Derivations for GFX node

Wouter Kouw

November 24, 2020

1 Problem

Suppose we have the following linear second-order differential equation:

$$x''(t) + \theta_2 x'(t) + \theta_1 x(t) = \eta u(t) + w(t) \tag{1}$$

Reduce this to a multivariate first-order differential equation with the following substitutions:

$$z_1(t) = x(t) \tag{2a}$$
$$z_2(t) = x'(t)\,, \tag{2b}$$

which produces:

$$z_1'(t) = z_2(t) \tag{3a}$$
$$z_2'(t) = -\theta_2 z_2(t) - \theta_1 z_1(t) + \eta u(t) + w(t)\,. \tag{3b}$$

We can re-write this into a matrix form:

$$\begin{bmatrix} z_1'(t) \\ z_2'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\theta_1 & -\theta_2 \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)\,, \tag{4}$$

where $z_t = \begin{bmatrix} z_1(t) & z_2(t) \end{bmatrix}^\top$.

1.1 Discretization

We can perform an approximate discretization using Euler-Maruyama, at non-overlapping time points t_k and t_{k+1}. The Wiener process $w(t)$ becomes:

$$w(t) \approx \frac{\beta_{t_{k+1}} - \beta_{t_k}}{\Delta t_k}\,, \tag{5}$$

where $\Delta t_k = t_{k+1} - t_k$. I rename the increment $\beta_{t_{k+1}} - \beta_{t_k}$ to be w_t, which, if the time-points t_{k+1} and t_k do not overlap, follows a Gaussian distribution (Def. 4.1 [1]):

$$w_t \sim \mathcal{N}(0, \tau^{-1}\Delta t_k)\,. \tag{6}$$

Henceforth, we will assume all Δt_k are equal and use subscripts $t+1$ and t. The states $z'(t)$ are approximated as $(z_{t+1} - z_t)/\Delta t$. The control signal $u(t)$ is observed and directly maps to u_t. Using this discretization, we get:

$$\left(\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} - \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix}\right)/\Delta t = \begin{bmatrix} 0 & 1 \\ -\theta_1 & -\theta_2 \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta \end{bmatrix} u_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{w_t}{\Delta t} \tag{7a}$$

$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} - \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \Delta t \\ -\theta_1 \Delta t & -\theta_2 \Delta t \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta \Delta t \end{bmatrix} u_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_t \tag{7b}$$

$$\underbrace{\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix}}_{z_{t+1}} = \underbrace{\begin{bmatrix} 1 & \Delta t \\ -\theta_1 \Delta t & -\theta_2 \Delta t + 1 \end{bmatrix}}_{A(\theta)} \underbrace{\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix}}_{z_t} + \underbrace{\begin{bmatrix} 0 \\ \eta \Delta t \end{bmatrix}}_{B(\eta)} u_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_t \tag{7c}$$

The above discrete-time state transition can be cast to a Gaussian distribution as:

$$z_{t+1} \sim \mathcal{N}(A(\theta) z_t + B(\eta) u_t, Q)\,. \tag{8}$$

for

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & \tau^{-1} \Delta t \end{bmatrix}, \; A(\theta) = \begin{bmatrix} 1 & \Delta t \\ -\theta_1 \Delta t & -\theta_2 \Delta t + 1 \end{bmatrix}, \text{ and } B(\eta) = \begin{bmatrix} 0 \\ \eta \Delta t \end{bmatrix}. \tag{9}$$

The expectation of the inverse of Q is obtained by a noise injection:

$$\mathbb{E}_{q(\tau)} Q^{-1} = \mathbb{E}_{q(\tau)} \begin{bmatrix} 0 & 0 \\ 0 & \tau^{-1} \Delta t \end{bmatrix}^{-1} \approx \mathbb{E}_{q(\tau)} \begin{bmatrix} \epsilon & 0 \\ 0 & \tau^{-1} \Delta t \end{bmatrix}^{-1} = \begin{bmatrix} \epsilon^{-1} & 0 \\ 0 & \tau/\Delta t \end{bmatrix} \triangleq m_\Lambda\,. \tag{10}$$

(wk) It might be important to have a covariance matrix Q without the 0 on the diagonal. That depends on whether the noise should also apply to the substitution in Equation 3a.

The matrix A can be constructed from a vector $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$ as:

$$A(\theta) = S + s\theta^\top, \quad S = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad s = \begin{bmatrix} 0 \\ -\Delta t \end{bmatrix}. \tag{11}$$

1.2 Recognition model

I choose the following recognition factors:

$$z_{t-1} \sim \mathcal{N}(m_{z_{t-1}}, V_{z_{t-1}}) \tag{12a}$$
$$z_t \sim \mathcal{N}(m_{z_t}, V_{z_t}) \tag{12b}$$
$$\theta \sim \mathcal{N}(m_\theta, V_\theta) \tag{12c}$$
$$\eta \sim \mathcal{N}(m_\eta, v_\eta) \tag{12d}$$
$$\tau \sim \Gamma(a_\tau, b_\tau)\,. \tag{12e}$$

2 Messages

In computing the messages, the following results are useful:

$$\mathbb{E}_{q(\theta)}A(\theta) = \mathbb{E}_{q(\theta)}\left[S + s\theta^\top\right] = S + sm_\theta^\top = A(m_\theta) \tag{13a}$$

$$\mathbb{E}_{q(\eta)}B(\eta) = \mathbb{E}_{q(\eta)}\begin{bmatrix}0\\ \Delta t\end{bmatrix}\eta = \begin{bmatrix}0\\ \Delta t\end{bmatrix}m_\eta = B(m_\eta)\,. \tag{13b}$$

Message to z_t

$$\overrightarrow{\nu}(z_t) = \mathbb{E}_{q(z_{t-1})q(\theta)q(\eta)q(\tau)}\log\mathcal{N}\big(z_t \mid A(\theta)z_{t-1} + B(\eta)u_t, Q\big) \tag{14a}$$

$$\propto -\frac{1}{2}\mathbb{E}\big[\big(z_t - A(\theta)z_{t-1} - B(\eta)u_t\big)^\top Q^{-1}\big(z_t - A(\theta)z_{t-1} - B(\eta)u_t\big)\big] \tag{14b}$$

$$\begin{aligned}\propto -\frac{1}{2}\mathbb{E}\big[&\underbrace{z_t^\top Q^{-1}z_t}_{①} - \underbrace{(A(\theta)z_{t-1})^\top Q^{-1}z_t}_{②} - \underbrace{(B(\eta)u_t)^\top Q^{-1}z_t}_{③}\\ &- \underbrace{z_t^\top Q^{-1}A(\theta)z_{t-1}}_{④} - \underbrace{z_tQ^{-1}B(\eta)u_t}_{⑤}\big]\,.\end{aligned} \tag{14c}$$

With the terms:

$$① = z_t^\top m_\Lambda z_t \tag{15a}$$

$$② = (A(m_\theta)m_{z_{t-1}})^\top m_\Lambda z_t \tag{15b}$$

$$③ = (B(m_\eta)u_t)^\top m_\Lambda z_t \tag{15c}$$

$$④ = z_t^\top m_\Lambda A(m_\theta)m_{z_{t-1}} \tag{15d}$$

$$⑤ = z_t^\top m_\Lambda B(m_\eta)u_t\,. \tag{15e}$$

Plugging those back in:

$$\begin{aligned}\overrightarrow{\nu}(z_t) \propto -\frac{1}{2}\big[&z_t^\top\underbrace{m_\Lambda}_{\Phi}z_t - \big(A(m_\theta)m_{z_{t-1}} + B(m_\eta)u_t\big)m_\Lambda z_t\\ &- z_t^\top m_\Lambda\underbrace{(A(m_\theta)m_{z_{t-1}} + B(m_\eta)u_t)}_{\phi}\big]\end{aligned} \tag{16}$$

$$\sim\mathcal{N}(\phi, \Phi^{-1})\,. \tag{17}$$

Message to z_{t-1}

Message to θ

$$\overrightarrow{\nu}(\theta) = \mathbb{E}_{q(z_t)q(z_{t-1})q(\eta)q(\tau)} \log \mathcal{N}\big(z_t \mid A(\theta)z_{t-1} + B(\eta)u_t, Q\big) \tag{18a}$$

$$\propto -\frac{1}{2}\mathbb{E}\big[\big(z_t - A(\theta)z_{t-1} - B(\eta)u_t\big)^\top Q^{-1}\big(z_t - A(\theta)z_{t-1} - B(\eta)u_t\big)\big] \tag{18b}$$

$$\begin{aligned}\propto -\frac{1}{2}\mathbb{E}\big[&- \underbrace{(A(\theta)z_{t-1})^\top Q^{-1} z_t}_{①} + \underbrace{(A(\theta)z_{t-1})^\top Q^{-1} A(\theta) z_{t-1}}_{②} \\ &+ \underbrace{(A(\theta)z_{t-1})^\top Q^{-1} B(\eta)u_t}_{③} - \underbrace{z_t^\top Q^{-1} A(\theta) z_{t-1}}_{④} \\ &+ \underbrace{(B(\eta)u_t)^\top Q^{-1} A(\theta) z_{t-1}}_{⑤}\big]\,. \end{aligned} \tag{18c}$$

With the terms:

$$① = \mathbb{E}_{q(z_{t-1})q(z_t)q(\tau)}\big[(A(\theta)z_{t-1})^\top Q^{-1} z_t\big] \tag{19a}$$
$$= ((S + s\theta^\top)m_{z_{t-1}})^\top m_\Lambda m_{z_t} \tag{19b}$$
$$= m_{z_{t-1}}^\top \theta s^\top m_\Lambda m_{z_t} \tag{19c}$$
$$= \theta^\top m_{z_{t-1}} s^\top m_\Lambda m_{z_t} \tag{19d}$$
$$③ = \mathbb{E}_{q(z_{t-1})q(\tau)q(\eta)}\big[A(\theta)z_{t-1})^\top Q^{-1} B(\eta)u_t\big] \tag{19e}$$
$$= (S + s\theta^\top m_{z_{t-1}})^\top m_\Lambda B(m_\eta)u_t \tag{19f}$$
$$= m_{z_{t-1}}^\top \theta s^\top m_\Lambda B(m_\eta)u_t \tag{19g}$$
$$= \theta^\top m_{z_{t-1}} s^\top m_\Lambda B(m_\eta)u_t \tag{19h}$$
$$④ = \mathbb{E}_{q(z_{t-1})q(z_t)q(\tau)}\big[z_t^\top Q^{-1} A(\theta) z_{t-1}\big] \tag{19i}$$
$$= m_{z_t}^\top m_\Lambda (S + s\theta^\top) m_{z_{t-1}} \tag{19j}$$
$$= m_{z_t}^\top m_\Lambda s m_{z_{t-1}}^\top \theta \tag{19k}$$
$$⑤ = \mathbb{E}_{q(\eta)q(\tau)q(z_{t-1})}\big[(B(\eta)u_t)^{-1} Q^{-1} A(\theta) z_{t-1}\big] \tag{19l}$$
$$= (B(m_\eta)u_t)^\top m_\Lambda (S + s\theta^\top) m_{z_{t-1}} \tag{19m}$$
$$= (B(m_\eta)u_t)^\top m_\Lambda s m_{z_{t-1}}^\top \theta\,. \tag{19n}$$

The message (2) is more difficult:

$$
\begin{aligned}
(2) &= \mathbb{E}_{q(z_{t-1})q(\tau)}\big[(A(\theta)z_{t-1})^\top Q^{-1} A(\theta) z_{t-1}\big] && \text{(20a)}\\
&= \mathbb{E}\big[(z_{t-1}^\top (S + s\theta^\top)^\top m_\Lambda (S + s\theta^\top) z_{t-1}\big] && \text{(20b)}\\
&\propto \mathbb{E}\big[\underbrace{z_{t-1}^\top \theta s^\top m_\Lambda S z_{t-1}}_{(A)}\\
&\quad + \underbrace{z_{t-1}^\top \theta s^\top m_\Lambda s\theta^\top z_{t-1}}_{(B)}\\
&\quad + \underbrace{z_{t-1}^\top S^\top m_\Lambda s\theta^\top z_{t-1}}_{(C)}\big]\\
(A) &= \mathbb{E}_{q(z_{t-1})}\big[z_{t-1}^\top \theta s^\top m_\Lambda S z_{t-1}\big] && \text{(20c)}\\
&= \mathbb{E}_{q(z_{t-1})} \mathrm{tr}\big(\theta s^\top m_\Lambda S z_{t-1} z_{t-1}^\top\big) && \text{(20d)}\\
&= \mathrm{tr}\Big(\theta s^\top m_\Lambda S \big(m_{z_{t-1}} m_{z_{t-1}}^\top + V_{z_{t-1}}\big)\Big) && \text{(20e)}\\
(B) &= \mathbb{E}_{q(z_{t-1})}\big[z_{t-1}^\top \theta s^\top m_\Lambda s\theta^\top z_{t-1}\big] && \text{(20f)}\\
&= \mathbb{E}_{q(z_{t-1})} \mathrm{tr}\big(\theta s^\top m_\Lambda s\theta^\top z_{t-1} z_{t-1}^\top\big) && \text{(20g)}\\
&= \mathrm{tr}\Big(\theta s^\top m_\Lambda s\theta^\top \big(m_{z_{t-1}} m_{z_{t-1}}^\top + V_{z_{t-1}}\big)\Big) && \text{(20h)}\\
(C) &= \mathbb{E}_{q(z_{t-1})}\big[z_{t-1}^\top S^\top m_\Lambda s\theta^\top z_{t-1}\big] && \text{(20i)}\\
&= \mathbb{E}_{q(z_{t-1})} \mathrm{tr}\big(S^\top m_\Lambda s\theta^\top z_{t-1} z_{t-1}^\top\big) && \text{(20j)}\\
&= \mathrm{tr}\Big(S^\top m_\Lambda s\theta^\top \big(m_{z_{t-1}} m_{z_{t-1}}^\top + V_{z_{t-1}}\big)\Big)\,. && \text{(20k)}
\end{aligned}
$$

Collecting terms (1), (3) and (2A) produces:

Message to η

Message to τ

References

[1] Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.