1 Kernel expression

The Gaussian process in consideration is a mixture of independent processes. In specific, there is one global process $\mathcal{GP}_g$ and a collection of local processes $\{\mathcal{GP}_{l_i}\}$. A local non-stationary weight is added to each of the processes, to balance between exploring with global process and exploiting with local processes. The weight is modeled as Gaussian functions, to give the expression below:

$$f(\mathbf{x}) = e^{-\frac{\|\mathbf{x}-\boldsymbol{\psi_g}\|_2^2}{2*\sigma_g^2}} f_g(\mathbf{x}) + \sum_i e^{-\frac{\|\mathbf{x}-\boldsymbol{\psi_l}\|_2^2}{2*\sigma_{l_i}^2}} f_{l_i}(\mathbf{x}),$$

$$f_g \sim \mathcal{GP}_g, \quad f_{l_i} \sim \mathcal{GP}_{l_i},$$

$\boldsymbol{\psi}$ denotes the position of the center of the influence region of a process.

Additivity of Gaussian processes results in the sum being Gaussian processes. If we further assume no correlation between the $\mathcal{GP}$ s, then we may describe $\mathcal{GP}_{tot}$: $f(\mathbf{x}) \sim \mathcal{GP}_{tot}$ uniquely with mean function and covariance kernel as:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\frac{\|\mathbf{x_1}-\boldsymbol{\psi_g}\|_2^2 + \|\mathbf{x_2}-\boldsymbol{\psi_g}\|_2^2}{2\sigma_g^2}\right) k_g(\mathbf{x}_1, \mathbf{x}_2)$$

$$+ \sum_i \exp\left(\frac{\|\mathbf{x_1}-\boldsymbol{\psi_g}\|_2^2 + \|\mathbf{x_2}-\boldsymbol{\psi_g}\|_2^2}{2\sigma_{l_i}^2}\right) k_{l_i}(\mathbf{x}_1, \mathbf{x}_2),$$

$$m(\mathbf{x}) = m_g(\mathbf{x}) + \sum_i m_{l_i}(\mathbf{x})$$

Our assumptions are: 1). Local Kernels are at a same place, and the areas of influence are isotropic. 2). input of $\mathbf{x}$ is vaguely standardized to $[-1, 1]_d$, which can be used for setting the priors of sub-kernel hyperparameters and position/weights hyperparameters. 3). Global kernel has near uniform weight, which can be simulated by placing at $\boldsymbol{\psi_g} = [0.5]_d$ and σ_g being large, e.g. taken to be 10. 4). Global weight σ_g isn't a hyperparameter, while local weights σ_{l_i}s are hyperparameters. If necessary, we can use a unified σ_l. 5). To emphasize the local/global weights, $\sigma_l \ll \sigma_g$ might also be necessary. The practical hyperparameter might be σ_l/σ_g, constrained to $(0, 1)$, or its logarithm constrained to $(-\infty, 0)$, instead of σ_l.