Space-Varying Covariance Matrices

C. Barata[(a)*]   J. M. Lemos[(b)]   J. S. Marques[(a)]
[(a)]ISR, Instituto Superior Técnico, Lisboa, Portugal
[(b)] INESC-ID, Instituto Superior Técnico, Lisboa, Portugal

January 2018

This report describes the estimate of the covariance matrices $\mathbf{\Sigma}_k$, $k = 1, ..., K$, associated to each of the velocity fields T_k. Two formulations will be used:

1. Each field has a global matrix $\mathbf{\Sigma}_k \in \mathbb{R}^{2\times 2}$.
2. A matrix $\mathbf{\Sigma}_k^n \in \mathbb{R}^{2\times 2}$ is defined for each node n of a velocity field k.

This report uses the same notation as that of the ARGUS report from December 23, 2010. All of the references to equations are with respect to the aforementioned report, unless stated otherwise.

1 Global $\mathbf{\Sigma}_k$ for each field

Differentiating $U(\theta, \hat{\theta})$ (1.32) with respect to $\mathbf{\Sigma}_\alpha$.

$$\frac{\partial U(\theta,\hat{\theta})}{\partial \mathbf{\Sigma}_\alpha} = \frac{\partial}{\partial \mathbf{\Sigma}_\alpha} E[\log p(\mathcal{X},\mathcal{K}|\theta)|\mathcal{X},\hat{\theta}] + \frac{\partial}{\partial \mathbf{\Sigma}_\alpha} \log p(\theta)\,. \tag{1}$$

Without considering the prior $p(\theta)$:

$$\frac{\partial}{\partial \mathbf{\Sigma}_\alpha} E[\log p(\mathcal{X},\mathcal{K}|\theta)|\mathcal{X},\hat{\theta}] = -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha} \sum_{s=1}^{S}\sum_{t=1}^{L_s}\sum_{k=1}^{K} w_k^s(t)[\log(|\mathbf{\Sigma}_k|) + \|v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s)\|^2_{\mathbf{\Sigma}_k}], \tag{2}$$

*This work was supported by the FCT project [PTDC/EEIPRO/0426/2014] and supported by FCT plurianual funding [UID/EEA/50009/2019].

where $v_t^s = x_t^s - x_{t-1}^s$. Using the **trace trick** we get

$$\begin{aligned}
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha}\sum_{s=1}^{S}\sum_{t=1}^{L_s}\sum_{k=1}^{K} w_k^s(t)[\log(|\mathbf{\Sigma}_k|) + trace(\mathbf{\Sigma}_k^{-1}(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T) \\
&= -\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\frac{\partial}{\partial \mathbf{\Sigma}_\alpha}\left[\log(|\mathbf{\Sigma}_\alpha|) + trace(\mathbf{\Sigma}_\alpha^{-1}(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T)\right] \\
&= -\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\left[\mathbf{\Sigma}_\alpha^{-1} - (\mathbf{\Sigma}_\alpha^{-1}\mathbf{G}_t^s\mathbf{\Sigma}_\alpha^{-1})^T\right]
\end{aligned}$$

, (3)

where

$$\mathbf{G}_t^s = (v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T. \tag{4}$$

The derivatives were obtained using the following properties

$$\frac{\partial}{\partial \mathbf{H}}\log(|\mathbf{H}|) = (\mathbf{H}^{-1})^T, \tag{5}$$

$$\frac{\partial}{\partial \mathbf{H}} trace(\mathbf{A}\mathbf{H}^{-1}\mathbf{B}) = -(\mathbf{H}^{-1}\mathbf{B}\mathbf{A}\mathbf{H}^{-1})^T. \tag{6}$$

In our case, $\mathbf{H} = \mathbf{\Sigma}_\alpha$, $\mathbf{A} = \mathbf{I}$, and $\mathbf{B} = \mathbf{G}$.

In order to maximize $E[\log p(\mathcal{X}, \mathcal{K}|\theta)|\mathcal{X}, \hat{\theta}]$ with respect to $\mathbf{\Sigma}_\alpha$ we compute

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{\Sigma}_\alpha} E[\log p(\mathcal{X}, \mathcal{K}|\theta)|\mathcal{X}, \hat{\theta}] &= 0 \\
-\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\left[\mathbf{\Sigma}_\alpha^{-1} - (\mathbf{\Sigma}_\alpha^{-1}\mathbf{G}_t^s\mathbf{\Sigma}_\alpha^{-1})^T\right] &= 0 \\
\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\left[\mathbf{\Sigma}_\alpha^{-1} - \mathbf{\Sigma}_\alpha^{-1}(\mathbf{G}_t^s)^T\mathbf{\Sigma}_\alpha^{-1}\right] &= 0 \\
\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\left[\mathbf{\Sigma}_\alpha\mathbf{\Sigma}_\alpha^{-1}\mathbf{\Sigma}_\alpha - \mathbf{\Sigma}_\alpha\mathbf{\Sigma}_\alpha^{-1}(\mathbf{G}_t^s)^T\mathbf{\Sigma}_\alpha^{-1}\mathbf{\Sigma}_\alpha\right] &= 0 \\
\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\left[\mathbf{I}\mathbf{\Sigma}_\alpha - \mathbf{I}(\mathbf{G}_t^s)^T\mathbf{I}\right] &= 0
\end{aligned}$$

$$\mathbf{\Sigma}_\alpha = \frac{\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)(\mathbf{G}_t^s)^T}{\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)} \tag{7}$$

2 Node-specific $\mathbf{\Sigma}_k^n$ for each field

In this formulation, each velocity field k has N matrices $\mathbf{\Sigma}_k^n$, defined at the nodes $n \in \{1, 2, ..., N\}$. For each field, the covariance matrix at position x_{t-1} is interpolated from the N matrices, similarly to (1.12)

$$\mathbf{\Sigma}_k(x_{t-1}) = \sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1}). \tag{8}$$

Since $0 \leq \phi^n(x) \leq 1$, the interpolated matrix is still semidefinite positive.

Using this new formulation, we now want to compute $\frac{\partial U(\theta,\hat{\theta})}{\partial \mathbf{\Sigma}_\alpha^\gamma}$. Without considering a prior, this leads to

$$\begin{aligned}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma} E[\log p(\mathcal{X},\mathcal{K}|\theta)|\mathcal{X},\hat{\theta}] &= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma}\sum_{s=1}^{S}\sum_{t=1}^{L_s}\sum_{k=1}^{K} w_k^s(t)[\log(|\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s)|) + \\ &+ (v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T(\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s))^{-1}(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))], \end{aligned} \tag{9}$$

Using the **trace trick** we get

$$\begin{aligned} &= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma}\sum_{s=1}^{S}\sum_{t=1}^{L_s}\sum_{k=1}^{K} w_k^s(t)[\log(|\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s)|) \\ &+ trace((\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s))^{-1}(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T] \\ &= -\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma}\log(|\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s)|) \\ &+ w_\alpha^s(t)\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma} trace\left(\left(\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1}^s)\right)^{-1}(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))(v_t^s - \mathbf{T_k}\Phi(x_{t-1}^s))^T\right), \end{aligned} \tag{10}$$

To obtain the derivative it is necessary to use the **chain rule**

$$\frac{\partial f(\mathbf{P}(\mathbf{H}))}{\partial \mathbf{H}} = \frac{\partial f}{\partial \mathbf{P}}.\frac{\partial \mathbf{P}}{\partial \mathbf{H}}. \tag{11}$$

In our case $\mathbf{H} = \mathbf{\Sigma}_\alpha^\gamma$, $\mathbf{P}(\mathbf{H}) = \sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1})$, and f represents the $\log(|.|)$ and $trace(.)$ expressions. Looking separately to each of the terms we get

$$\begin{aligned}\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma}\log(|\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1})|) &= \frac{\partial \log(|\mathbf{P}|)}{\partial \mathbf{P}}.\frac{\partial \mathbf{P}}{\partial \mathbf{\Sigma}_\alpha^\gamma} \\ &= \mathbf{P}^{-1}\phi^\gamma(x_{t-1})\mathbf{I} \\ &= \phi^\gamma(x_{t-1})\left(\sum_{n=1}^{N}\mathbf{\Sigma}_k^n\phi^n(x_{t-1})\right)^{-1} \end{aligned} \tag{12}$$

and

$$\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma} trace\left(\left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1}^s)\right)^{-1} \mathbf{G}_t^s\right) = \frac{\partial trace(\mathbf{P}^{-1}\mathbf{G})}{\partial \mathbf{P}} \cdot \frac{\partial \mathbf{P}}{\partial \mathbf{\Sigma}_\alpha^\gamma}$$

$$= -(\mathbf{P}^{-1}\mathbf{G}\mathbf{I}\mathbf{P}^{-1})^T \phi^\gamma(x_{t-1})\mathbf{I}$$

$$= -\phi^\gamma(x_{t-1})\left(\left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1})\right)^{-1} \mathbf{G}\left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1})\right)^{-1}\right)^T , \quad (13)$$

where $\mathbf{G}$ is defined as in (6) of this report.

Combining the two terms, we get

$$\frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma} E[\log p(\mathcal{X}, \mathcal{K}|\theta)|\mathcal{X}, \hat{\theta}] = -\frac{1}{2}\sum_{s=1}^{S}\sum_{t=1}^{L_s} w_\alpha^s(t)\phi^\gamma(x_{t-1}^s)[\left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1}^s)\right)^{-1} -$$

$$- \left(\left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1}^s)\right)^{-1} \mathbf{G}_t^s \left(\sum_{n=1}^{N} \mathbf{\Sigma}_k^n \phi^n(x_{t-1}^s)\right)^{-1}\right)^T] \quad (14)$$

Setting $\frac{\partial U(\theta,\hat{\theta})}{\partial \mathbf{\Sigma}_\alpha^\gamma} = 0$ does not have an explicit solution. Thus, it is necessary to apply an iterative algorithm to solve it. Two options can be:

1. **Gradient method** - however it requires the projection of the estimated matrix in to a space of semidefinite positive matrices, in order to guarantee this property.

2. **Newton Raphson** - two methodologies are available in this case. The one that uses the first derivative, and the one the that uses the first and second derivatives. For the latter, it will be necessary to determine the expression of the second derivative.

3 Comment on the prior

The estimation of the covariance matrix for each of the formulations was performed without taking into account a prior. Although such formulation is possible, it would also be interesting to add a prior in the estimation. A relevant prior could be one that expressed the uncertainty of each node, *e.g*, setting diagonal covariance matrices with high values in nodes that are not supported by any observations.