Gaussian Process Modeling for Polymer Dynamics

Pankaj Chouhan, Sachin Shanbhag Department of Scientific Computing

Objectives

Using a surrogate model (SM), we seek to mimic the input-output relationship, $\boldsymbol{x} \rightarrow \boldsymbol{h}$, of computationally expensive molecular models of polymer rheology. We build an SM, $\boldsymbol{h} = \hat{m}(\boldsymbol{x})$, where $\hat{m} : \mathbb{R}^d \rightarrow \mathbb{R}^N$, with $\boldsymbol{x}$ representing the structure of the polymer mixture, and the output $\boldsymbol{h}$ representing the rheology. The things we considered when building the SM:

1. Computational Cost
2. Prediction Accuracy
3. Functional Output

Input Data

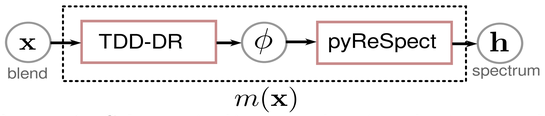

Figure 1: Schematic diagram showing the true model $\hat{m}(\boldsymbol{x})$ (dotted line), which includes the TDD-DR model and the pyReSpect program. It takes in polymer blend information in the form of $\boldsymbol{x} = [Z_1, Z_2, \rho_1, \rho_2, w_1]$ and yields the relaxation spectrum.

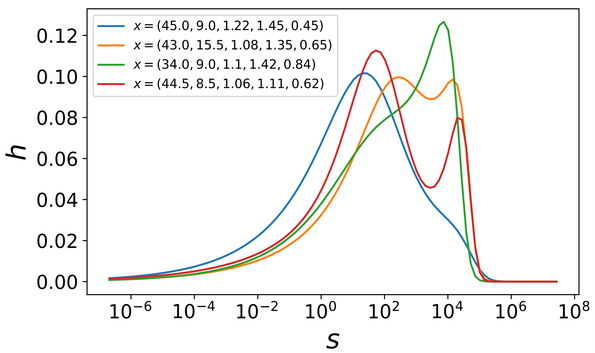

Figure 2: Four different realization of relaxation spectrum $\boldsymbol{h}(s)$, observed on a grid s of size 100, $N = 100$.

References

[1] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.

[2] Pankaj Chouhan and Sachin Shanbhag. Surrogate modeling with gaussian processes for an inverse problem in polymer dynamics. *International Journal of Computational Methods*, page 2143003, 2022.

GPsep

Given n different input observations, and each observation being a vector-valued function defined over a grid s of size N, a naïve GP regression implementation cost $\mathcal{O}(n^3N^3)$ [1].

In [2], we used a separable kernel

$$k(\boldsymbol{x}_i, s; \boldsymbol{x}_j, s') = k_{\boldsymbol{x}}(\boldsymbol{x}_i, \boldsymbol{x}_j).k_s(s, s') \qquad (1)$$

This reduces the computational cost

$$\mathcal{O}(n^3) + \mathcal{O}(N^3) \qquad (2)$$

Using a separable kernel has the following drawback

1. Constant predictive uncertainty $\hat{\sigma}^2_*$.
2. Can address only stationary processes.

KLGP and KLSVGP

1. Karhunen-Loève expansion-based gaussian process (KLGP), a modified model based on the functional principal component analysis.
2. Using KL expansion, $\boldsymbol{h}$ can be written as

$$\boldsymbol{h}(s, \boldsymbol{x}) = \bar{\boldsymbol{h}}(s) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i(\boldsymbol{x})\phi_i(s) \qquad (3)$$

   where ϕ_i and λ_i are eigenvalues and eigenvectors of empirical co-variance matrix built using training labels $\boldsymbol{h}$. $\xi_i \sim N(0, 1)$ is a $i.i.d$ random variable, that we fit using a GP.
3. We pick J principal modes such that 99% of the original variance is captured.
4. Using KLGP, one large $\mathbb{R}^d \rightarrow \mathbb{R}^N$ problem is split into J small $\mathbb{R}^d \rightarrow \mathbb{R}$ problems. Thus, the final training cost becomes $J\mathcal{O}(n^3) + \mathcal{O}(N^3)$.
5. KLSVGP, a model that allows batch training of the KLGP model, uses m 'inducing' points to summarize n training points, where $n >> m$. The cost of training KLSVGP is $\mathcal{O}(Jm^3)$.

Results

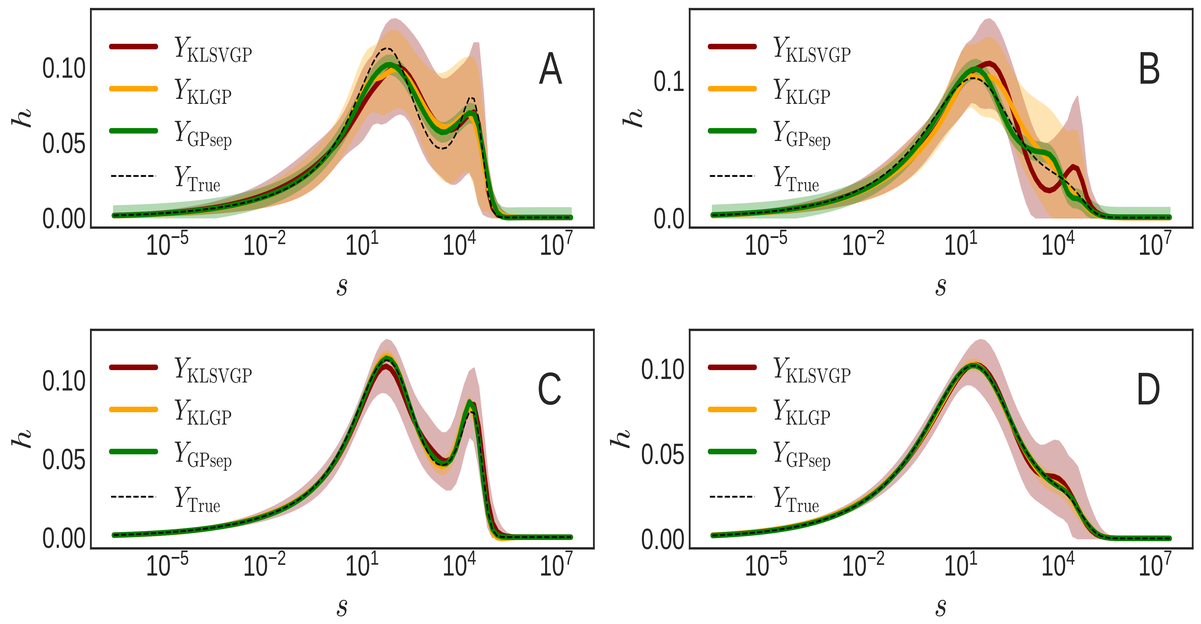

Figure 3: This figure illustrates the predicted mean along with 95% confidence region for two test points $\boldsymbol{x}_* = \{[34.0, 9.0, 1.10, 1.42, 0.84], [44.5, 8.5, 1.06, 1.11, 0.62]\}$ in first and second column respectively. The first and second row corresponds to the training set of size $n = \{102, 1608\}$ respectively. The dashed black line represents the true relaxation spectrum $\boldsymbol{h}$.

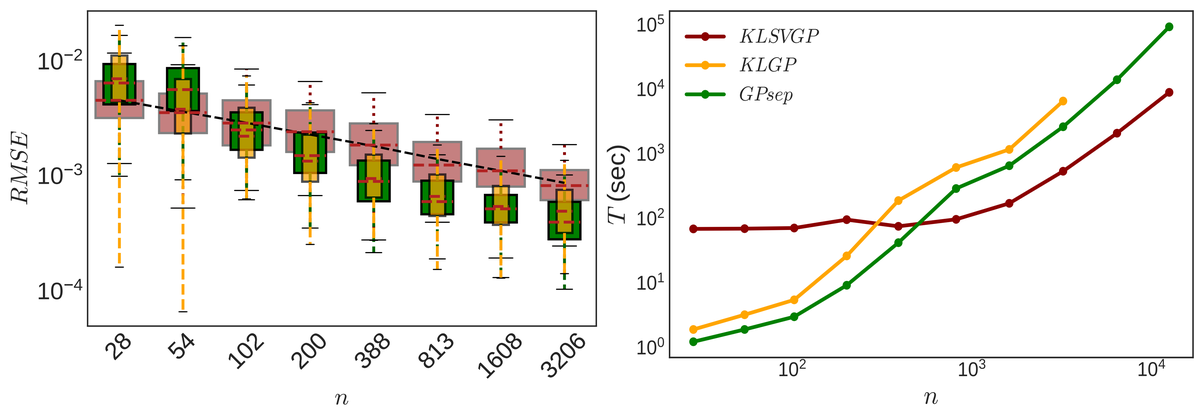

Figure 4: This figure illustrates the variation of root mean square error(RMSE) and computational time with n. As n increases, the model learns a better input-output mapping. However, it comes at the expense of more computational resources. The asymptotic limit of computational complexity can be observed for larger n.

Future Research

Explore KLGP and KLSVGP models on a real-world dataset, preferably on a dataset where a separable kernel is not suited. Furthermore, implement a parallel version of KLSVGP and KLGP.

Contact Information

Email pc19d@fsu.edu
Email sshanbhag@fsu.edu
Phone +1 (850)-339-6651