Supplementary material: A Mutually-Dependent Hadamard Kernel for Modelling Latent Variable Couplings

Sami Remes, Markus Heinonen, Samuel Kaski

A Optimisation of kernel parameters

The factor $q(\mathbf{Z})$ corresponding to the GP parameters of the proposed Hadamard kernel is updated by finding point estimates by maximizing the variational lower bound. The relevant part of the bound is given by

$$\mathcal{L}(\mathbf{Z}) = \sum_s \langle \log p(\mathbf{u}^{(s)}|\mathbf{Z})\rangle + \log p(\mathbf{Z}) = -\frac{1}{2}(S \log|\mathbf{K}| + \sum_s \langle \mathbf{u}^{(s)T}\mathbf{K}^{-1}\mathbf{u}^{(s)}\rangle + \mathrm{Tr}(\mathbf{Z}^T\mathbf{K}_z^{-1}\mathbf{Z}))$$

where $\mathbf{K} = \mathbf{Z}\mathbf{Z}^T \circ \mathbf{K}_Q + \mathbf{\Omega}$, and its gradient by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Z}_{ij}} = \frac{1}{2}\,\mathrm{Tr}\left(\left[\mathbf{K}^{-1}(\sum_s \langle \mathbf{u}^{(s)}\mathbf{u}^{(s)T}\rangle)\mathbf{K}^{-1} - S\mathbf{K}^{-1}\right]\frac{\partial \mathbf{K}}{\partial \mathbf{Z}_{ij}}\right) - \frac{1}{2}[\mathbf{K}_z^{-1}\mathbf{Z}]_{ij}$$

where $\mathbf{K}_z = K_z \otimes I_Q$ is a block matrix of full size $(QN \times QN)$, and

$$\frac{\partial \mathbf{K}}{\partial \mathbf{Z}_{ij}} = \frac{\partial(\mathbf{Z}\mathbf{Z}^T \circ \mathbf{K}_Q + \mathbf{\Omega})}{\partial \mathbf{Z}_{ij}} = (\mathbf{Z}\mathbf{1}_{ij}^T + \mathbf{1}_{ij}\mathbf{Z}^T) \circ \mathbf{K}_Q.$$

The cost function in the whitened domain can be evaluated as $\mathcal{L}(\mathbf{L}\hat{\mathbf{Z}})$ and gradient as

$$\frac{\partial \mathcal{L}}{\partial \hat{\mathbf{Z}}} = \frac{\partial L}{\partial \mathbf{Z}}\frac{\partial \mathbf{Z}}{\partial \hat{\mathbf{Z}}} = \mathbf{L}^T\frac{\partial \mathcal{L}}{\partial \mathbf{Z}}.$$

We can similarly optimize noise precision ω_u and lengthscales ℓ_u.