A note on the prior of GPDM latent variables

Jack M. Wang
Stanford University

The distribution over latent variables is obtained by marginalizing over the weights $\mathbf{A}$:

$$p(\mathbf{X} \,|\, \bar{\alpha}) = p(\mathbf{x}_1) p(\mathbf{X}_{2:N} | \mathbf{x}_1, \bar{\alpha}) = p(\mathbf{x}_1) \int p(\mathbf{X}_{2:N} \,|\, \mathbf{x}_1, \mathbf{A}, \bar{\alpha}) \, p(\mathbf{A}) \, d\mathbf{A}, \quad (1)$$

where $\bar{\alpha}$ is a vector of kernel hyperparameters. Incorporating assumptions of the Markov property and the independence of output dimensions gives

$$p(\mathbf{X}_{2:N} \,|\mathbf{x}_1, \bar{\alpha}) = \int \prod_{t=2}^{N} p(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, \mathbf{A}, \bar{\alpha}) p(\mathbf{A} \,|\, \bar{\alpha}) \, d\mathbf{A} = \int \prod_{t=2}^{N} \prod_{i=1}^{d} p(x_{t,i} \,|\, \mathbf{x}_{t-1}, \mathbf{A}, \bar{\alpha}) \, d\mathbf{A}. \quad (2)$$

Since $x_{:,i}$ only depends on $\mathbf{a}_{:,i}$, factoring $p(\mathbf{A})$ and rearranging the integration order gives

$$\begin{aligned} p(\mathbf{X}_{2:N} \,|\mathbf{x}_1, \bar{\alpha}) &= \int \prod_{t=2}^{N} \prod_{i=1}^{d} p(x_{t,i} \,|\, \mathbf{x}_{t-1}, \mathbf{a}_{:,i}, \bar{\alpha}) p(\mathbf{a}_{:,i}) \, d\mathbf{A} \\ &= \int_{\mathbf{a}_{:,1}} \cdots \int_{\mathbf{a}_{:,d}} \prod_{i=1}^{d} \prod_{t=2}^{N} p(x_{t,i} \,|\, \mathbf{x}_{t-1}, \mathbf{a}_{:,i}, \bar{\alpha}) p(\mathbf{a}_{:,i}) \, d\mathbf{a}_{:,d} \ldots d\mathbf{a}_{:,1} \\ &= \prod_{i=1}^{d} \int_{\mathbf{a}_{:,i}} \prod_{t=2}^{N} p(x_{t,i} \,|\, \mathbf{x}_{t-1}, \mathbf{a}_{:,i}, \bar{\alpha}) p(\mathbf{a}_{:,i}) \, d\mathbf{a}_{:,i}. \end{aligned} \quad (3)$$

The RHS of (3) is identical in form to the likelihood of of a GPLVM with input $\mathbf{X}_{1:N-1}$ and output $\mathbf{X}_{2:N}$. Therefore we obtain

$$p(\mathbf{X} \,|\, \bar{\alpha}) = \frac{p(\mathbf{x}_1)}{\sqrt{(2\pi)^{(N-1)d} |\mathbf{K}_X|^d}} \exp\left(-\frac{1}{2} \mathrm{tr}\left(\mathbf{K}_X^{-1} \mathbf{X}_{2:N} \mathbf{X}_{2:N}^T \right) \right), \quad (4)$$

where $\mathbf{X}_{2:N} = [\mathbf{x}_2, \ldots, \mathbf{x}_N]^T$, and $\mathbf{K}_X$ is the $(N-1) \times (N-1)$ kernel matrix constructed from $\mathbf{X}_{1:N-1} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N-1}]^T$.