Supplementary derivations for

"Variational message passing for online polynomial NARMAX identification"

Wouter Kouw

September 24, 2021

PNARMAX

Let y_k be an output, u_k be an input and e_k be an error at time k. Consider the following set of dynamics:

$$y_k = \theta^\top \phi\left(u_k, u_{k-1}, \dots, u_{k-M_1}, y_{k-1}, \dots, y_{k-M_2}, e_{k-1}, \dots, e_{k-M_3}\right) + e_k \tag{1}$$

where $e_k \sim \mathcal{N}(0, \tau^{-1})$ is a zero mean zero auto-correlation Gaussian noise component. The function f, parameterized by θ, is a nonlinear regression from input, output and noise components unto the current output. The constants M_1, M_2, and M_3 refer to the delays, for a total model order of $M = M_1 + 1 + M_2 + M_3$.

We will use the following shorthands:

$$\begin{aligned}
\mathbf{u}_{k-1} &\triangleq (u_{k-1}, \dots, u_{k-M_1}) && \text{(2a)}\\
\mathbf{y}_{k-1} &\triangleq (y_{k-1}, \dots, u_{k-M_2}) && \text{(2b)}\\
\mathbf{e}_{k-1} &\triangleq (e_{k-1}, \dots, e_{k-M_3}) && \text{(2c)}\\
\phi_k &\triangleq \phi\left(\mathbf{u}_{k-1}, \mathbf{y}_{k-1}, \mathbf{e}_{k-1}\right) && \text{(2d)}\\
\Phi_k &\triangleq \phi_k \phi_k^\top \, . && \text{(2e)}
\end{aligned}$$

Generative model

We cast the polynomial NARMAX dynamics to a Gaussian likelihood function:

$$p(y_k \mid \theta, u_k, \mathbf{u}_{k-1}, \mathbf{y}_{k-1}, \mathbf{e}_{k-1}, \tau) = \mathcal{N}\Big(y_k \mid \theta^\top \phi_k, \ \tau^{-1}\Big) \, . \tag{3}$$

Using the following priors

$$p(\theta) \triangleq \mathcal{N}\big(\theta \mid \mu_0, \Lambda_0^{-1}\big) \, , \qquad p(\tau) \triangleq \Gamma\big(\tau \mid \alpha_0, \beta_0\big) \, , \tag{4}$$

we form a generative model for the signal up to time T:

$$p(y_{1:T}, \theta, \tau \mid u_{1:T}) = \underbrace{p(\theta)p(\tau)}_{\text{priors}} \prod_{k=1}^{T} \underbrace{p(y_k \mid \theta, u_k, \mathbf{u}_{k-1}, \mathbf{y}_{k-1}, \mathbf{e}_{k-1})}_{\text{likelihood}} . \tag{5}$$

Recognition model

We employ a mean-field factorisation: $q_k(\theta, \tau) = q_k(\theta)q_k(\tau)$ where

$$q_k(\theta) \triangleq \mathcal{N}(\theta \mid \mu_k, \Lambda_k^{-1}) , \qquad q_k(\tau) \triangleq \mathit{\Gamma}(\tau \mid \alpha_k, \beta_k) . \tag{6}$$

Note that we employ mean-covariance and shape-rate parameterisations.

Message derivations

The PNARMAX factor node sends out variational messages to its coefficients θ and its precision parameter τ. The general mathematical formula for a variational message in a factor graph is [1]:

$$\nu(x_i) \propto \exp\left(\mathbb{E}_{q(x_{j\neq i})}\big[\log p(x_i, \dots)\big]\right) , \tag{7}$$

where $p(x_i, \dots)$ represents the factor node function, which in our case is the likelihood (Equation 3).

Message to θ

In the following, I ignore terms that do not depend on θ (denoted with C).

$$\log \nu(\theta) = \mathbb{E}_{q(\tau)} \log \mathcal{N}\Big(y_k \mid \theta^\top \phi_k, \tau^{-1}\Big) + \mathsf{C} \tag{8a}$$

$$= -\frac{1}{2}\mathbb{E}_{q(\tau)}\big[\tau\big]\Big(y_k - \theta^\top \phi_k\Big)^2 + \mathsf{C} \tag{8b}$$

$$= -\frac{1}{2}\frac{\alpha_k}{\beta_k}\Big(y_k^2 - 2y_k\theta^\top \phi_k + \theta^\top \phi_k \phi_k^\top \theta\Big) + \mathsf{C} \tag{8c}$$

$$= -\frac{1}{2}\Big(- 2\theta^\top \underbrace{\frac{\alpha_k}{\beta_k} y_k \phi_k}_{Wm} + \theta^\top \underbrace{(\frac{\alpha_k}{\beta_k}\Phi_k)}_{W}\theta\Big) + \mathsf{C} . \tag{8d}$$

We recognise both a linear function, $\theta^\top (Wm)$, and a quadratic function, $\theta^\top W \theta$. Consider for a moment a multivariate Gaussian probability density function, $\mathcal{N}(x \mid m, W^{-1})$, in the log-domain and ignore the normalisation terms as well as all terms in the exponent that don't depend on x:

$$\log \mathcal{N}(x \mid m, W^{-1}) \propto -\frac{1}{2}\Big(- 2x^\top W m + x^\top W x\Big) . \tag{9}$$

With this, we recognise a Gaussian distribution in Equation 8d. If we left-multiply Wm with the inverse precision W^{-1}, we obtain the mean; $m = W^{-1}Wm$. This yields:

$$\nu(\theta) \propto \mathcal{N}\Big(\theta \mid \big(\frac{\alpha_k}{\beta_k}\Phi_k\big)^{-1}\big(\frac{\alpha_k}{\beta_k}y_k\phi_k\big), \big(\frac{\alpha_k}{\beta_k}\Phi_k\big)^{-1}\Big)\,. \tag{10}$$

Message to τ

Terms that do not depend on τ are moved to the constant term C.

$$\begin{aligned}
\log \nu(\tau) &= \mathbb{E}_{q(\theta)} \log \mathcal{N}\Big(y_k \mid \theta^\top \phi_k, \tau^{-1}\Big) + \mathsf{C} && \text{(11a)}\\
&= \frac{1}{2}\log\tau - \frac{\tau}{2}\mathbb{E}_{q(\theta)}\Big[y_k^2 - 2y_k\theta^\top\phi_k + \theta^\top\phi_k\phi_k^\top\theta\Big] + \mathsf{C} && \text{(11b)}\\
&= \frac{1}{2}\log\tau - \tau\frac{1}{2}\Big(y_k^2 - 2y_k\mathbb{E}_{q(\theta)}[\theta^\top]\phi_k + \phi_k^\top\mathbb{E}_{q(\theta)}[\theta\theta^\top]\phi_k\Big) + \mathsf{C} && \text{(11c)}\\
&= \underbrace{\frac{1}{2}}_{a-1}\log\tau - \tau\underbrace{\frac{1}{2}\Big(y_k^2 - 2y_k\mu_k^\top\phi_k + \phi_k^\top(\mu_k\mu_k^\top + \Lambda_k^{-1})\phi_k\Big)}_{b} + \mathsf{C}\,. && \text{(11d)}
\end{aligned}$$

Consider the Gamma probability density function, with a shape-rate parameterisation, in the log-domain:

$$\log\Gamma(x \mid a, b) = (a-1)\log x - xb + \mathsf{C}\,. \tag{12}$$

We recognise a log-term and a linear term in Equation 11d. We can therefore say that the variational message towards τ is proportional to a Gamma distribution:

$$\nu(\tau) \propto \Gamma\Big(\tau \mid \frac{3}{2}, \frac{1}{2}\Big(y_k^2 - 2y_k\mu_k^\top\phi_k + \phi_k^\top(\mu_k\mu_k^\top + \Lambda_k^{-1})\phi_k\Big)\Big)\,. \tag{13}$$

References

[1] Justin Dauwels. On variational message passing on factor graphs. In *IEEE International Symposium on Information Theory*, pages 2546–2550, 2007.