Gaussian Process Regression Subject to Chance Constraints

Kewei Wang, Georg Stadler

Courant Institute of Mathematical Sciences, New York University

Goal

Incorporate probabilistic bound and monotonicity constraints in Gaussian processes and linear inverse problems, maintaining the posterior distribution as Gaussian. This is in contrast to alternative methodologies that utilize nonlinear transformations or truncated Gaussian distributions.

Gaussian Process Regressions (GPR)

Forward model: $\boldsymbol{y} = Fu + \boldsymbol{\eta}$, $F : u \mapsto (u(x^{(1)}), \dots, u(x^{(n_{\text{ob}})}))$.
Observation pairs: $(\boldsymbol{X}, \boldsymbol{y}) = (x^{(i)}, y^{(i)})_{i=1}^{n_{\text{ob}}}$.
Noise/errors: $\boldsymbol{\eta} \sim \mathcal{N}(0, \Gamma_{\text{n}}) = \mathcal{N}(0, \sigma_{\text{n}}^2 I)$.
Given prior GP $u_{\text{prior}} \sim \mathcal{N}(u_0, \mathcal{K})$, the posterior GP used for prediction is:

$$u_{\text{post}} \sim \mathcal{N}(u_{\text{post}}^m, \mathcal{K}_{\text{post}}), \text{ where}$$

- $\mathcal{K}_{\text{post}}(x, x') = \mathcal{K}(x, x') - \mathcal{K}(x, \boldsymbol{X})(\mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) + \Gamma_{\text{n}})^{-1}\mathcal{K}(\boldsymbol{X}, x')$.
- $u_{\text{post}}^m(x) = u_0(x) + \mathcal{K}(x, \boldsymbol{X})(\mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) + \Gamma_{\text{n}})^{-1}(\boldsymbol{y} - Fu_0)$.

Radial basis functions as the kernel:

$$\mathcal{K}(x, x') = \sigma^2 \exp\left(\frac{-\|x - x'\|_2^2}{2\ell^2}\right).$$

The covariance of $\boldsymbol{y}$ is thus

$$K := \mathcal{K}(X, X) + \Gamma_{\text{n}} = \sigma^2 \exp\left(\frac{-\|x - x'\|_2^2}{2\ell^2}\right) + \sigma_{\text{n}}^2 I.$$

GPR: Find optimal parameters of $\mathcal{K}$ by minimizing an objective $\mathcal{J}(\theta)$:

$$\min_{\theta} \quad \mathcal{J}(\theta) := -\log p(\boldsymbol{y}|\theta).$$

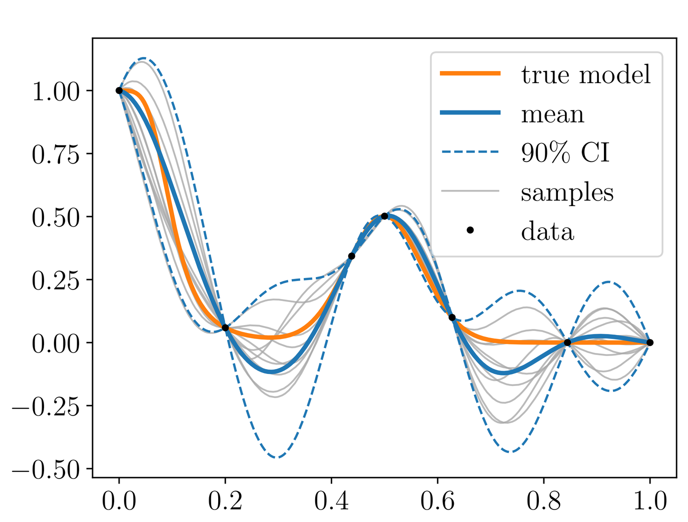
- $\theta := (\ell, \sigma, \sigma_{\text{n}})$: kernel parameters.
- $p(\boldsymbol{y}|\theta) = \frac{1}{2}\left((\boldsymbol{y} - u_0(\boldsymbol{X}))^T K^{-1}(\boldsymbol{y} - u_0(\boldsymbol{X})) + \log|K| + n_{\text{ob}} \log(2\pi)\right)$: marginal likelihood of θ.

Example: GPR for Positive Functions

Using GPR to fit a positive function with a few observations.

$$f(x) = \frac{1}{1 + (10x)^4} + \frac{1}{2}\exp\left(-100\left(x - \frac{1}{2}\right)^2\right), \ x \in \mathcal{D} = [0, 1].$$

GPR (figure on the right) is not satisfactory since most samples are not positive.
Existing approaches to enforce bound constraints:



- Truncated Gaussian distributions: $\mathcal{TN}(m, \Sigma) \propto \mathbf{1}_{[a,b]}\mathcal{N}(m, \Sigma)$.
- Warping: use GPR for $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{y})$ ($\Phi : \mathbb{R} \to [a, b]$), and use $\Phi(u_{\text{post}})$ as the surrogate model.

Both give **non-Gaussian** posteriors.

Joint Chance Constraints

Bound constraints can be incorporated into GPR in the form of joint chance constraints:

$$\varphi(\theta) := \mathbb{P}\left(\underline{u}(x) \le u_{\text{post}}(x; \theta) \le \overline{u}(x) \text{ for a.a. } x \in \mathcal{D}\right) \ge p.$$

Unlike other approaches, this keeps everything Gaussian.

Spherical-radial Decomposition (SRD)

General chance constraints: $\varphi(\theta) = \mathbb{P}(g(\theta, \xi) \le 0) \ge p$, where g is a scalar function corresponding to the joint constraint, and ξ is the (known) uncertainty.
Simple Monte Carlo approach for estimating $\varphi(\theta)$:

$$\varphi(\theta) \approx \bar{\varphi}_N(\theta) := \frac{1}{N}\sum_{i=1}^N \mathbf{1}_{\{g(\theta, z_i) \le 0\}},$$

which cannot give $\nabla\varphi(\theta)$.

SRD for $\xi \sim \mathcal{N}(m, \Sigma)$, $\Sigma = LL^T$, $L \in \mathbb{R}^{n \times k}$:

$$\xi = m + \tau L v,$$
$$\tau \sim \chi_k, \quad \text{Chi distribution with parameter } k,$$
$$v \sim \mathcal{U}(\mathbb{S}^{k-1}), \quad \text{uniform distribution on } \mathbb{S}^{k-1}.$$

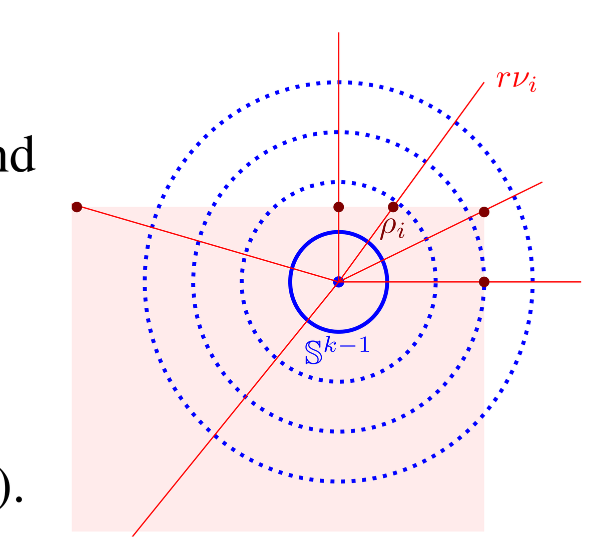
Using SRD to evaluate $\varphi(\theta)$:

$$\varphi(\theta) = \int_{v \in \mathbb{S}^{k-1}} \mu_{\chi_k}(\{r \ge 0 : g(\theta, m + rLv)\}) d\mu_{\mathcal{U}}(v).$$

Approximation:
if $\{z \in \mathbb{R}^n : g(\theta, z) \le 0\}$ is convex, and $g(\theta, m) < 0$, then

$$\varphi(\theta) \approx \tilde{\varphi}_N(\theta) := \frac{1}{N}\sum_{i=1}^N F_{\chi_k}(\rho(\theta, v_i)).$$

- $\rho(\theta, v) := \sup_{r \ge 0}\{g(\theta, m + rLv) \le 0\}$.
- $\mu_{\chi_k}(\{r \ge 0 : g(\theta, m + rLv)\}) = F_{\chi_k}(\rho(\theta, v))$.



Derivative:

$$\nabla\tilde{\varphi}_N(\theta) = -\frac{1}{N}\sum_{i=1}^N \frac{f_{\chi_k}(\rho(\theta, v_i))}{\langle \nabla_z g(\theta, m + \rho(\theta, v_i)Lv_i), Lv_i \rangle} \nabla_\theta g(\theta, m + \rho(\theta, v_i)Lv_i).$$

Low-rank Approximation for SRD

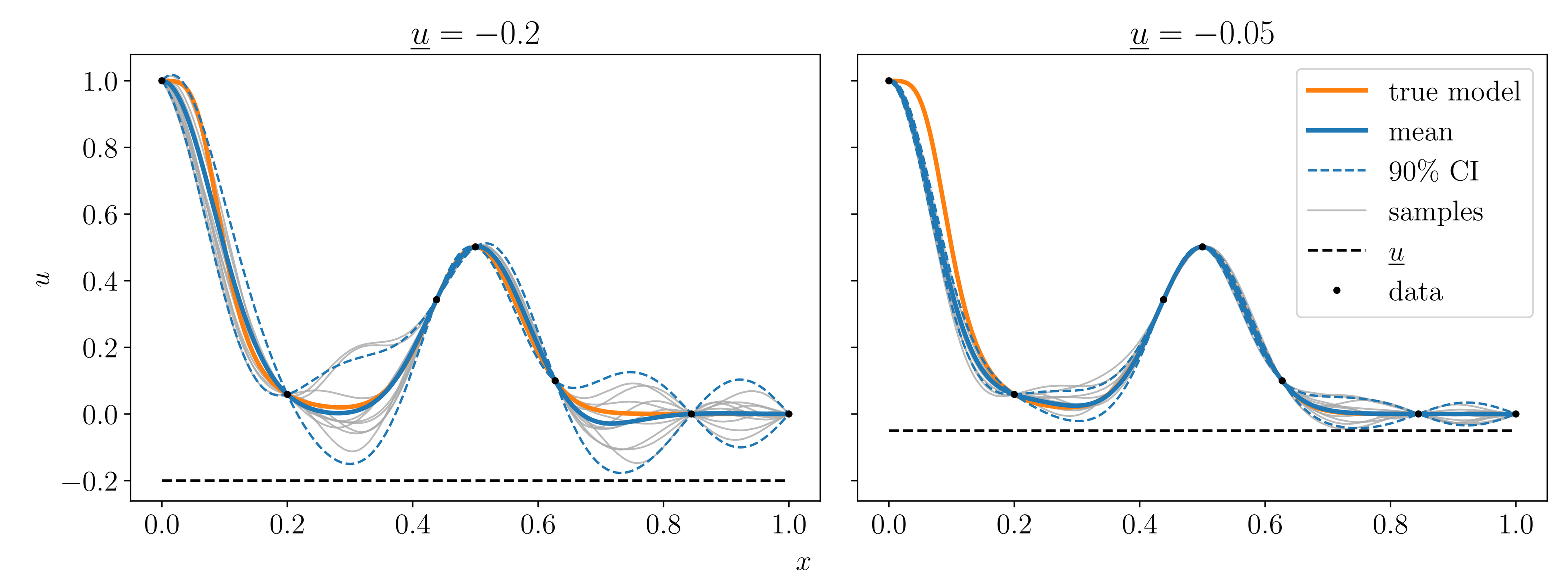
If ξ is the discretization of a GP with $L \in \mathbb{R}^{n \times n}$, n is large but the spectrum of $\Sigma = LL^T$ decays fast, making low-rank approximation possible:

$$\Sigma = LL^T \approx L_k L_k^T = \Sigma_k.$$

Thus use uniform samples from $\mathbb{S}^{k-1}$ instead of $\mathbb{S}^{n-1}$, accelerating the computation and making it disretization-invariant.

GPR with Chance Constraints

Using chance constrained GPR to fit the same function as the unconstrained case, and incorporate the chance constraint $\mathbb{P}(u_{\text{post}} \ge \underline{u} \text{ a.e.}) \ge p$ with $p = 0.95$ and different lower bound $\underline{u}$.



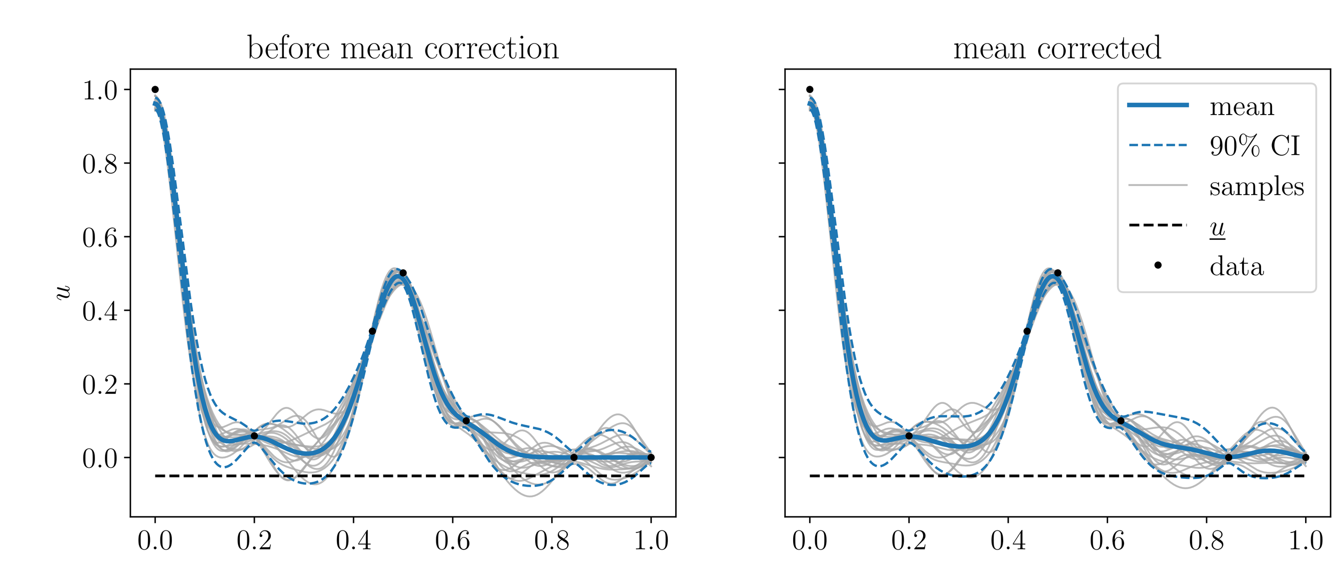
Mean Correction for Inverse Problems

A similar approach works for general Bayesian inverse problems by correcting the posterior mean:

$$\bar{u}_{\text{post}}^m = \underset{u \in H}{\operatorname{argmin}} \quad \frac{1}{2}\|Fu - \boldsymbol{y}\|_{\Gamma_{\text{n}}}^2 + \frac{1}{2}\|u - u_0\|_{\mathcal{K}}^2,$$
$$s.t. \quad \varphi(u_{\text{post}}) = \mathbb{P}(u_{\text{post}}(x) \ge \underline{u}(x) \text{ for a.a. } x) \ge p.$$

This is equivalent to minimizing $\frac{1}{2}\|u - u_{\text{post}}^m\|_{\Gamma_{\text{post}}}^2$, and the corrected posterior GP is $\bar{u}_{\text{post}} \sim \mathcal{N}(\bar{u}_{\text{post}}^m, \mathcal{K}_{\text{post}})$.

Example using the same f, with $\underline{u} = -0.05$ and $p = 0.7$:



References

[1] R. Henrion, G. Stadler, and F. Wechsung, Optimal Control under Uncertainty with Joint Chance State Constraints: Almost-Everywhere Bounds, Variance Reduction, and Application to (Bi) linear Elliptic PDEs, SIAM/ASA Journal on Uncertainty Quantification, 13 (2025), pp. 1028–1053.

[2] A. Pensoneault, X. Yang, and X. Zhu, Nonnegativity-enforced Gaussian process regression, Theoretical and Applied Mechanics Letters, 10 (2020), pp. 182–187.

Acknowledgements

Partially supported by the U.S. Department of Energy, under Award #DE-SC0024721, and the National Science Foundation under #2411229.

Email: kw3645@nyu.edu