Elias Olofsson 2021-05-18

Task 2: M-step in a linear Gaussian state space model

Consider a linear Gaussian state space model,

$$z_t = A z_{t-1} + \varepsilon \quad , \qquad \varepsilon \sim N(0, \Sigma)$$

$$x_t = B z_t + \eta \quad , \qquad \eta \sim N(0, \Gamma) .$$

Derive the M-step for the latent state noise Σ by maximizing the expected log-likelihood $\mathbb{E}[\log p(X, Z)]$ with respect to Σ, where $X = \{x_t \mid t \in 1, \dots, T\}$ and $Z = \{z_t \mid t \in 1, \dots, T\}$ are the sets of all latent states and observation from time 1 to T.

Solution:

$$\Sigma^* = \underset{\Sigma}{\arg\max} \; \mathbb{E}_q\left[\log p(X, Z)\right]$$

$$\left[\begin{array}{l}
\text{Bayes' rule, Conditional indep. \& } 1^{st} \text{ order Markov property:} \\
p(X,Z) = p(X \mid Z)\, p(Z) \\
\qquad\quad\;\; = p(z_1)\, p(x_1 \mid z_1) \prod_{t=2}^{T} p(x_t \mid z_t)\, p(z_t \mid z_{t-1})
\end{array}\right]$$

$$= \underset{\Sigma}{\arg\max} \left(\mathbb{E}_q\left[\log p(z_1)\right] + \mathbb{E}_q\left[\sum_{t=2}^{T} \log p(z_t \mid z_{t-1})\right] + \mathbb{E}_q\left[\sum_{t=1}^{T} \log p(x_t \mid z_t)\right] \right)$$

$$\left[\begin{array}{l}
\text{Probability assumptions for a linear SSM:} \\
z_1 \sim N(\mu_0, \Sigma) \;, \quad z_t \sim N(A z_{t-1}, \Sigma) \;, \quad x_t \sim N(B z_{t-1}, \Gamma)
\end{array}\right]$$

$$= \underset{\Sigma}{\arg\max} \; \mathbb{E}_q\left[\cancel{-\frac{M}{2}\log(2\pi)} - \frac{1}{2}\log\det(\Sigma) - \frac{1}{2}(z_1-\mu_0)^T \Sigma^{-1} (z_1 - \mu_0)\right]$$

$$+ \mathbb{E}_q\left[\cancel{-\frac{M(T-1)}{2}\log(2\pi)} - \frac{(T-1)}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{t=2}^{T}(z_t - A z_{t-1})^T \Sigma^{-1} (z_t - A z_{t-1})\right]$$

$$\cancel{+ \mathbb{E}_q\left[-\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log\det(\Gamma) - \frac{1}{2}\sum_{t=1}^{T}(x_t - B z_t)^T \Gamma^{-1} (x_t - B z_t)\right]}$$

(drop terms indep. of Σ, i.e. no influence on argmax.)

$$= \underset{\Sigma}{\operatorname{argmax}} \quad \mathbb{E}_q\left[-\frac{1}{2}\log\det(\Sigma) - \frac{1}{2}(z_1-\mu_0)^T\Sigma^{-1}(z_1-\mu_0)\right]$$

$$+ \mathbb{E}_q\left[-\frac{(T-1)}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{t=2}^{T}(z_t - Az_{t-1})^T\Sigma^{-1}(z_t - Az_{t-1})\right]$$

$$= \underset{\Sigma}{\operatorname{argmax}} \quad \mathbb{E}_q\left[-\frac{T}{2}\log\det(\Sigma)\right] + \mathbb{E}_q\left[-\frac{1}{2}(z_1-\mu_0)^T\Sigma^{-1}(z_1-\mu_0)\right]$$

$$+ \mathbb{E}_q\left[-\frac{1}{2}\sum_{t=2}^{T}(z_t - Az_{t-1})^T\Sigma^{-1}(z_t - Az_{t-1})\right]$$

(Since the expectation value is evaluated over random variables x & z, we can eliminate the $\mathbb{E}_q[\cdots]$ of the first term :)

$$= \underset{\Sigma}{\operatorname{argmax}} \quad -\frac{T}{2}\log\det(\Sigma) - \frac{1}{2}\mathbb{E}_q\left[(z_1-\mu_0)^T\Sigma^{-1}(z_1-\mu_0)\right]$$

$$-\frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_q\left[(z_t - Az_{t-1})^T\Sigma^{-1}(z_t - Az_{t-1})\right]$$

$$= \underset{\Sigma}{\operatorname{argmax}} \quad -\frac{T}{2}\log\det(\Sigma)$$

$$-\frac{1}{2}\left(\mathbb{E}[z_1^T\Sigma^{-1}z_1] - \mathbb{E}[z_1^T\Sigma^{-1}\mu_0] - \mathbb{E}[\mu_0^T\Sigma^{-1}z_1] + \mathbb{E}[\mu_0^T\Sigma^{-1}\mu_0]\right)$$

$$-\frac{1}{2}\sum_{t=2}^{T}\left(\mathbb{E}[z_t^T\Sigma^{-1}z_t] - \mathbb{E}[z_t^T\Sigma^{-1}Az_{t-1}] - \mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}z_t] + \mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}A^Tz_{t-1}]\right)$$

[We need to extract the Σ^{-1}:s from the expectation values. Use the matrix identity :

$$x^TAy = Tr[Ayx^T] \qquad \text{where } x, y \text{ column vector, } A \text{ matrix}$$

$$\Rightarrow \quad \mathbb{E}[z_t^T\Sigma^{-1}z_t] = Tr[\Sigma^{-1}\mathbb{E}[z_tz_t^T]]$$

$$\mathbb{E}[z_t^T\Sigma^{-1}Az_{t-1}] = Tr[\Sigma^{-1}A\,\mathbb{E}[z_tz_{t-1}^T]]$$

$$\mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}z_t] = Tr[A^T\Sigma^{-1}\mathbb{E}[z_{t-1}z_t^T]]$$

$$\mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}Az_{t-1}] = Tr[A^T\Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]]$$

]

etc.

We start with the second term of the argmax – expression :

$$-\frac{1}{2}\left(\mathbb{E}[z_1^T\Sigma^{-1}z_1] - \mathbb{E}[z_1^T\Sigma^{-1}\mu_0] - \mathbb{E}[\mu_0^T\Sigma^{-1}z_1] + \mathbb{E}[\mu_0^T\Sigma^{-1}\mu_0]\right) =$$

$$= -\frac{1}{2}\left(\mathbb{E}[z_1^T\Sigma^{-1}z_1] - \underbrace{\mathbb{E}[z_1^T]}_{=\mu_0^T}\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}\underbrace{\mathbb{E}[z_1]}_{=\mu_0} + \mu_0^T\Sigma^{-1}\mu_0\right)$$

$$= -\frac{1}{2}\left(\mathbb{E}[z_1^T\Sigma^{-1}z_1] - \mu_0^T\Sigma^{-1}\mu_0\right)$$

$$= -\frac{1}{2}\left(\mathrm{Tr}[\Sigma^{-1}\mathbb{E}[z_1z_1^T]] - \mathrm{Tr}[\Sigma^{-1}\mu_0\mu_0^T]\right)$$

$$= -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}\mathbb{E}[z_1z_1^T] - \Sigma^{-1}\mu_0\mu_0^T\right]$$

$$= -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}\left(\mathbb{E}[z_1z_1^T] - \mu_0\mu_0^T\right)\right]$$

$$= -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}\,\mathrm{Var}(z_1)\right]$$

$$= -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}\Sigma\right]$$

$$= -\frac{1}{2}\,\mathrm{Tr}[\mathbb{1}] = -\frac{M}{2} \qquad \text{const. w.r.t } \Sigma$$

$\Longrightarrow$ Can drop the 2^{nd} term.

Taking the third term of the argmax – expression :

$$-\frac{1}{2}\sum_{t=2}^{T}\left(\mathbb{E}[z_t^T\Sigma^{-1}z_t] - \mathbb{E}[z_t^T\Sigma^{-1}Az_{t-1}] - \mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}z_t] + \mathbb{E}[z_{t-1}^TA^T\Sigma^{-1}Az_{t-1}]\right)$$

$$= -\frac{1}{2}\sum_{t=2}^{T}\Big(\mathrm{Tr}[\Sigma^{-1}\mathbb{E}[z_tz_t^T]] - \mathrm{Tr}[\Sigma^{-1}A\,\mathbb{E}[z_tz_{t-1}^T]]$$
$$- \mathrm{Tr}[A^T\Sigma^{-1}\mathbb{E}[z_{t-1}z_t^T]] + \mathrm{Tr}[A^T\Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]]\Big)$$

$$\left[\ \mathbb{E}[z_{t-1}z_t^T] = \mathbb{E}[(z_tz_{t-1}^T)^T] = \mathbb{E}[z_tz_{t-1}^T]^T\ \right]$$

$$= -\frac{1}{2}\sum_{t=2}^{T}\Big(\mathrm{Tr}[\Sigma^{-1}\mathbb{E}[z_tz_t]] - \mathrm{Tr}[\Sigma^{-1}A\,\mathbb{E}[z_tz_{t-1}^T]]$$
$$- \mathrm{Tr}[A^T\Sigma^{-1}\mathbb{E}[z_tz_{t-1}^T]^T + \mathrm{Tr}[A^T\Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]]\Big)$$

Collecting all the terms :

$$\Sigma^* = \underset{\Sigma}{\arg\max} \; -\frac{T}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{t=2}^{T}\Big(Tr[\Sigma^{-1}\mathbb{E}[z_t z_t]] - Tr[\Sigma^{-1}A\,\mathbb{E}[z_t z_{t-1}^T]]$$

$$- Tr[A^T\Sigma^{-1}\mathbb{E}[z_t z_{t-1}^T]^T] + Tr[A^T\Sigma^{-1}A^T\,\mathbb{E}[z_{t-1}z_{t-1}^T]]\Big)$$

Finding the maximum w.r.t Σ :

$$\frac{\partial\, ELBO}{\partial \Sigma} = \frac{\partial}{\partial\Sigma}(\star) \overset{!}{=} 0$$

Differentiating term by term :

$$\frac{\partial}{\partial\Sigma}\Big[-\frac{T}{2}\log\det(\Sigma)\Big] = -\frac{T}{2}\frac{\partial}{\partial\Sigma}\log\det(\Sigma) =$$

$$= -\frac{T}{2}\frac{1}{\det(\Sigma)}\frac{\partial\det(\Sigma)}{\partial\Sigma} = -\frac{T}{2}\frac{1}{\det(\Sigma)}\det(\Sigma)\,\Sigma^{-1}$$

$$= -\frac{T}{2}\Sigma^{-1}$$

From the Matrix cook-book :

$$\Big[\frac{\partial\, Tr[AX^{-1}B]}{\partial X} = -X^{-1}BAX^{-1}\Big]$$

$$\frac{\partial}{\partial\Sigma}\Big[Tr[\Sigma^{-1}\mathbb{E}[z_t z_t^T]]\Big] = -\Sigma^{-1}\mathbb{E}[z_t z_t^T]\,\Sigma^{-1}$$

$$\frac{\partial}{\partial\Sigma}\Big[-Tr[\Sigma^{-1}A\,\mathbb{E}[z_t z_{t-1}^T]]\Big] = \Sigma^{-1}A\,\mathbb{E}[z_t z_{t-1}^T]\,\Sigma^{-1}$$

$$\frac{\partial}{\partial\Sigma}\Big[-Tr[A^T\Sigma^{-1}\mathbb{E}[z_t z_{t-1}^T]^T]\Big] = \Sigma^{-1}\mathbb{E}[z_t z_{t-1}^T]^T A^T\,\Sigma^{-1}$$

$$\frac{\partial}{\partial\Sigma}\Big[Tr[A^T\Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]]\Big] = -\Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]\,A^T\,\Sigma^{-1}$$

Putting it all together :

$$0 = -\frac{T}{2}\Sigma^{-1} - \frac{1}{2}\sum_{t=2}^{T}\Big(-\Sigma^{-1}\mathbb{E}[z_t z_t^T] + \Sigma^{-1}A\,\mathbb{E}[z_t z_{t-1}^T]\,\Sigma^{-1}$$

$$+\Sigma^{-1}\mathbb{E}[z_t z_{t-1}^T]^T A^T\,\Sigma^{-1} - \Sigma^{-1}A\,\mathbb{E}[z_{t-1}z_{t-1}^T]\,A^T\,\Sigma^{-1}\Big)$$

Left- and right-multiply with Σ, and multiply by 2:

$$0 = -T\Sigma - \sum_{t=2}^{T}\Big(-\mathbb{E}[z_t z_t^T] + A\,\mathbb{E}[z_t z_{t-1}^T] + \mathbb{E}[z_t z_{t-1}^T]^T A^T$$
$$- A\,\mathbb{E}[z_{t-1} z_{t-1}^T] A^T \Big)$$

$$= -T\Sigma + \sum_{t=2}^{T}\Big(\mathbb{E}[z_t z_t^T] - A\,\mathbb{E}[z_t z_{t-1}^T] - \big(A\,\mathbb{E}[z_t z_{t-1}^T]\big)^T$$
$$+ A\,\mathbb{E}[z_{t-1} z_{t-1}^T] A^T \Big)$$

$$\Rightarrow \quad \Sigma^* = \frac{1}{T}\Big[\sum_{t=2}^{T}\mathbb{E}[z_t z_t^T] - A\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}^T] - \Big(A\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}^T]\Big)^T$$
$$+ A\sum_{t=2}^{T}\mathbb{E}[z_{t-1} z_{t-1}^T] A^T \Big]$$

[Using the previously derived result for A:

$$A = \Big(\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}^T]\Big)\Big(\sum_{t=2}^{T}\mathbb{E}[z_{t-1} z_{t-1}^T]\Big)^{-1}$$]

$$= \frac{1}{T}\Big[\sum_{t=2}^{T}\mathbb{E}[z_t z_t^T] - \Big(\sum_{t=2}\mathbb{E}[z_t z_{t-1}^T]\Big)\Big(\sum_{t=2}\mathbb{E}[z_{t-1} z_{t-1}^T]\Big)^{-1}\Big(\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}^T]\Big)$$

$-$

....

[I believe that I must have mixed up a transpose or similar, since I can't get terms to cancel in a meaningful way.

I suspect that the final answer should be similar to

$$\Sigma^* = \frac{1}{T}\Big(\sum_{t=2}\mathbb{E}[z_t z_t^T] - \Big(\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}^T]\Big)\Big(\sum_{t=2}^{T}\mathbb{E}[z_{t-1} z_{t-1}^T]\Big)^{-1}\Big(\sum_{t=2}^{T}\mathbb{E}[z_t z_{t-1}]\Big)^T\Big)$$

<u>But</u>, since I have not arrived there, this is <u>not</u> a final answer to the question.]