Supplementary Material
Scalable Gaussian Processes for Characterizing Multidimensional Change Surfaces

William Herlands[1] **Andrew Wilson**[1] **Hannes Nickisch**[2] **Seth Flaxman**[3]
Daniel Neill[1] **Wilbert van Panhuis**[4] **Eric Xing**[1]
[1]Carnegie Mellon University [2]Philips Research Hamburg [3]University of Oxford [4]University of Pittsburgh

Initialization of $w(x)$ RKS Features

To initialize $w(x)$ defined by RKS features we first simplify our change surface model and assume that each latent function $f_1, ..., f_r$ from Eq. 8 is drawn from a Gaussian process with an RBF kernel. Since RBF kernels have many fewer hyperparameters than spectral mixture kernels, this enables the initialization to focus on $w(x)$. Algorithm 2 provides the procedure for initializing this simplified change surface model. Note that depending on the application domain, a model with latent functions defined by RBF kernels may be sufficient.

Algorithm 2 Initialize RKS $w(x)$ by optimizing a simplified model with RBF kernels

1: **for** $i = 1 : g$ **do**
2: Draw a, ω, b for RKS features in $w(x)$
3: Draw h random values for RBF kernels. Choose the best with maximum marginal likelihood
4: Partial optimization of $w(x)$ and RBF kernels
5: **end for**
6: Choose the best set of hyperparameters with maximum marginal likelihood
7: Optimize all hyperparameters until convergence

In the algorithm, we test multiple possible sets of values for $w(x)$ by drawing the hyperparameters a, ω, and b from their respective prior distributions g number of times. To recall the prior distributions from Section 3.1 were,

$$a \sim \mathcal{N}(0, \frac{\sigma_0}{m} I) \tag{1}$$

$$\omega_i \sim \mathcal{N}(0, \frac{1}{4\pi^2}\Lambda^{-1}) \tag{2}$$

$$b_i \sim \text{Uniform}(0, 2\pi) \tag{3}$$

We set reasonable values for hyperparameters in the prior distributions. Specifically, we let $\Lambda = (\frac{range(x)}{2})^2$, $\sigma_0 = \text{std}(y)$, and $\sigma_n = \frac{\text{mean}(|y|)}{10}$. These choices are similar to those used in Lázaro-Gredilla et al. (2010).

For each set of $w(x)$ hyperparameters that we sample, we sample sets of hyperparameters for the RBF kernels h number of times and select the set that yeilds the maximum marginal likelihood. Then we run an abbreviated optimization procedure over each set of $w(x)$ and RBF hyperparameters and finally select the joint set that yeilds the maximum marginal likelihood. Finally, we optimize all the resulting parameters until convergence.

References

Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C. E., and Figueiras-Vidal, A. R. (2010). Sparse spectrum gaussian process regression. *The Journal of Machine Learning Research*, 11:1865–1881.