Bounded Gaussian Process Regression

Andrew Pensoneault

The University of Iowa

April 24, 2019

Overview

Introduction

- In many applications, prior knowledge is often available on the lower and/or upper bound of a response.
  - This could be nonnegativity, physical constraints, etc
- Gaussian processes are a very powerful and flexible tool for regression
  - However they have unbounded support, and thus can lead to infeasable approximations
- The paper by Jize Zhang and Lizhen Lin [1] approached addressed this problem with three novel approaches
  - Applying the bound information in parameter selection
  - Applying the bound information in projection
  - Applying the bound information in both steps

Gaussian Process Regression

- A Gaussian Process (GP) is a distribution for random functions for which any evaluation of a finite set will follow a multivariate normal distribution
- We shall consider the problem in this talk of modeling some function $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$, $x \in \mathbb{R}^{n_k}$, with further assumption that

$$m(x) = 0, \;\; k(x, x') = \sigma^2 \exp \frac{-||x - x'||^2_\Theta}{2}$$

where $\Theta = \text{diag}([\theta_i^2]_{i=1}^{n_k})$ and $||x||_\Theta = x^T \Theta^{-1} x$

Gaussian Process Regression

- Define $X = [x_1, ..., x_N]$, $Y = [y_1, ..., y_N]$, $X^* = [x_1^*, ..., x_M^*]$, the inputs, responses, and data [2].
- Define $K = [k(x_i, x_j)]_{ij}$, and $K^{**} = [k(x_i^*, x_j^*)]_{ij}$ and $k^* = [k(x_i, x_j^*)]_{ij}$, then

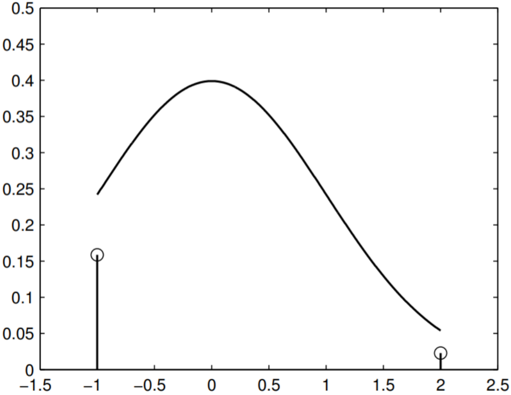
$$f(X^*) \sim \mathcal{N}(\mu_f(X^*), \sigma_f^2(X^*))$$

$$\mu_f(X^*) = {K^*}^T K^{-1} Y,$$

$$\sigma_f^2(X^*) = K^{**} - {K^*}^T K^{-1} K^*$$

Bounded Gaussian Process

- The approach taken here is to perform standard GP regression, and apply bounds after the fact
  - This leads to at each location, the random variable being the mixture of a Bernoulli random variable and truncated Gaussian random variable



Simulations

- Each GP has parameters selected using cross validation, which involved solving a nonlinear optimization problem
- A genetic algorithm was used to perform the minimization
- The problem chosen was the second example in the paper,

$$y = x^2 \sin(x^{-1}), x \in \left[-\frac{\pi}{8}, \frac{\pi}{8} \right]$$

$$l(x) = -x^2, u(x) = x^2$$

  which has both an upper and lower bound and is highly nonlinear and oscillatory.
- Pre-normalization was performed on both covariates and responses for training
- Training samples were selected from $N = 15$ Latin Hypercube sample, and $m = 1000$ uniform points were used for testing
- Evaluation was based on RMSE and R^2 coefficients

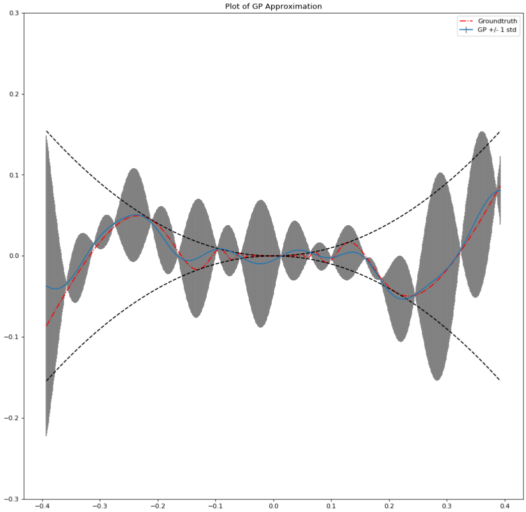
GP

- The first is an approach using standard Cross Validation, LOO-CV PRESS minimization, i.e.

$$\underset{\theta}{\text{argmin}} \quad \left\{ \sum_{i=1}^{N} (y_i - \mu_{f,-i})^2 \right\}$$

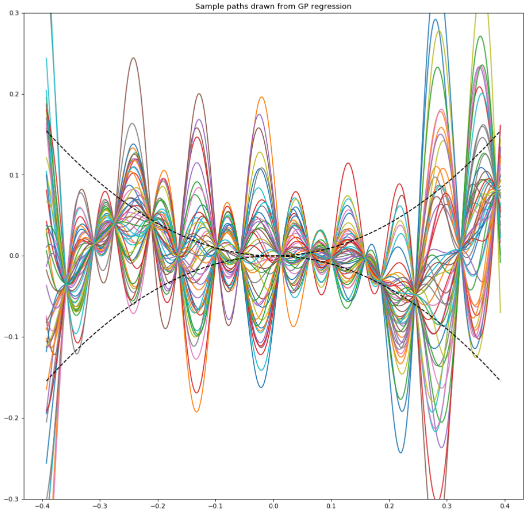
$$\begin{aligned}
\mu_{f,-i} =& y_i - \left[\mathbf{K}^{-1}\mathbf{Y}\right]_i / \left[\mathbf{K}^{-1}\right]_{ii} \\
\sigma^2_{f,-i} =& 1 / \left[\mathbf{K}^{-1}\right]_{ii} \\
\hat{\sigma} =& \left(\frac{1}{N}\mathbf{Y}^T\mathbf{K}^{-1}[\text{diag}(\mathbf{K}^{-1})]^{-1}\mathbf{K}^{-1}\mathbf{Y}\right)
\end{aligned}$$

GP



Plot of GP Approximation
Groundtruth
GP +/- 1 std
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

GP

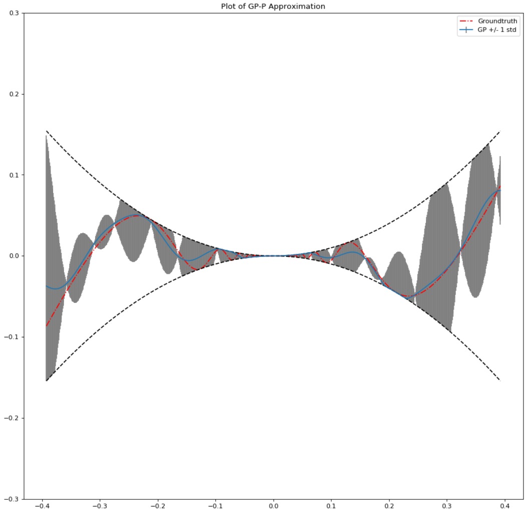


Sample paths drawn from GP regression
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

bGP-P

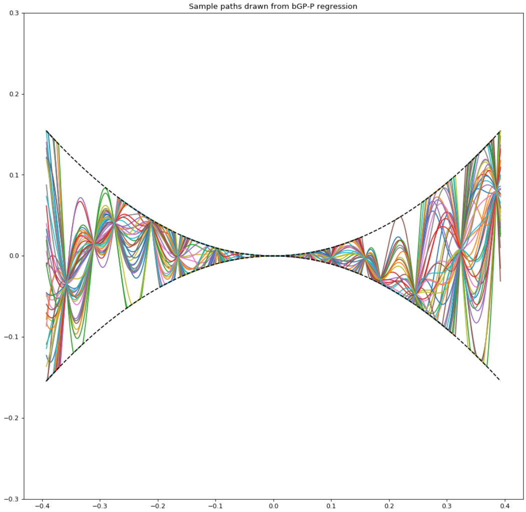
- This approach takes the above solution and projects it into our bounded space

bGP-P



Plot of GP-P Approximation
Groundtruth
GP +/- 1 std
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

bGP-P



Sample paths drawn from bGP-P regression
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

bGP-I

- With this algorithm, we also minimize the LOO-CV PRESS, however, instead of using the unconstrained prediction mean, we use the LOO-CV constrained mean, and additionally to mitigate sensitivity issues, we wish to place a constraint on σ^2 being reasonably close to the unconstrained estimate

$$
\begin{aligned}
\underset{\sigma^2,\theta}{\operatorname{argmin}} \quad & \left\{ \sum_{i=1}^{N} (y_i - \mu_{g,-i})^2 \right\} \\
\text{subject to} \quad & \sigma^2 \geqslant c_l\left(\frac{1}{N}\mathbf{Y}^T\mathbf{K}^{-1}[\mathrm{diag}(\mathbf{K}^{-1})]^{-1}\mathbf{K}^{-1}\mathbf{Y}\right) \\
& \sigma^2 \leqslant c_u\left(\frac{1}{N}\mathbf{Y}^T\mathbf{K}^{-1}[\mathrm{diag}(\mathbf{K}^{-1})]^{-1}\mathbf{K}^{-1}\mathbf{Y}\right).
\end{aligned}
$$

bGP-I

- In the paper the values of mean and covariance were derived to be as seen below, and to determine the $\mu_{g,-i}$, we will plug in $\mu_{f,-i}$ and $\sigma^2_{f,-i}$ into below

$$Z = \Phi\left(\alpha; 0, 1\right) - \Phi\left(\beta; 0, 1\right)$$

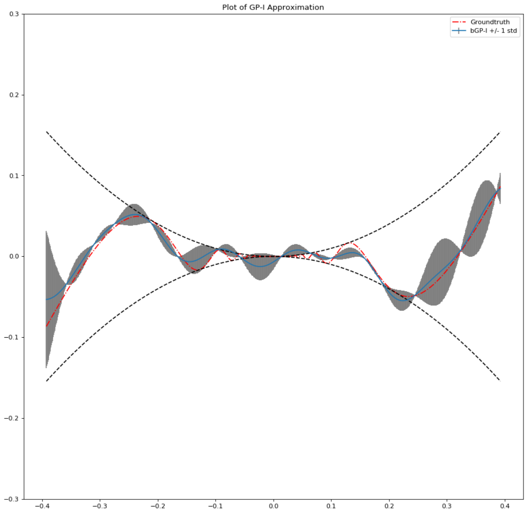
$$\alpha = \frac{l(x) - \mu_f}{\sigma_f}, \ \beta = \frac{u(x) - \mu_f}{\sigma_f}$$

$$\mu_g = Z \cdot \mu_f + \left[\phi\left(\alpha; 0, 1\right) - \phi\left(\beta; 0, 1\right)\right] \cdot \sigma_f + l(x) \cdot \Phi\left(\alpha; 0, 1\right) + u(x) \cdot \left[1 - \Phi\left(\beta; 0, 1\right)\right]$$

$$\sigma_g^2 = Z \cdot \left[\sigma_f^2 + \mu_f^2\right] + 2\mu_f\sigma_f \cdot \left[\phi\left(\alpha; 0, 1\right) - \phi\left(\beta; 0, 1\right)\right] + \sigma_f^2 \cdot \left[\alpha\phi\left(\alpha; 0, 1\right) - \beta\phi\left(\beta; 0, 1\right)\right]$$

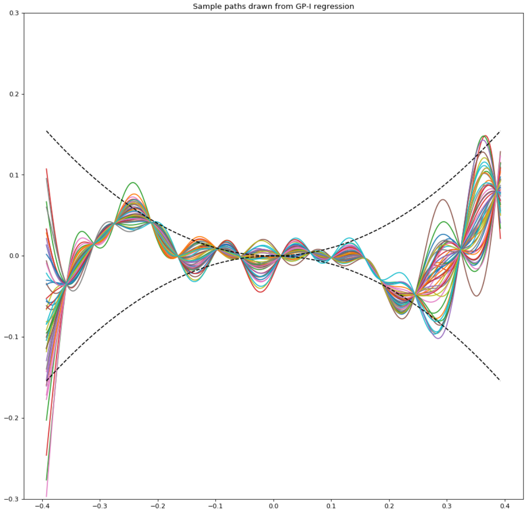
$$+ l(x)^2 \cdot \Phi\left(\alpha; 0, 1\right) + u(x)^2 \cdot \left[1 - \Phi\left(\beta; 0, 1\right)\right] - \mu_g^2$$

bGP-I



Plot of GP-I Approximation
Groundtruth
bGP-I +/- 1 std
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

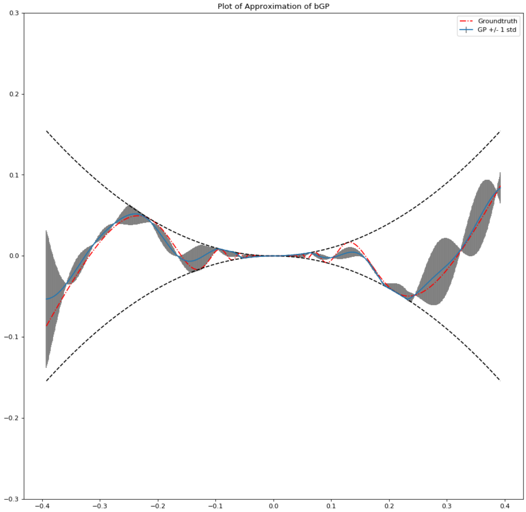
bGP-I


Sample paths drawn from GP-I regression
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

bGP

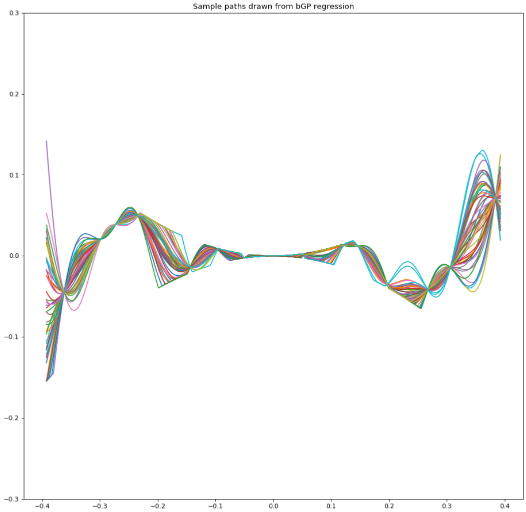
- The full bGP combines both previous approaches to utilize the bound information, first optimizing then projecting

bGP



Plot of Approximation of bGP
Groundtruth
GP +/- 1 std
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

bGP



Sample paths drawn from bGP regression
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

Discussion

	GP	bGP-P	bGP-I	bGP
RMSE	0.008703	0.008278	0.007717	0.006940
R^2	0.923077	0.930539	0.939431	0.951127

- From above, the inclusion of bound information decreases RMSE and increases R^2 coefficient
- This is consistent with the findings of the paper

Next Steps

- Further work would potentially exploring the extension to multidimensional outputs
- If a suitable approach could be used as a surrogate model in the EnKF for costly models
- A full Bayesian analysis in the place of point estimate could be investigated to eliminate the need for hyper-parameter selection in this new projected space

Github

- The code used in this presentation is available at `https://github.com/Andrewpensoneault/bounded_gp_regression`

References I

Lizhen Lin Jize Zhang.
Bounded regression with gaussian process projection, 2018.

C. E. Rasmussen C. K. I. Williams.
Gaussian processes for machine learning, 2006.