

Supplementary Material

Sparse Convolved Gaussian Processes for Multi-output Regression

A Convolution with Gaussian kernel functions

For the covariance matrix of the latent functions we employ

$$k_{u_r u_r}(\mathbf{x}, \mathbf{x}') = \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top \mathbf{L}_r (\mathbf{x} - \mathbf{x}') \right]$$

and for the smoothing kernel

$$k_{qr}(\boldsymbol{\tau}) = \frac{S_{qr} |\mathbf{L}_{qr}|^{1/2}}{(2\pi)^{p/2}} \exp \left[-\frac{1}{2} \boldsymbol{\tau}^\top \mathbf{L}_{qr} \boldsymbol{\tau} \right].$$

Applying successively the result for the multiplication of Gaussian distributions [3], the covariance functions in expressions (3) and (4) are also Gaussian covariances given by

$$\begin{aligned} \text{cov} [f_q(\mathbf{x}), f_s(\mathbf{x}')] &= \sum_{r=1}^R \frac{S_{qr} S_{sr} |\mathbf{L}_r^{-1}|^{1/2}}{|\mathbf{L}_{qr}^{-1} + \mathbf{L}_{sr}^{-1} + \mathbf{L}_r^{-1}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top (\mathbf{L}_{qr}^{-1} + \mathbf{L}_{sr}^{-1} + \mathbf{L}_r^{-1})^{-1} (\mathbf{x} - \mathbf{x}') \right] \\ \text{cov} [f_q(\mathbf{x}), u_r(\mathbf{z})] &= \frac{S_{qr} |\mathbf{L}_r^{-1}|^{1/2}}{|\mathbf{L}_{qr}^{-1} + \mathbf{L}_r^{-1}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{z})^\top (\mathbf{L}_{qr}^{-1} + \mathbf{L}_r^{-1})^{-1} (\mathbf{x} - \mathbf{z}) \right] \end{aligned}$$

B Matrix Derivatives

We follow the notation of [3] obtaining similar results to [7]. This notation allows us to apply the chain rule for matrix derivation in a straight-forward manner. Let's define $\mathbf{G} \doteq \text{vec } \mathbf{G}$, where vec is the vectorization operator over the matrix \mathbf{G} . For a function \mathcal{L} the equivalence between $\frac{\partial \mathcal{L}}{\partial \mathbf{G}}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{G}^\top}$ is given through $\frac{\partial \mathcal{L}}{\partial \mathbf{G}^\top} = \left(\left(\frac{\partial \mathcal{L}}{\partial \mathbf{G}} \right)^\top \right)^\top$. The log-likelihood function is given as

$$\mathcal{L}(\mathbf{Z}, \boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{D} + \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \mathbf{K}_{u,f}| - \frac{1}{2} \text{tr} \left[(\mathbf{D} + \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \mathbf{K}_{u,f})^{-1} \mathbf{y} \mathbf{y}^\top \right] + \text{const} \quad (10)$$

where we have redefined \mathbf{D} as $\mathbf{D} = [\mathbf{K}_{f,f} - \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \mathbf{K}_{u,f}] \odot \mathbf{M} + \boldsymbol{\Sigma}$, to keep a simpler notation. Using the matrix inversion lemma and its equivalent form for determinants, expression (10) can be written as

$$\mathcal{L}(\mathbf{Z}, \boldsymbol{\theta}) \propto \frac{1}{2} \log |\mathbf{K}_{u,u}| - \frac{1}{2} \log |\mathbf{A}| - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \text{tr} [\mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top] + \frac{1}{2} \text{tr} [\mathbf{D}^{-1} \mathbf{K}_{f,u} \mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top].$$

We can find $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{Z}}$ applying the chain rule to \mathcal{L} obtaining expressions for $\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{f,f}}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{f,u}}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{u,u}}$ and combining those with the relevant derivatives of the covariances wrt $\boldsymbol{\theta}$ and \mathbf{Z} ,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{G}} = \frac{\partial \mathcal{L}_A}{\partial \mathbf{A}} \frac{\partial \mathbf{A}}{\partial \mathbf{D}} \frac{\partial \mathbf{D}}{\partial \mathbf{G}} + \frac{\partial \mathcal{L}_D}{\partial \mathbf{D}} \frac{\partial \mathbf{D}}{\partial \mathbf{G}} + \left[\frac{\partial \mathcal{L}_A}{\partial \mathbf{A}} \frac{\partial \mathbf{A}}{\partial \mathbf{G}} + \frac{\partial \mathcal{L}_G}{\partial \mathbf{G}} \right] \delta_{GK} \quad (11)$$

where the subindex in \mathcal{L}_E stands for those terms of \mathcal{L} which depend on \mathbf{E} , \mathbf{G} is either $\mathbf{K}_{f,f}$, $\mathbf{K}_{u,f}$ or $\mathbf{K}_{u,u}$ and δ_{GK} is zero if \mathbf{G} is equal to $\mathbf{K}_{f,f}$ and one in other case. Next we present expressions for each partial derivative

$$\begin{aligned} \frac{\partial \mathcal{L}_A}{\partial \mathbf{A}} &= -\frac{1}{2} (\mathbf{C})^\top, \quad \frac{\partial \mathbf{A}}{\partial \mathbf{D}} = -(\mathbf{K}_{u,f} \mathbf{D}^{-1} \otimes \mathbf{K}_{u,f} \mathbf{D}^{-1}), \quad \frac{\partial \mathcal{L}_D}{\partial \mathbf{D}} = -\frac{1}{2} ((\mathbf{D}^{-1} \mathbf{H} \mathbf{D}^{-1})^\top)^\top \\ \frac{\partial \mathbf{D}}{\partial \mathbf{K}_{f,f}} &= \text{diag}(\mathbf{M}), \quad \frac{\partial \mathbf{D}}{\partial \mathbf{K}_{u,f}} = -\text{diag}(\mathbf{M}) [(\mathbf{I} \otimes \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1}) + (\mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \otimes \mathbf{I}) \mathbf{T}_D], \\ \frac{\partial \mathbf{D}}{\partial \mathbf{K}_{u,u}} &= \text{diag}(\mathbf{M}) (\mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \otimes \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1}), \quad \frac{\partial \mathbf{A}}{\partial \mathbf{K}_{u,f}} = (\mathbf{K}_{u,f} \mathbf{D}^{-1} \otimes \mathbf{I}) + (\mathbf{I} \otimes \mathbf{K}_{u,f} \mathbf{D}^{-1}) \mathbf{T}_A \\ \frac{\partial \mathbf{A}}{\partial \mathbf{K}_{u,u}} &= \mathbf{I}, \quad \frac{\partial \mathcal{L}_{K_{u,f}}}{\partial \mathbf{K}_{u,f}} = \left((\mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top \mathbf{D}^{-1})^\top \right)^\top, \quad \frac{\partial \mathcal{L}_{K_{u,u}}}{\partial \mathbf{K}_{u,u}} = \frac{1}{2} ((\mathbf{K}_{u,u}^{-1})^\top)^\top, \end{aligned}$$

where $\mathbf{C} = \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top \mathbf{D}^{-1} \mathbf{K}_{f,u} \mathbf{A}^{-1}$, $\mathbf{H} = \mathbf{D} - \mathbf{y} \mathbf{y}^\top + \mathbf{K}_{f,u} \mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top + (\mathbf{K}_{f,u} \mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top)^\top$ and \mathbf{T}_D and \mathbf{T}_A are *vectorized transpose matrices* [3] and we have not included

their dimensions to keep the notation clearer. We can replace the above expressions in (11) to find the corresponding derivatives, so

$$\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{f,f}:} = \frac{1}{2} \left[((\mathbf{C}):)^\top (\mathbf{K}_{u,f} \mathbf{D}^{-1} \otimes \mathbf{K}_{u,f} \mathbf{D}^{-1}) - \frac{1}{2} ((\mathbf{D}^{-1} \mathbf{H} \mathbf{D}^{-1}):)^\top \right] \text{diag}(\mathbf{M}:) \quad (12a)$$

$$= -\frac{1}{2} ((\mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1}):)^\top \text{diag}(\mathbf{M}:) = -\frac{1}{2} (\text{diag}(\mathbf{M}:) (\mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1}):)^\top \quad (12b)$$

$$= -\frac{1}{2} ((\mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1} \odot \mathbf{M}):)^\top = -\frac{1}{2} (\mathbf{Q}:)^\top \quad (12c)$$

where in (12a) $\mathbf{J} = \mathbf{H} - \mathbf{K}_{f,u} \mathbf{C} \mathbf{K}_{u,f}$ and $\mathbf{Q} = (\mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1} \odot \mathbf{M})$. We have used the property $(\mathbf{B}:)^\top (\mathbf{F} \otimes \mathbf{P}) = ((\mathbf{P}^\top \mathbf{B} \mathbf{F}):)^\top$ in (12a) and the property $\text{diag}(\mathbf{B}:) \mathbf{F}: = (\mathbf{B} \odot \mathbf{F}):$, to go from (12b) to (12c). We also have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{K}_{u,f}:} &= \frac{1}{2} (\mathbf{Q}:)^\top [(\mathbf{I} \otimes \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1}) + (\mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1} \otimes \mathbf{I}) \mathbf{T}_D] - \frac{1}{2} (\mathbf{C}:)^\top \\ &\quad [(\mathbf{K}_{u,f} \mathbf{D}^{-1} \otimes \mathbf{I}) + (\mathbf{I} \otimes \mathbf{K}_{u,f} \mathbf{D}^{-1}) \mathbf{T}_A] + \left((\mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top \mathbf{D}^{-1}): \right)^\top \\ &= \left((\mathbf{K}_{u,u}^{-1} \mathbf{K}_{u,f} \mathbf{Q} - \mathbf{C} \mathbf{K}_{u,f} \mathbf{D}^{-1} + \mathbf{A}^{-1} \mathbf{K}_{u,f} \mathbf{D}^{-1} \mathbf{y} \mathbf{y}^\top \mathbf{D}^{-1}): \right)^\top \end{aligned} \quad (13)$$

where in (13), $(\mathbf{Q}:)^\top (\mathbf{F} \otimes \mathbf{I}) \mathbf{T}_D = (\mathbf{Q}:)^\top \mathbf{T}_D (\mathbf{I} \otimes \mathbf{F}) = (\mathbf{T}_D^\top \mathbf{Q}:)^\top (\mathbf{I} \otimes \mathbf{F}) = (\mathbf{Q}:)^\top (\mathbf{I} \otimes \mathbf{F})$. A similar analysis is formulated for the term involving \mathbf{T}_A . Finally, results for $\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{u,f}:}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{\Sigma}:}$ are obtained as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{u,u}:} = -\frac{1}{2} ((\mathbf{K}_{u,u}^{-1} - \mathbf{C} - \mathbf{K}_{u,u}^{-1} \mathbf{K}_{u,f} \mathbf{Q} \mathbf{K}_{f,u} \mathbf{K}_{u,u}^{-1}):)^\top, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{\Sigma}:} = -\frac{1}{2} (\mathbf{Q}:)^\top.$$