (x_t - \mu_t) + \frac{1}{2}(x_t - \mu_t)^{\mathrm{T}} \frac{\partial^2 j_\varphi}{\partial x^2} (x_t - \mu_t) \\
\left\langle f(x_{t-1}, \theta) \right\rangle &= f(\mu_{t-1}, \mu_\theta) + \frac{\partial f}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) + O^{(2)} \\
\left\langle f(x_{t-1}, \theta)^{\mathrm{T}} Q_x^{-1} f(x_{t-1}, \theta) \right\rangle &= f(\mu_{t-1}, \mu_\theta)^{\mathrm{T}} Q_x^{-1} f(\mu_{t-1}, \mu_\theta) \\
& \quad + 2\, f(\mu_{t-1}, \mu_\theta)^{\mathrm{T}} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) + (x_{t-1} - \mu_{t-1})^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) \\
& \quad + j_\theta(\mu_{t-1}) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} (x_{t-1} - \mu_{t-1}) + \frac{1}{2}(x_{t-1} - \mu_{t-1})^{\mathrm{T}} \frac{\partial^2 j_\theta}{\partial x^2} (x_{t-1} - \mu_{t-1})
\end{aligned}
$$

NB: the Taylor expansion of the observation function is done around $\mu_t^{(k)}$, as opposed to the evolution function which is expanded around $\mu_{t-1}^{(k)}$.

Then the integrand becomes:

$$
\begin{aligned}
h\left(x_t, x_{t-1}\right) = & -\frac{1}{2}\left[x_t{}^{\mathrm{T}}\Omega_t{}^{-1}x_t - 2n_t{}^{\mathrm{T}}\Omega_t{}^{-1}x_t + n_t{}^{\mathrm{T}}\Omega_t{}^{-1}n_t\right] \\
& -\frac{\alpha}{2}\Bigg[x_t{}^{\mathrm{T}}Q_x{}^{-1}x_t - 2x_t{}^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_{t-1},\mu_\theta\right) - 2x_t{}^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) \\
& \qquad + f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}}Q_x{}^{-1}f\left(\mu_{t-1},\mu_\theta\right) \\
& \qquad + 2\,f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + \left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial f}{\partial x}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) \\
& \qquad + j_\theta\left(\mu_{t-1}\right) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\left(x_{t-1}-\mu_{t-1}\right) + \frac{1}{2}\left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\left(x_{t-1}-\mu_{t-1}\right)\Bigg] \\
& -\frac{\sigma}{2}\Bigg[y_t{}^{\mathrm{T}}Q_y{}^{-1}y_t - 2y_t{}^{\mathrm{T}}Q_y{}^{-1}g\left(\mu_t,\mu_\varphi\right) - 2y_t{}^{\mathrm{T}}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) \\
& \qquad + g\left(\mu_t,\mu_\varphi\right)^{\mathrm{T}}Q_y{}^{-1}g\left(\mu_t,\mu_\varphi\right) \\
& \qquad + 2\,g\left(\mu_t,\mu_\varphi\right)^{\mathrm{T}}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + \left(x_t-\mu_t\right)^{\mathrm{T}}\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) \\
& \qquad + j_\varphi\left(\mu_t\right) + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}}\left(x_t-\mu_t\right) + \frac{1}{2}\left(x_t-\mu_t\right)^{\mathrm{T}}\frac{\partial^2 j_\varphi}{\partial x^2}\left(x_t-\mu_t\right)\Bigg] \\
= & -\frac{1}{2}x_t{}^{\mathrm{T}}\left[\Omega_t{}^{-1} + \alpha Q_x{}^{-1} + \sigma\left(\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\varphi}{\partial x^2}\right)\right]x_t \\
& -\frac{1}{2}\Bigg[-2n_t{}^{\mathrm{T}}\Omega_t{}^{-1} + \alpha\left(-2\,f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}}Q_x{}^{-1} - 2\left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\frac{\partial f}{\partial x}Q_x{}^{-1}\right) \\
& \quad + \sigma\left(-2y_t{}^{\mathrm{T}}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + 2\,g\left(\mu_t,\mu_\varphi\right)^{\mathrm{T}}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} - 2\mu_t{}^{\mathrm{T}}\frac{\partial g}{\partial x}Q_y{}^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{\partial j_\varphi}{\partial x}^{\mathrm{T}} - \mu_t{}^{\mathrm{T}}\frac{\partial^2 j_\varphi}{\partial x^2}\right)\Bigg]x_t \\
& -\frac{\alpha}{2}\left[\left(2\,f\left(\mu_{t-1},\mu_\theta\right)^{\mathrm{T}}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\right)\left(x_{t-1}-\mu_{t-1}\right) + \left(x_{t-1}-\mu_{t-1}\right)^{\mathrm{T}}\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right]\left(x_{t-1}-\mu_{t-1}\right)\right] \\
& + K^0 \\
= & -\frac{1}{2}x_t{}^{\mathrm{T}}Ex_t - F^{\mathrm{T}}x_t + K\left(x_{t-1}\right) + K^0
\end{aligned}
$$

Then, the log β-messages $\beta_{t-1}\left(x_{t-1}\right)$ simply becomes:

$$
\begin{aligned}
\beta_{t-1}\left(x_{t-1}\right) &= \frac{1}{2}F^{\mathrm{T}}E^{-1}F + K\left(x_{t-1}\right) + K^{00} \\
&= \frac{1}{2}\left(F^{00} + F^0 x_{t-1}\right)^{\mathrm{T}}E^{-1}\left(F^{00} + F^0 x_{t-1}\right) + K\left(x_{t-1}\right) + K^{00} \quad , \\
&= \frac{1}{2}x_{t-1}{}^{\mathrm{T}}F^{0\mathrm{T}}E^{-1}F^0 x_{t-1} + F^{00\mathrm{T}}E^{-1}F^0 x_{t-1} + K\left(x_{t-1}\right) + K^{000}
\end{aligned}
$$

where $F^{0}x_{t-1}$ and F^{00} correspond to the respective parts of the vector $F = F^{00} + F^{0}x_{t-1}$ that explicitly depend and do not depend on x_{t-1}:

$$\begin{cases} F^{0} = -\alpha Q_x^{-1}\dfrac{\partial f}{\partial x}^{\mathrm{T}} \\ F^{00} = -\Omega_t^{-1}n_t + \alpha Q_x^{-1}\left(\dfrac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1} - f\left(\mu_{t-1},\mu_\theta\right)\right) + \sigma\left(\dfrac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right) - y_t\right) - \dfrac{\partial g}{\partial x}Q_y^{-1}\dfrac{\partial g}{\partial x}^{\mathrm{T}}\mu_t + \dfrac{1}{2}\dfrac{\partial j_\phi}{\partial x} - \dfrac{1}{2}\dfrac{\partial^2 j_\phi}{\partial x^2}\mu_t\right) \\ E = \Omega_t^{-1} + \alpha Q_x^{-1} + \sigma\left(\dfrac{\partial g}{\partial x}Q_y^{-1}\dfrac{\partial g}{\partial x}^{\mathrm{T}} + \dfrac{1}{2}\dfrac{\partial^2 j_\varphi}{\partial x^2}\right) \end{cases},$$

Then:

$$
\begin{aligned}
\beta_{t-1}\left(x_{t-1}\right) = & \frac{\alpha^2}{2} x_{t-1}{}^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x{}^{-1} E^{-1} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} x_{t-1} \\
& - \frac{\alpha}{T_e} \left[-\Omega_t{}^{-1} n_t + \alpha Q_x{}^{-1} \left(\frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1} - f\left(\mu_{t-1}, \mu_\theta\right) \right) \right. \\
& \qquad \left. + \sigma \left(\frac{\partial g}{\partial x} Q_y{}^{-1} \left(g\left(\mu_t, \mu_\phi\right) - y_t \right) - \left(\frac{\partial g}{\partial x} Q_y{}^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\phi}{\partial x^2} \right) \mu_t + \frac{1}{2} \frac{\partial j_\phi}{\partial x} \right) \right]^{\mathrm{T}} E^{-1} Q_y{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} x_{t-1} \\
& - \frac{\alpha}{2} \left[\left(2 f\left(\mu_{t-1}, \mu_\theta\right)^{\mathrm{T}} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} \right) \left(x_{t-1} - \mu_{t-1}\right) + \left(x_{t-1} - \mu_{t-1}\right)^{\mathrm{T}} \left[\frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\theta}{\partial x^2} \right] \left(x_{t-1} - \mu_{t-1}\right) \right] \\
= & -\frac{\alpha}{2} x_{t-1}{}^{\mathrm{T}} \left[\frac{\partial f}{\partial x} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial^2 j_\theta}{\partial x^2} - \alpha \frac{\partial f}{\partial x} Q_x{}^{-1} E^{-1} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \right] x_{t-1} \\
& - \alpha \left[\left[-\Omega_t{}^{-1} n_t + \alpha Q_x{}^{-1} \left(\frac{\partial f}{\partial x}^{\mathrm{T}} \mu_{t-1} - f\left(\mu_{t-1}, \mu_\theta\right) \right) \right. \right. \\
& \qquad \left. + \sigma \left(\frac{\partial g}{\partial x} Q_y{}^{-1} \left(g\left(\mu_t, \mu_\phi\right) - y_t - \frac{\partial g}{\partial x}^{\mathrm{T}} \mu_t \right) - \frac{1}{2} \frac{\partial^2 j_\phi}{\partial x^2} \mu_t + \frac{1}{2} \frac{\partial j_\phi}{\partial x} \right) \right]^{\mathrm{T}} E^{-1} Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \\
& \qquad \left. + \left(\left(f\left(\mu_{t-1}, \mu_\theta\right)^{\mathrm{T}} - \mu_{t-1}{}^{\mathrm{T}} \frac{\partial f}{\partial x} \right) Q_x{}^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2} \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} - \frac{1}{2} \mu_{t-1}{}^{\mathrm{T}} \frac{\partial^2 j_\theta}{\partial x^2} \right) \right] x_{t-1} \\
& + K^0 \\
= & -\frac{1}{2} x_{t-1}{}^{\mathrm{T}} G\, x_{t-1} - H^{\mathrm{T}} x_{t-1} + K^0
\end{aligned}
$$

Using again the QF transform yields the update rules for the β-messages $\beta_{t-1}\left(x_{t-1}\right)$:

$$n_{t-1} = \alpha\,\Omega_{t-1}\left[\frac{\partial f}{\partial x}Q_x^{-1}E^{-1}\left[\Omega_t^{-1}n_t + \alpha Q_x^{-1}\left(f\left(\mu_{t-1}^{(k)},\mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}^{(k)}\right)\right.\right.$$

$$\left.-\sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right) - y_t - \frac{\partial g}{\partial x}^{\mathrm{T}}\mu_t^{(k)}\right) + \frac{1}{2}\left(\frac{\partial j_\phi}{\partial x} - \frac{\partial^2 j_\phi}{\partial x^2}\mu_t^{(k)}\right)\right)\right]$$

$$\left. - \frac{\partial f}{\partial x}Q_x^{-1}\left(f\left(\mu_{t-1}^{(k)},\mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}}\mu_{t-1}^{(k)}\right) - \frac{1}{2}\left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2}\mu_{t-1}^{(k)}\right)\right]$$

$$\Omega_{t-1} = \frac{1}{\alpha}\left[\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2} - \alpha\frac{\partial f}{\partial x}Q_x^{-1}E^{-1}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\right]^{-1}$$

We still make use of some algebra to simplify Eq. (9):

$$
\begin{aligned}
n_{t-1} &= \alpha\,\Omega_{t-1}\Bigg[\left(\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}} - \alpha\frac{\partial f}{\partial x}Q_x^{-1}E^{-1}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right)\mu_{t-1}^{(k)} \\
&\quad + \frac{\partial f}{\partial x}Q_x^{-1}E^{-1}\left[\Omega_t^{-1}n_t + \alpha Q_x^{-1}f\left(\mu_{t-1}^{(k)},\mu_\theta\right) - \sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right) - y_t - \frac{\partial g}{\partial x}^{\mathrm{T}}\mu_t^{(k)}\right) + \frac{1}{2}\left(\frac{\partial j_\phi}{\partial x} - \frac{\partial^2 j_\phi}{\partial x^2}\mu_t^{(k)}\right)\right)\right] \\
&\quad - \frac{\partial f}{\partial x}Q_x^{-1}f\left(\mu_{t-1}^{(k)},\mu_\theta\right) - \frac{1}{2}\frac{\partial j_\theta}{\partial x}\Bigg] \\
&= \mu_{t-1}^{(k)} + \alpha\,\Omega_{t-1}\Bigg[\frac{\partial f}{\partial x}Q_x^{-1}E^{-1}\Big[\Omega_t^{-1}n_t + \alpha Q_x^{-1}f\left(\mu_{t-1}^{(k)},\mu_\theta\right) \\
&\quad - \sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right) - y_t - \frac{\partial g}{\partial x}^{\mathrm{T}}\mu_t^{(k)}\right) + \frac{1}{2}\left(\frac{\partial j_\phi}{\partial x} - \frac{\partial^2 j_\phi}{\partial x^2}\mu_t^{(k)}\right)\right)\Big] \\
&\quad - \frac{\partial f}{\partial x}Q_x^{-1}f\left(\mu_{t-1}^{(k)},\mu_\theta\right) - \frac{1}{2}\frac{\partial j_\theta}{\partial x}\Bigg]
\end{aligned}
$$

A further simplification can be afforded by noting that:

$$n_{t-1} = \mu_{t-1}^{(k)} + \alpha\Omega_{t-1}\left[\frac{\partial f}{\partial x}Q_x^{-1}E^{-1}\left[\left(\sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\phi}{\partial x^2}\right)+\Omega_t^{-1}+\alpha Q_x^{-1}\right)\mu_t^{(k)} + \Omega_t^{-1}\left(n_t-\mu_t^{(k)}\right)+\alpha Q_x^{-1}\left(f\left(\mu_{t-1}^{(k)},\mu_\theta\right)-\mu_t^{(k)}\right)\right.\right.$$

$$\left.\left. -\sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right)-y_t\right)+\frac{1}{2}\frac{\partial j_\phi}{\partial x}\right)\right] - \frac{\partial f}{\partial x}Q_x^{-1}f\left(\mu_{t-1}^{(k)},\mu_\theta\right)-\frac{1}{2}\frac{\partial j_\theta}{\partial x}\right]$$

$$= \mu_{t-1}^{(k)} + \alpha\Omega_{t-1}\left[\frac{\partial f}{\partial x}Q_x^{-1}\left(\mu_t^{(k)} - f\left(\mu_{t-1}^{(k)},\mu_\theta\right)\right) - \frac{1}{2}\frac{\partial j_\theta}{\partial x}\right.$$

$$\left. + \frac{\partial f}{\partial x}Q_x^{-1}E^{-1}\left[\Omega_t^{-1}\left(n_t-\mu_t^{(k)}\right)+\alpha Q_x^{-1}\left(f\left(\mu_{t-1}^{(k)},\mu_\theta\right)-\mu_t^{(k)}\right)-\sigma\left(\frac{\partial g}{\partial x}Q_y^{-1}\left(g\left(\mu_t,\mu_\phi\right)-y_t\right)+\frac{1}{2}\frac{\partial j_\phi}{\partial x}\right)\right]\right]$$

Again, we use the bilinear approximation of the observation and evolution functions to provide a numerical estimation of the mean-field trace terms $j_\varphi(x)$ and $j_\theta(x)$. These are derived using Taylor expansions of $g_\varphi(x)$ and $f_\theta(x)$ respectively around μ_t and μ_{t-1}, exactly as it is done in the forward pass (sections 3.1 and 3.2).

3.4 Backward pass: the *variational posterior pdf*

3.4.1 The αβ-message passing scheme

The variational posterior is simply derived using the following fusion of α- and β-messages:

$$q(x_t) \equiv p(x_t|y_{1:T}) = \alpha_t(x_t)\,\beta_t(x_t)$$

This a Gaussian pdf whose sufficient statistics are given by:

$$\begin{aligned} \mu_t &= \Psi_{t,t}\left(R_{t|t}^{-1} m_t + \Omega_t^{-1} n_t\right) \\ \Psi_{t,t} &= \left[R_{t|t}^{-1} + \Omega_t^{-1}\right]^{-1} \end{aligned}.$$

3.4.2 The inter-time step covariance matrix

This covariance matrix is required for the update equations of the evolution parameters (see section 2). It is derived from the explicit calculus of the following joint pdf:

$$\begin{aligned} p\left(x_t, x_{t+1} \middle| y_{1:T}\right) &\propto p\left(x_t \middle| y_{1:t}\right) p\left(x_{t+1} \middle| x_t\right) p\left(y_{t+1} \middle| x_{t+1}\right) p\left(y_{t+2:T} \middle| x_{t+1}\right) \\ &= \alpha_t\left(x_t\right) p\left(x_{t+1} \middle| x_t\right) p\left(y_{t+1} \middle| x_{t+1}\right) \beta_{t+1}\left(x_{t+1}\right) \\ &\xrightarrow{VB} \alpha_t\left(x_t\right) \exp\left\langle \log p\left(x_{t+1} \middle| x_t\right) + \log p\left(y_{t+1} \middle| x_{t+1}\right)\right\rangle \beta_{t+1}\left(x_{t+1}\right) \\ &\approx N\left(\begin{pmatrix} \mu_t \\ \mu_{t+1} \end{pmatrix}, \begin{bmatrix} \Psi_{t,t} & \Psi_{t,t+1} \\ \Psi_{t,t+1}^{\mathrm{T}} & \Psi_{t+1,t+1} \end{bmatrix}\right) \end{aligned}$$

the last line being the consequence of the Laplace approximation.

Then, we can write:

$$\begin{aligned} \log p\left(x_{t+1}, x_t \middle| y_{1:T}\right) = &-\frac{1}{2}\left(x_t - m_t\right)^{\mathrm{T}} R_{t|t}^{-1}\left(x_t - m_t\right) \\ &-\frac{1}{2}\alpha\left[x_{t+1}^{\mathrm{T}} Q_x^{-1} x_{t+1} - 2x_{t+1}^{\mathrm{T}} Q_x^{-1}\left\langle f\left(x_t,\theta\right)\right\rangle + \left\langle f\left(x_t,\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(x_t,\theta\right)\right\rangle\right] \\ &-\frac{1}{2}\sigma\left[y_{t+1}^{\mathrm{T}} Q_y^{-1} y_{t+1} - 2y_{t+1}^{\mathrm{T}} Q_y^{-1}\left\langle g\left(x_{t+1},\phi\right)\right\rangle + \left\langle g\left(x_{t+1},\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(x_{t+1},\phi\right)\right\rangle\right] \\ &-\frac{1}{2}\left(x_{t+1} - n_{t+1}\right)^{\mathrm{T}} \Omega_{t+1}^{-1}\left(x_{t+1} - n_{t+1}\right) \end{aligned}$$

We then use the usual Taylor expansions, and then focus on the quadratic terms, since they are the only terms of interest here:

$$
\begin{aligned}
\log p\left(x_{t+1}, x_t \middle| y_{1:T}\right) = & -\frac{1}{2}\left[x_t^{\mathrm{T}} R_{t|t}^{-1} x_t - 2 m_t^{\mathrm{T}} R_{t|t}^{-1} x_t + m_t^{\mathrm{T}} R_{t|t}^{-1} m_t \right] \\
& -\frac{\alpha}{2}\Bigg[x_{t+1}^{\mathrm{T}} Q_x^{-1} x_{t+1} - 2 x_{t+1}^{\mathrm{T}} Q_x^{-1} f\left(\mu_t, \mu_\theta\right) - 2 x_{t+1}^{\mathrm{T}} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) \\
& \qquad + f\left(\mu_t, \mu_\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(\mu_t, \mu_\theta\right) \\
& \qquad + 2 f\left(\mu_t, \mu_\theta\right)^{\mathrm{T}} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) + \left(x_t - \mu_t\right)^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) \\
& \qquad + j_\theta\left(\mu_t\right) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}} \left(x_t - \mu_t\right) + \frac{1}{2}\left(x_t - \mu_t\right)^{\mathrm{T}} \frac{\partial^2 j_\theta}{\partial x^2}\left(x_t - \mu_t\right) \Bigg] \\
& -\frac{\sigma}{2}\Bigg[y_{t+1}^{\mathrm{T}} Q_y^{-1} y_{t+1} - 2 y_{t+1}^{\mathrm{T}} Q_y^{-1} g\left(\mu_{t+1}, \mu_\phi\right) - 2 y_{t+1}^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \left(x_{t+1} - \mu_{t+1}\right) \\
& \qquad + g\left(\mu_{t+1}, \mu_\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_{t+1}, \mu_\phi\right) \\
& \qquad + 2 g\left(\mu_{t+1}, \mu_\phi\right)^{\mathrm{T}} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \left(x_{t+1} - \mu_{t+1}\right) + \left(x_{t+1} - \mu_{t+1}\right)^{\mathrm{T}} \frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \left(x_{t+1} - \mu_{t+1}\right) \\
& \qquad + j_\phi\left(\mu_{t+1}\right) + \frac{\partial j_\phi}{\partial x}^{\mathrm{T}} \left(x_{t+1} - \mu_{t+1}\right) + \frac{1}{2}\left(x_{t+1} - \mu_{t+1}\right)^{\mathrm{T}} \frac{\partial^2 j_\phi}{\partial x^2}\left(x_{t+1} - \mu_{t+1}\right) \Bigg] \\
& -\frac{1}{2}\left[x_{t+1}^{\mathrm{T}} \Omega_{t+1}^{-1} x_{t+1} - 2 n_{t+1}^{\mathrm{T}} \Omega_{t+1}^{-1} x_{t+1} + n_{t+1}^{\mathrm{T}} \Omega_{t+1}^{-1} n_{t+1} \right] \\
= & -\frac{1}{2} x_t^{\mathrm{T}} \left[R_{t|t}^{-1} + \alpha\left(\frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2} \right) \right] x_t \\
& -\frac{1}{2} x_{t+1}^{\mathrm{T}} \left[\Omega_{t+1}^{-1} + \alpha Q_x^{-1} + \sigma\left(\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\phi}{\partial x^2} \right) \right] x_{t+1} \\
& + \alpha\left[\frac{1}{2} x_{t+1}^{\mathrm{T}} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} x_t + \frac{1}{2} x_t^{\mathrm{T}} \frac{\partial f}{\partial x} Q_x^{-1} x_{t+1} \right] + K^0
\end{aligned}
$$

Provided that:

$$
\begin{aligned}
\log p\left(x_{t+1}, x_t \middle| y_{1:T}\right) &= -\frac{1}{2}\begin{pmatrix} x_t-\mu_t \\ x_{t+1}-\mu_{t+1}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix} \Psi_{t,t} & \Psi_{t,t+1} \\ \Psi_{t,t+1}{}^{\mathrm{T}} & \Psi_{t+1,t+1}\end{pmatrix}^{-1}\begin{pmatrix} x_t-\mu_t \\ x_{t+1}-\mu_{t+1}\end{pmatrix} \\
&= -\frac{1}{2}\begin{pmatrix} x_t \\ x_{t+1}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix} B_t & -\alpha F_t Q_x{}^{-1} \\ -\alpha Q_x{}^{-1}F_t^T & E_{t+1}\end{pmatrix}\begin{pmatrix} x_t \\ x_{t+1}\end{pmatrix} \quad + K^0 \\
B_t &= R_{t|t}{}^{-1} + \alpha\left(F_t Q_x{}^{-1}F_t^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right) \qquad , \\
E_{t+1} &= \Omega_{t+1}{}^{-1} + \alpha Q_x{}^{-1} + \sigma\left(G_{t+1}Q_y{}^{-1}G_{t+1}{}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\phi}{\partial x^2}\right) \\
G_{t+1} &= \left.\frac{\partial g}{\partial x}\right|_{\mu_{t+1}} \\
F_t &= \left.\frac{\partial f}{\partial x}\right|_{\mu_t}
\end{aligned}
$$

the use of the Schur complements yields the expression of the inter-time step covariance matrix:

$$
\begin{aligned}
\Psi_{t,t+1} &= B_t^{-1}F_t Q_x{}^{-1}\left[\frac{1}{\alpha}E_{t+1} - \alpha Q_x{}^{-1}F_t^{\mathrm{T}}B_t^{-1}F_t Q_x{}^{-1}\right]^{-1} \\
B_t &= R_{t|t}{}^{-1} + \alpha\left(F_t Q_x{}^{-1}F_t^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right) \\
E_{t+1} &= \Omega_{t+1}{}^{-1} + \alpha Q_x{}^{-1} + \sigma\left(G_{t+1}Q_y{}^{-1}G_{t+1}{}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\phi}{\partial x^2}\right) \\
\Omega_{t+1} &= \frac{1}{\alpha}\left[F_t Q_x{}^{-1}F_t^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2} - \alpha F_t Q_x{}^{-1}E_{t+2}{}^{-1}Q_x{}^{-1}F_t^{\mathrm{T}}\right]^{-1}
\end{aligned}
$$

For the last time, we use the bilinear approximation of the observation and evolution functions to provide a numerical estimation of the mean-field trace terms $j_\varphi(x)$ and $j_\theta(x)$. These are derived using Taylor expansions of $g_\varphi(x)$ and $f_\theta(x)$ respectively around μ_{t+1} and μ_t, exactly as it is done in the backward pass (sections 3.3).

3.5 The initial conditions

The initial condition determination is essentially motivated by relaxing the tight stochastic innovations precisions (which might be enforced for nearly-deterministic dynamical models) at the beginning of the time series.

$$\begin{aligned}\log q(x_0) &= -\frac{\alpha}{2}\left\langle x_1^{\mathrm{T}} Q_x^{-1} x_1 - 2x_1^{\mathrm{T}} Q_x^{-1} f(x_0,\theta) + f(x_0,\theta)^{\mathrm{T}} Q_x^{-1} f(x_0,\theta)\right\rangle - \frac{1}{2}(x_0-\mu_{00})^{\mathrm{T}} \Sigma_{00}^{-1}(x_0-\mu_{00}) \\ &= -\frac{\alpha}{2}\left(-2\mu_1^{\mathrm{T}} Q_x^{-1}\left\langle f(x_0,\theta)\right\rangle + \left\langle f(x_0,\theta)^{\mathrm{T}} Q_x^{-1} f(x_0,\theta)\right\rangle\right) - \frac{1}{2}\left(x_0^{\mathrm{T}}\Sigma_{00}^{-1}x_0 + -2\mu_{00}^{\mathrm{T}}\Sigma_{00}^{-1}x_0\right) + K\end{aligned}$$

Taylor-expand $f(x_0,\theta)$ around μ_θ:

$$f(x_0,\theta) = f(x_0,\mu_\theta) + \frac{\partial f}{\partial\theta}^{\mathrm{T}}(\theta-\mu_\theta) + O^{(2)}$$

Then:

$$\begin{aligned}\left\langle f(x_0,\theta)\right\rangle &= f(x_0,\mu_\theta) \\ \left\langle f(x_0,\theta)^{\mathrm{T}} Q_x^{-1} f(x_0,\theta)\right\rangle &= f(x_0,\mu_\theta)^{\mathrm{T}} Q_x^{-1} f(x_0,\mu_\theta) + \underbrace{tr\left[\frac{\partial f}{\partial\theta} Q_x^{-1} \frac{\partial f}{\partial\theta}^{\mathrm{T}} \Sigma_\theta\right]}_{j_\theta(x)}\end{aligned}$$

Taylor-expand $f(x_0,\theta)$ and $j_\theta(x)$ around μ_0:

$$\begin{aligned}f(x_0,\mu_\theta) &= f(\mu_0,\mu_\theta) + \frac{\partial f}{\partial x}^{\mathrm{T}}(x_0-\mu_0) + O^{(2)} \\ j_\theta(x) &= j_\theta(\mu_0) + \frac{\partial j_\theta}{\partial x}^{\mathrm{T}}(x_0-\mu_0) + \frac{1}{2}(x_0-\mu_0)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}(x_0-\mu_0)\end{aligned}$$

Then the required first- and second-order expectations write:

$$\left\langle f\left(x_0,\theta\right)\right\rangle = f\left(\mu_0,\mu_\theta\right)+\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)+O^{(2)}$$

$$\begin{aligned}\left\langle f\left(x_0,\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(x_0,\theta\right)\right\rangle &= f\left(\mu_0,\mu_\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(\mu_0,\mu_\theta\right)\\ &+2\, f\left(\mu_0,\mu_\theta\right)^{\mathrm{T}} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)+\left(x_0-\mu_0\right)^{\mathrm{T}}\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)\\ &+j_\theta\left(\mu_0\right)+\frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)+\frac{1}{2}\left(x_0-\mu_0\right)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\left(x_0-\mu_0\right)\end{aligned}$$

Then $\log q\left(x_0\right)$ writes:

$$\begin{aligned}\log q\left(x_0\right) &= -\frac{1}{2}\left(x_0^{\mathrm{T}}\Sigma_{00}^{-1}x_0+-2\mu_{00}^{\mathrm{T}}\Sigma_{00}^{-1}x_0\right)\\ &-\frac{\alpha}{2}\left(-2\,\mu_1^{\mathrm{T}}Q_x^{-1}f\left(\mu_0,\mu_\theta\right)-2\,\mu_1^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)\right.\\ &\quad+f\left(\mu_0,\mu_\theta\right)^{\mathrm{T}}Q_x^{-1}f\left(\mu_0,\mu_\theta\right)\\ &\quad+2\,f\left(\mu_0,\mu_\theta\right)^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)+\left(x_0-\mu_0\right)^{\mathrm{T}}\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)\\ &\quad\left.+j_\theta\left(\mu_0\right)+\frac{\partial j_\theta}{\partial x}^{\mathrm{T}}\left(x_0-\mu_0\right)+\frac{1}{2}\left(x_0-\mu_0\right)^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\left(x_0-\mu_0\right)\right)+K\\ &=-\frac{1}{2}x_0^{\mathrm{T}}\left[\Sigma_{00}^{-1}+\alpha\left(\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}+\frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right)\right]x_0\\ &-\frac{1}{2}\left[-2\mu_{00}^{\mathrm{T}}\Sigma_{00}^{-1}\right.\\ &\quad\left.+\alpha\left(-2\,\mu_1^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}+2\,f\left(\mu_0,\mu_\theta\right)^{\mathrm{T}}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}-2\mu_0^{\mathrm{T}}\frac{\partial f}{\partial x}Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}+\frac{\partial j_\theta}{\partial x}^{\mathrm{T}}-\mu_0^{\mathrm{T}}\frac{\partial^2 j_\theta}{\partial x^2}\right)\right]x_0+K'\\ &=-\frac{1}{2}x_0^{\mathrm{T}}B\,x_0-A^{\mathrm{T}}x_0+K'=-\frac{1}{2}\left(x_0-\nu\right)^{\mathrm{T}}B\left(x_0-\nu\right)+K''\end{aligned}$$

Where the last line is derived from the QF transform, given that:

$$\begin{aligned}
\upsilon &= -B^{-1}A \\
B &= \Sigma_{00}{}^{-1} + \alpha\left(\frac{\partial f}{\partial x}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{1}{2}\frac{\partial^2 j_\theta}{\partial x^2}\right) \\
A &= \left[-\Sigma_{00}{}^{-1}\mu_{00} + \alpha\left(-\frac{\partial f}{\partial x}Q_x{}^{-1}\mu_1 + \frac{\partial f}{\partial x}Q_x{}^{-1}\left(f\left(\mu_0,\mu_\theta\right) - \frac{\partial f}{\partial x}^{\mathrm{T}}\mu_0\right) + \frac{1}{2}\left(\frac{\partial j_\theta}{\partial x} - \frac{\partial^2 j_\theta}{\partial x^2}^{\mathrm{T}}\mu_0\right)\right)\right]
\end{aligned}$$

This finally yields:

$$\begin{aligned}
\mu_0{}^{(k+1)} - \mu_0{}^{(k)} &= \Sigma_0{}^{(k+1)}\left[\Sigma_{00}{}^{-1}\left(\mu_{00} - \mu_0{}^{(k)}\right) + \alpha\left(\frac{\partial f}{\partial x}Q_x{}^{-1}\left(\mu_1 - f\left(\mu_0,\mu_\theta\right)\right) - \frac{\partial^2 \tilde{f}}{\partial x \partial\theta}\left[Q_x{}^{-1}\otimes\Sigma_\theta\right]\frac{\partial \tilde{f}}{\partial\theta}\right)\right] \\
\Psi_{0,0}{}^{(k+1)} &= \left[\alpha\left(\frac{\partial f}{\partial x}Q_x{}^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}} + \frac{\partial^2 \tilde{f}}{\partial x\partial\theta}\left[Q_x{}^{-1}\otimes\Sigma_\theta\right]\frac{\partial^2 \tilde{f}}{\partial x\partial\theta}^{\mathrm{T}}\right) + \Sigma_{00}{}^{-1}\right]^{-1}
\end{aligned}$$

We chose a notation that renders the initial condition posterior $q(x_0) = N(\mu_0, \Psi_{0,0})$ consistent with the hidden-states variational posterior $q(x_{1:n_t})$.

Note that the uncertainty associated with the initial condition on the states enforces a careful application of the first prediction update (ie $\alpha^*(x_1)$), as well as the evolution parameters and stochastic innovations precision VB update:

$$\log\alpha^*\left(x_1\right) \approx \left(x_1 - m_1^*\right)^{\mathrm{T}} R_{1|0}{}^{-1}\left(x_1 - m_1^*\right) + K ,$$

where:

$$\begin{aligned}
R_{1|0} &= \frac{\partial f}{\partial x}\Psi_{0,0}\frac{\partial f}{\partial x}^{T} + \frac{1}{\alpha}Q_x . \\
m_1^* &= f\left(\mu_0,\mu_\theta\right)
\end{aligned}$$

The associated update follows from the usual formulae given bellow (see filtered density).

Precision parameters

4.1 Measurement noise

The variational posterior pdf of the measurement noise precision σ writes:

$$
\begin{aligned}
\log q(\sigma) &= -\frac{\sigma}{2}\left\langle \sum_{t=1}^{T}\left(y_t - g\left(x_t,\phi\right)\right)^{\mathrm{T}} Q_y^{-1}\left(y_t - g\left(x_t,\phi\right)\right)\right\rangle + \frac{pT}{2}\log\sigma - b_{\sigma 0}\sigma + \left(a_{\sigma 0}-1\right)\log\sigma \\
&= -\left(\frac{1}{2}\left\langle \sum_{t=1}^{T}\left(y_t - g\left(x_t,\phi\right)\right)^{\mathrm{T}} Q_y^{-1}\left(y_t - g\left(x_t,\phi\right)\right)\right\rangle + b_{\sigma 0}\right)\sigma + \left(\frac{pT}{2} + a_{\sigma 0} - 1\right)\log\sigma
\end{aligned}
.
$$

We use the previous Laplace approximation to the bracketed term:

$$
\begin{aligned}
\left\langle \sum_{t=1}^{T}\left(y_t - g\left(x_t,\phi\right)\right)^{\mathrm{T}} Q_y^{-1}\left(y_t - g\left(x_t,\phi\right)\right)\right\rangle \approx \sum_{t=1}^{T}\Bigg\langle & y_t^{\mathrm{T}} Q_y^{-1} y_t - 2 y_t^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) - 2 y_t^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial\phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\
& + g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) + 2 g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1}\frac{\partial g}{\partial\phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\
& + \left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial g}{\partial\phi} Q_y^{-1}\frac{\partial g}{\partial\phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) \\
& + j_x\left(\mu_\phi\right) + \frac{\partial j_x}{\partial\phi}^{\mathrm{T}}\left(\phi-\mu_\phi\right) + \frac{1}{2}\left(\phi-\mu_\phi\right)^{\mathrm{T}}\frac{\partial^2 j_x}{\partial\phi^2}\left(\phi-\mu_\phi\right)\Bigg\rangle \\
= \sum_{t=1}^{T}\Bigg(& y_t^{\mathrm{T}} Q_y^{-1} y_t - 2 y_t^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) + g\left(\mu_t,\mu_\phi\right)^{\mathrm{T}} Q_y^{-1} g\left(\mu_t,\mu_\phi\right) \\
& + tr\left[\left(\frac{\partial g}{\partial\phi} Q_y^{-1}\frac{\partial g}{\partial\phi}^{\mathrm{T}} + \frac{\partial^2\tilde{g}}{\partial x\partial\phi}\left[Q_y^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{g}}{\partial x\partial\phi}^{\mathrm{T}}\right)\Sigma_\phi\right] + tr\left[\frac{\partial g}{\partial x} Q_y^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}}\Psi_{t,t}\right]\Bigg) \\
= \sum_{t=1}^{T} & \left(y_t - g\left(\mu_t,\mu_\phi\right)\right)^{\mathrm{T}} Q_y^{-1}\left(y_t - g\left(\mu_t,\mu_\phi\right)\right) \\
& + \sum_{t=1}^{T} tr\left[\left(\frac{\partial g}{\partial\phi} Q_y^{-1}\frac{\partial g}{\partial\phi}^{\mathrm{T}} + \frac{\partial^2\tilde{g}}{\partial x\partial\phi}\left[Q_y^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{g}}{\partial x\partial\phi}^{\mathrm{T}}\right)\Sigma_\phi\right] + \sum_{t=1}^{T} tr\left[\frac{\partial g}{\partial x} Q_y^{-1}\frac{\partial g}{\partial x}^{\mathrm{T}}\Psi_{t,t}\right]
\end{aligned}
$$

Therefore, the variational posterior of σ is of the form: $q(\sigma) = Ga\left(\sigma \middle| a_\sigma, b_\sigma\right)$ with:

$$\begin{cases} a_\sigma = a_{\sigma 0} + \dfrac{pT}{2} \\ b_\sigma = b_{\sigma 0} + \dfrac{1}{2}\sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathrm{T}} Q_y^{-1} \hat{\varepsilon}_t + \dfrac{1}{2}\sum_{t=1}^{T} tr\left[\left(\dfrac{\partial g}{\partial \phi} Q_y^{-1} \dfrac{\partial g}{\partial \phi}^{\mathrm{T}} + \dfrac{\partial^2 \tilde{g}}{\partial x \partial \phi}\left[Q_y^{-1} \otimes \Psi_{t,t}\right] \dfrac{\partial^2 \tilde{g}}{\partial x \partial \phi}^{\mathrm{T}}\right)\Sigma_\phi\right] + \dfrac{1}{2}\sum_{t=1}^{T} tr\left[\dfrac{\partial g}{\partial x} Q_y^{-1} \dfrac{\partial g}{\partial x}^{\mathrm{T}} \Psi_{t,t}\right] \end{cases}$$

where $\hat{\varepsilon}_t = y_t - g\left(\mu_t, \mu_\phi\right)$

4.2 Stochastic innovations (state noise)

Similarly:

$$\begin{aligned} \log q(\alpha) &= -\frac{\alpha}{2}\left\langle \sum_{t=1}^{T}\left(x_t - f\left(x_{t-1},\theta\right)\right)^{\mathrm{T}} Q_x^{-1}\left(x_t - f\left(x_{t-1},\theta\right)\right)\right\rangle + \frac{nT}{2}\log\alpha - b_{\alpha 0}\alpha + \left(a_{\alpha 0} - 1\right)\log\alpha \\ &= -\left(\frac{1}{2}\left\langle \sum_{t=1}^{T}\left(x_t - f\left(x_{t-1},\theta\right)\right)^{\mathrm{T}} Q_x^{-1}\left(x_t - f\left(x_{t-1},\theta\right)\right)\right\rangle + b_{\alpha 0}\right)\alpha + \left(\frac{nT}{2} + a_{\alpha 0} - 1\right)\log\alpha \end{aligned},$$

Using the Laplace approximation for the bracketed terms:

$$\begin{aligned}
\left\langle \sum_{t=1}^{T}\left(x_t - f\left(x_{t-1},\theta\right)\right)^{\mathrm{T}} Q_x^{-1}\left(x_t - f\left(x_{t-1},\theta\right)\right)\right\rangle &\approx \sum_{t=0}^{T-1}\left(\left\langle x_{t+1}{}^{\mathrm{T}} Q_x^{-1} x_{t+1}\right\rangle - 2\left\langle x_{t+1}{}^{\mathrm{T}} Q_x^{-1} f\left(x_t,\theta\right)\right\rangle + \left\langle f\left(x_t,\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(x_t,\theta\right)\right\rangle\right) \\
&= \sum_{t=1}^{T} \mu_t{}^{\mathrm{T}} Q_x^{-1}\mu_t + \sum_{t=1}^{T} tr\left[Q_x^{-1}\Psi_{t,t}\right] \\
&\quad -2\sum_{t=0}^{T-1}\left(\mu_{t+1}{}^{\mathrm{T}} Q_x^{-1} f\left(\mu_t,\mu_\theta\right) + tr\left[Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t+1}\right]\right) \\
&\quad +\sum_{t=0}^{T-1}\left(f\left(\mu_t,\mu_\theta\right)^{\mathrm{T}} Q_x^{-1} f\left(\mu_t,\mu_\theta\right) + tr\left[\left(\frac{\partial f}{\partial\theta} Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} + \frac{\partial^2\tilde{f}}{\partial x\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{f}}{\partial x\partial\theta}^{\mathrm{T}}\right)\Sigma_\theta\right]\right. \\
&\qquad\left. + tr\left[\frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t}\right]\right) \\
&= \sum_{t=1}^{T}\left(\mu_t - f\left(\mu_{t-1},\mu_\theta\right)\right)^{\mathrm{T}} Q_x^{-1}\left(\mu_t - f\left(\mu_{t-1},\mu_\theta\right)\right) \\
&\quad +\sum_{t=0}^{T-1} tr\left[\left(\frac{\partial f}{\partial\theta} Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} + \frac{\partial^2\tilde{f}}{\partial x\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{f}}{\partial x\partial\theta}^{\mathrm{T}}\right)\Sigma_\theta\right] \\
&\quad +\sum_{t=1}^{T-1} tr\left[\left(Q_x^{-1} + \frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\right)\Psi_{t,t}\right] + tr\left[Q_x^{-1}\Psi_{T,T}\right] + tr\left[\frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{0,0}\right] \\
&\quad -2\sum_{t=1}^{T-1} tr\left[Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t+1}\right]
\end{aligned}$$

Therefore, the variational posterior of α is of the form: $q(\alpha) = Ga\left(\alpha \middle| a_\alpha, b_\alpha\right)$ with:

$$\left\{\begin{aligned}
a_\alpha &= a_{\alpha 0} + \frac{nT}{2} \\
b_\alpha &= b_{\alpha 0} + \frac{1}{2}\sum_{t=0}^{T}\hat{\eta}_t{}^{\mathrm{T}} Q_x^{-1}\hat{\eta}_t + \frac{1}{2}\sum_{t=0}^{T-1} tr\left[\left(\frac{\partial f}{\partial\theta} Q_x^{-1}\frac{\partial f}{\partial\theta}^{\mathrm{T}} + \frac{\partial^2\tilde{f}}{\partial x\partial\theta}\left[Q_x^{-1}\otimes\Psi_{t,t}\right]\frac{\partial^2\tilde{f}}{\partial x\partial\theta}^{\mathrm{T}}\right)\Sigma_\theta\right] \\
&\quad + \frac{1}{2}\sum_{t=1}^{T-1} tr\left[\left(Q_x^{-1} + \frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\right)\Psi_{t,t}\right] + \frac{1}{2} tr\left[\frac{\partial f}{\partial x} Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{0,0} + Q_x^{-1}\Psi_{T,T}\right] - \sum_{t=1}^{T-1} tr\left[Q_x^{-1}\frac{\partial f}{\partial x}^{\mathrm{T}}\Psi_{t,t+1}\right]
\end{aligned}\right.$$

where $\hat{\eta}_t = \mu_t - f\left(\mu_{t-1},\mu_\theta\right)$ and $\Psi_{0,1} = 0$.

Free energy evaluation

The free energy is defined as:

$$
\begin{aligned}
F &= \left\langle \log p\left(\vartheta, y_{1:T} \left| m\right.\right)\right\rangle_{\prod_i q(\vartheta_i)} + \sum_i S\left(q\left(\vartheta_i\right)\right) \\
&\approx \log p\left(\mu, y_{1:T} \left| m\right.\right) + \frac{1}{2}\sum_{i=1}^{n_\vartheta} tr\left[\frac{\partial^2}{\partial \vartheta_i^{\,2}} \log p\left(\vartheta, y_{1:T} \left| m\right.\right)\bigg|_{\vartheta_i=\mu_i} \Sigma_i \right] + \sum_{i=1}^{n_\vartheta} S\left(q\left(\vartheta_i\right)\right),
\end{aligned}
$$

where ϑ_i is the *i*th subset of unknown variable which posterior has to be characterized.

Let us write the joint pdf associated to the generative model m:

$$
p\left(x_{1:T}, \theta, \varphi, \alpha, \sigma, y_{1:T} \left| x_0, m\right.\right) = p(\alpha)p(\sigma)p(\theta)p(\varphi)\prod_{t=1}^{T} p\left(x_t \left| x_{t-1}, \theta, \alpha\right.\right)p\left(y_t \left| x_t, \theta, \alpha\right.\right)
$$

Given that:

$$
\begin{aligned}
\left\langle \log \prod_{t=1}^{T} p\left(x_t \left| x_{t-1}, \theta, \alpha\right.\right) p\left(y_t \left| x_t, \theta, \alpha\right.\right)\right\rangle &= -\frac{1}{2}\left\langle \alpha \sum_{t=1}^{T}\left(x_t - f\left(x_{t-1}, \theta\right) \right)^{\mathrm{T}} Q_x^{-1}\left(x_t - f\left(x_{t-1}, \theta\right) \right)\right\rangle \\
&\quad -\frac{1}{2}\left\langle \sigma \sum_{t=1}^{T}\left(y_t - g\left(x_t, \phi\right) \right)^{\mathrm{T}} Q_y^{-1}\left(y_t - g\left(x_t, \phi\right) \right)\right\rangle \\
&\quad \underbrace{+\frac{nT}{2}\left(\left\langle \log \alpha\right\rangle - \log\left|Q_x\right| - \log\left(2\pi\right)\right)}_{\text{normalization constant}} \underbrace{+\frac{pT}{2}\left(\left\langle \log \sigma\right\rangle - \log\left|Q_y\right| - \log\left(2\pi\right)\right)}_{\text{normalization constant}}
\end{aligned}
$$

The expectations involved are derived from the sufficient statistics calculated during the VB updates:

$$\left\langle \sigma \sum_{t=1}^{T} \left(y_t - g\left(x_t,\phi\right) \right)^{\mathrm{T}} Q_y^{-1} \left(y_t - g\left(x_t,\phi\right) \right) \right\rangle = \langle\sigma\rangle \sum_{t=1}^{T} \hat{\varepsilon}_{\mathrm{t}}^{\mathrm{T}} Q_y^{-1} \hat{\varepsilon}_{\mathrm{t}}$$

$$+\langle\sigma\rangle \sum_{t=1}^{T} tr\left[\left(\frac{\partial g}{\partial \phi} Q_y^{-1} \frac{\partial g}{\partial \phi}^{\mathrm{T}} + \frac{\partial^2 \tilde{g}}{\partial x \partial \phi} \left[Q_y^{-1} \otimes \Psi_{t,t} \right] \frac{\partial^2 \tilde{g}}{\partial x \partial \phi}^{\mathrm{T}} \right) \Sigma_\phi \right]$$

$$+\langle\sigma\rangle \sum_{t=1}^{T} tr\left[\frac{\partial g}{\partial x} Q_y^{-1} \frac{\partial g}{\partial x}^{\mathrm{T}} \Psi_{t,t} \right]$$

$$\left\langle \alpha \sum_{t=1}^{T} \left(x_t - f\left(x_{t-1},\theta\right) \right)^{\mathrm{T}} Q_x^{-1} \left(x_t - f\left(x_{t-1},\theta\right) \right) \right\rangle = \langle\alpha\rangle \sum_{t=0}^{T} \hat{\eta}_t^{\mathrm{T}} Q_x^{-1} \hat{\eta}_t$$

$$+\langle\alpha\rangle \sum_{t=0}^{T-1} tr\left[\left(\frac{\partial f}{\partial \theta} Q_x^{-1} \frac{\partial f}{\partial \theta}^{\mathrm{T}} + \frac{\partial^2 \tilde{f}}{\partial x \partial \theta} \left[Q_x^{-1} \otimes \Psi_{t,t} \right] \frac{\partial^2 \tilde{f}}{\partial x \partial \theta}^{\mathrm{T}} \right) \Sigma_\theta \right]$$

$$+\langle\alpha\rangle \sum_{t=1}^{T-1} tr\left[\left(Q_x^{-1} + \frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \right) \Psi_{t,t} \right]$$

$$+\langle\alpha\rangle tr\left[\frac{\partial f}{\partial x} Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \Psi_{0,0} + Q_x^{-1} \Psi_{T,T} \right] - 2\sum_{t=1}^{T-1} tr\left[Q_x^{-1} \frac{\partial f}{\partial x}^{\mathrm{T}} \Psi_{t,t+1} \right]$$

Then the free energy writes:

$$\begin{aligned}
F_{T_e} \approx & \left\langle \log \prod_{t=1}^{T} p\left(x_t \middle| x_{t-1},\theta,\alpha\right) p\left(y_t \middle| x_t,\theta,\alpha\right) \right\rangle \\
& \underbrace{-\frac{1}{2}\left(\mu_\theta-\mu_{\theta 0}\right)^{\mathrm{T}} Q_{\theta 0}^{-1}\left(\mu_\theta-\mu_{\theta 0}\right) - \frac{1}{2} tr\left[Q_{\theta 0}^{-1}\Sigma_\theta\right] - \frac{1}{2}\log\left|Q_{\theta 0}\right| - \frac{n_\theta}{2}\log\left(2\pi\right)}_{\text{expected log prior of } \theta} \\
& \underbrace{-\frac{1}{2}\left(\mu_\phi-\mu_{\phi 0}\right)^{\mathrm{T}} Q_{\phi 0}^{-1}\left(\mu_\phi-\mu_{\phi 0}\right) - \frac{1}{2} tr\left[Q_{\phi 0}^{-1}\Sigma_\phi\right] - \frac{1}{2}\log\left|Q_{\phi 0}\right| - \frac{n_\phi}{2}\log\left(2\pi\right)}_{\text{expected log prior of } \phi} \\
& \underbrace{-\frac{1}{2}\left(\mu_0-\mu_{00}\right)^{\mathrm{T}} \Sigma_{00}^{-1}\left(\mu_0-\mu_{00}\right) - \frac{1}{2} tr\left[\Sigma_{00}^{-1}\Psi_{0,0}\right] - \frac{1}{2}\log\left|\Sigma_{00}\right| - \frac{n}{2}\log\left(2\pi\right)}_{\text{hidden=states initial condition}} \\
& \underbrace{-b_{\alpha 0}\mu_\alpha + \left(a_{\alpha 0}-1\right)\left(\psi\left(a_\alpha\right)-\log\left(b_\alpha\right)\right) + a_{\alpha 0}\log\left(b_{\alpha 0}\right) - \log\Gamma\left(a_{\alpha 0}\right)}_{\text{expected log prior of } \alpha} \\
& \underbrace{-b_{\sigma 0}\mu_\sigma + \left(a_{\sigma 0}-1\right)\left(\psi\left(a_\sigma\right)-\log\left(b_\sigma\right)\right) + a_{\sigma 0}\log\left(b_{\sigma 0}\right) - \log\Gamma\left(a_{\sigma 0}\right)}_{\text{expected log prior of } \sigma} \\
& + S\left(q\left(x_{1:T}\right)\right) + S\left(q\left(x_0\right)\right) + S\left(q\left(\theta\right)\right) + S\left(q\left(\phi\right)\right) + S\left(q\left(\alpha\right)\right) + S\left(q\left(\sigma\right)\right)
\end{aligned}$$

Therefore, all that remains to be calculated are the entropy terms, ie:

$$\begin{aligned}
S(q(x_{1:T})) &= \underbrace{\frac{1}{2}\sum_{t=1}^{T-1}\left(\log\begin{vmatrix}\Psi_{t+1,t+1} & \Psi_{t+1,t}{}^{\mathsf{T}} \\ \Psi_{t+1,t} & \Psi_{t,t}\end{vmatrix} - \log\left|\Psi_{t,t}\right|\right) + \frac{1}{2}\log\left|\Psi_{1,1}\right| + \frac{nT}{2}(\log(2\pi)+1)}_{\text{makes use of the conditional independence structure of the states variational posterior (see MSM)}} \\
S(q(x_0)) &= \frac{n}{2}(\log(2\pi)+1) + \frac{1}{2}\log\left|\Psi_{0,0}\right| \\
S(q(\theta)) &= \frac{n_\theta}{2}(\log(2\pi)+1) + \frac{1}{2}\log\left|\Sigma_\theta\right| \\
S(q(\varphi)) &= \frac{n_\varphi}{2}(\log(2\pi)+1) + \frac{1}{2}\log\left|\Sigma_\varphi\right| \\
S(q(\alpha)) &= \log\Gamma(a_\alpha) - \log(b_\alpha) + (1-a_\alpha)\Psi(a_\alpha) + a_\alpha \\
S(q(\sigma)) &= \log\Gamma(a_\sigma) - \log(b_\sigma) + (1-a_\sigma)\Psi(a_\sigma) + a_\sigma
\end{aligned}$$

This completes the technical description of the VB algorithm for nonlinear state-space models.

Appendix

Let $Q(x) = -\frac{1}{2}(x-a)^{\mathrm{T}} B(x-a) - A^{\mathrm{T}}(x-a)$.

Then: $Q(x) = -\frac{1}{2}(x-\mu)^{\mathrm{T}} B(x-\mu) + \frac{1}{2} A^{\mathrm{T}} B^{-1} A$ with: $\mu = a - B^{-1} A$.

Quadratic form (*QF*)

Let $Q(x, y)$ be defined as:

$$
\begin{aligned}
Q(x, y) &= x^{\mathrm{T}} A_1 x + y^{\mathrm{T}} A_2 y - y^{\mathrm{T}} B^{\mathrm{T}} x - x^{\mathrm{T}} B y \\
&= \begin{pmatrix} x \\ y \end{pmatrix}^{\mathrm{T}} \underbrace{\begin{pmatrix} A_1 & -B \\ -B^{\mathrm{T}} & A_2 \end{pmatrix}}_{U} \begin{pmatrix} x \\ y \end{pmatrix}
\end{aligned}
$$

Then, if we partition $U^{-1} = \begin{pmatrix} A^0 & \Psi \\ B^0 & C^0 \end{pmatrix}$, then the Schur complements yield:

$$\psi = A_1^{-1} B \left[A_2 - B^{\mathrm{T}} A_1^{-1} B \right]^{-1}$$

Schur complements

$$\left(P^{-1} + B^{\mathrm{T}} R^{-1} B\right)^{-1} B^{\mathrm{T}} R^{-1} = P B^{\mathrm{T}} \left(B P B^{\mathrm{T}} + R\right)^{-1}$$

Woodbury matrix inversion lemna

If x behaves as a Gamma-variate, i.e.: $x \sim Ga(x|a, b)$, where (a, b) are the shape and scale parameters, respectively, then, the expectation, variance and entropy of x are given by:

$$
\begin{cases}
\langle x \rangle = \dfrac{a}{b} \\
\left\langle \left(x - \langle x \rangle\right)^2 \right\rangle = \dfrac{a}{b^2} \\
S\left(Ga(x|a, b)\right) = \log \Gamma(a) - \log b + (1-a)\Psi(a) + a
\end{cases}
$$

Gamma variates