Update overview

Date: August 23, 2021

» New covariance adjustments

» External comparisons (GPy, PCGP with matrix completion)

Old covariance adjustments

Previously we proposed an adjusted covariance matrix to handle the presence of missing data, i.e.

$$\mathbf{R}_k + \operatorname{diag}(\boldsymbol{w}_k),$$

where

$$
\begin{aligned}
w_{ik} &:= \operatorname{Var}(\boldsymbol{g}_{ik}(\theta) - \tilde{\boldsymbol{g}}_{ik}(\theta)) \\
&= 1 - \Phi_{\cdot k}^{\mathsf{T}} C_{J(i),\cdot}^{\mathsf{T}} C_{J(i),J(i)}^{-1} C_{J(i),\cdot} \Phi_{\cdot k},
\end{aligned}
$$

Updated covariance adjustments

We are proposing new covariance adjustments, being

$$\mathbf{R}_k + \beta_k \mathrm{diag}(\mathbf{v}_k),$$

where

$$v_{ik} = \frac{w_{ik}}{(1 - w_{ik})^\alpha}, \quad \alpha \in [0, 1].$$

Intuition:

» If complete data is observed for ith row, $v_{ik} = 0$.
  No adjustment is introduced.

» If no data is observed, $w_{ik} = 1$. And if $\alpha > 0$, $v_{ik} = \infty$.
  The row is disregarded.

The choices of α, β_k

Hyperparameter $\alpha \in [0, 1]$

» controls the rate of increase of variance v_{ik} as w_{ik} increases.

Hyperparameter $\log \beta_k \in (-12, 12)$

» controls the inflation (or deflation) of additional variances for the kth surrogate component.

» When β_k is large, any rows with missing data will be disregarded.

» β_k's can be included in hyperparameter estimation (through minimizing the negative loglikelihood)

Numerical experiments to determine α, β_k

» Functions: borehole, OTL circuit, Wing weight, and piston

» Number of locations: 25

» Number of training parameters: 50

» x are sampled uniformly in $[0, 1]^{d_x}$

» θ are sampled from latin hypercube sampling in $[0, 1]^{d_\theta}$

» Test parameters are sampled uniformly in $[0, 1]^{d_\theta}$

» Failures are **random**

Notice the optimized β value in the legend.

Numerical results for α, β_k choices
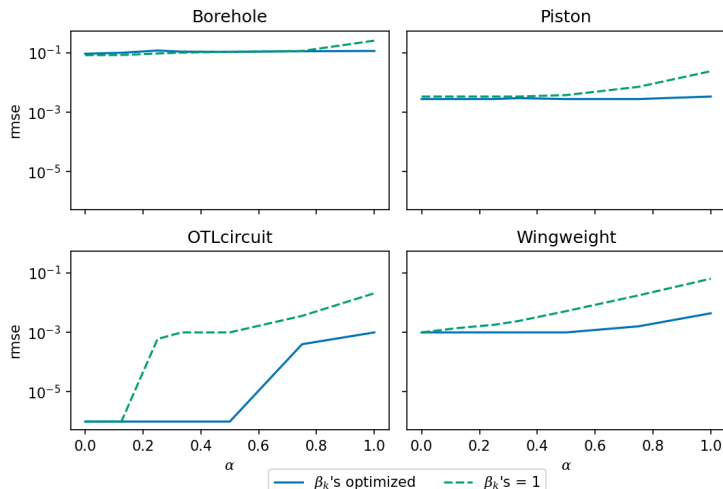
Summary for α, β_k choices

» Suggest $\alpha = 0.3$.

» Suggest β_k's to be included in optimization.

External comparisons for emulation strategies

» GPy, PCGPwM, and PCGPwMatComp are compared
» Functions: borehole, OTL circuit, Wing weight, and piston
» Number of locations: 10
» Number of training parameters: 10, 25, 50, 100, 250, 500, 1000, 1500
» Failures are **random**
» Comparison metrics:
  » RMSE, MAE, median absolute error, mean error (bias)
  » 90% coverage, 90% prediction interval width
  » interval score, CRPS