Gaussian Process (GP)

Descriptions of Gaussian Process (GP) in this discussion follow that in Wang et al.[1] and Acar and Rais-Rohani[2]. The GP metamodel is a group of output variables $f_N = \{f_n(x_n^1, x_n^2, \dots, x_n^1)\}_{n=1}^N$ with a Gaussian joint probability distribution

$$P(f_N|C_N, X_N) = \frac{1}{\sqrt{(2\pi)^N |C_N|}} \exp[-\frac{1}{2}(f_N - \mu)^T {C_N}^{-1}(f_N - \mu)] \quad (2.2)$$

where $X_N \equiv \{x_n\}_{n=1}^N$ are N pairs of L-dimensional input variables $x_n = {x_n}^1, {x_n}^2, \dots, {x_n}^L$, C_N is the covariance matrix with elements of $C_{ij} = C(x_i, x_j)$, and μ is the mean output vector.

Elements of the covariance matrix C_N are calculated from

$$C_{ij} = \theta_1 exp\left[-\frac{1}{2}\sum_{l=1}^{L}\frac{\left(x_i^{(l)} - x_j^{(l)}\right)^2}{r_l^2}\right] + \theta_2 \quad (2.3)$$

$$C_{ij} = \theta_1 exp\left[-\frac{1}{2}\sum_{l=1}^{L}\frac{\left(x_i^{(l)} - x_j^{(l)}\right)^2}{r_l^2}\right] + \theta_2 + \delta_{ij}\theta_3 \quad (2.4)$$

where $\theta_1, \theta_2, \theta_3$, and $r_l$are referred to as "hyperparameters" with r_l being the length scale. θ_3 is an independent noise parameter and δ_{ij} is Kronecker's delta (equal to one when i = j and zero otherwise). These hyperparameters are selected to maximize logarithmic likelihood of the predictions matching the training data. This is given by

$$LL = -\frac{1}{2}log|C_N| - \frac{1}{2}{f_N}^T {C_N}^{-1} f_N - \frac{N}{2}\log(2\pi) + \ln(P(\theta)) \quad (2.5)$$

where $P(\theta)$is the prior distribution of the hyperparameters. This is usually uniform because no prior knowledge is available and can be equated to zero for optimization.

Eq. (2.3) and Eq. (2.4) define the interpolation and regression modes of the Gaussian process model, respectively. The former passes through all training points while the latter provides a smoother surface to help with noisy data. The prediction surface with noise filtered

out is less complex and might not pass through all training points but this has better predictions at non-training points.

The response value at a prediction point $x_p = (x_p^1, x_p^2, \dots, x_p^L)$ is estimated as

$$\hat{f}(x_p) = k^T {C_N}^{-1} f_N \tag{2.6}$$

where $k = \left[C(x_1, x_p), \dots, C(x_N, x_p)\right]$. Standard deviation at the prediction point is available without requiring additional simulations or tests and can be calculated from

$$\sigma_{\hat{f}}(\boldsymbol{x}_p) = \kappa - k^T {C_N}^{-1} k \tag{2.7}$$

where $\kappa = C(x_p, x_p)$.

The MATLAB toolbox from Rasmussen and Williams[3] is used to develop the GP metamodels.

References

[1] Wang, L. Beeson, D. Wiggs, G. Rayasam, M. "A Comparison Of Meta-modeling Methods Using Practical Industry Requirements," Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Newport, RI, 2006.

[2] Acar, E. and Rais-Rohani, M. "Ensemble of Metamodels with Optimized Weight Factors, *"Structural and Multidisciplinary Optimization,* Vol. 37, No. 3, 2008, pp. 279-294.

[3] Rasmussen, C., Williams, C. "Gaussian Processes for Machine Learning," MIT, Cambridge, Massachusetts, 2006.