SAND2026-18227C

Bayesian Weight Space Gaussian Process Regression in Practice

Habib N. Najm

Sandia National Laboratories
Livermore, CA, USA
hnnajm@sandia.gov

ICERM Workshop
Bayesian Inverse Problems and UQ
Brown University, Providence RI
Mar 2-6, 2026

Acknowledgement

Joint work with:

Mridula Kuppa, University of Illinois at Urbana-Champaign

Khachik Sargsyan, Shinae Kim, Sandia National Labs

This work was supported by:

- DOE Office of Science (SC); Basic Energy Sciences; Chemical Sciences, Geosciences, and Biosciences (CSGB)
- DOE SC; Advanced Scientific Computing Research (ASCR) SciDAC; FASTMath institute

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

Outline

Motivation

- Gaussian process (GP) constructions have been used extensively for Bayesian learning in function space — Rasmussen & Williams '05
  - Extended to Non-linear observation models
    - Warped GP — Snelson '04, Rios '19
    - GP latent variable models, & Deep GP — Lawrence '05, Damianou '13
  - Integral observations — O'Callagahan '11
  - Multi-task learning, heteroscedastic noise — Boyle '04, Kersting '07
  - Physics-informed/constrained GP — Yang '18, Swiler '20
- Weight-space *vs.* function-space view
  - Parametric *vs.* non-parametric
  - Seamless integration in computational/physical models
  - Facile use with
    - – linear/nonlinear hierarchical models
    - – integral observations
    - – constraints
    - – model error embedding

Problem of Interest from Quantum Chemistry

- Density functional theory (DFT) defines the potential energy of a molecule i as an integral in 3D space. A typical example:

$$y_i = \int_{\mathbb{R}^3} g_i(\boldsymbol{r}) f(u_i(\boldsymbol{r}))\mathrm{d}\boldsymbol{r}, \qquad i = 1, \ldots, N$$

with $g_i(\boldsymbol{r}), u_i(\boldsymbol{r}) \in \mathbb{R}$, and $f()$ to be estimated from data on $\boldsymbol{y} \in \mathbb{R}^N$.

- Given basis functions $\phi_k(),\ k = 0, \ldots, m$, write $f(\cdot)$ as the expansion

$$f(\cdot) := f(\cdot, \boldsymbol{w}) = \sum_{k=0}^{m} w_k \phi_k(\cdot)$$

- For Q_i quadrature points, and $\boldsymbol{u}_i := (u_i(\boldsymbol{r}_1), \ldots, u_i(\boldsymbol{r}_{Q_i}))$, the integral defines the linear map $\mathcal{L}_i() : \mathbb{R}^{Q_i} \to \mathbb{R}$, with

$$y_i = \mathcal{L}_i(f(\boldsymbol{u}_i, \boldsymbol{w}))$$

where $f(\boldsymbol{u}_i, \boldsymbol{w}),\ \phi_k(\boldsymbol{u}_i) \in \mathbb{R}^{Q_i}$

Problem of Interest from Quantum Chemistry

- Then, with a database of N molecules, let $Q = \sum_{i=1}^{N} Q_i$,

$$\boldsymbol{x} := (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N) = (u_{11}, \cdots, u_{1Q_1}, \ldots, u_{N1}, \cdots, u_{NQ_N}) \in \mathbb{R}^Q$$

and $\mathcal{L}() : \mathbb{R}^Q \to \mathbb{R}^N$, we have the linear problem in $\boldsymbol{w}$

$$\begin{aligned} \boldsymbol{y} &= \mathcal{L}(f(\boldsymbol{x}, \boldsymbol{w})) \\ f(\boldsymbol{x}, \boldsymbol{w}) &= \sum_{k=0}^{m} w_k \boldsymbol{\phi}_k(\boldsymbol{x}) \end{aligned}$$

with $\phi_k(\boldsymbol{x}),\ f(\boldsymbol{x}, \boldsymbol{w}) \in \mathbb{R}^Q$, and

$$\boldsymbol{y} = \sum_{k=0}^{m} w_k \mathcal{L}(\phi_k(\boldsymbol{x})) = A\boldsymbol{w}$$

where $A = [A_0, \ldots, A_{m-1}] \in \mathbb{R}^{N \times m}$, and $A_k = \mathcal{L}(\boldsymbol{\phi}_k(\boldsymbol{x}))$.

Some Comments on the DFT Fitting Problem

- We have no explicit point-wise information on the shape of $f()$
- For each molecule i, the y_i data point provides filtered information on $f()$ according to the distribution of the $(u_{i1}, \ldots, u_{iQ_i})$ values
  - A sharp $p(u_i)$ localized at u^* provides good information on $f(u^*)$
  - A broad $p(u_i)$ provides only spatially-averaged information on $f()$
- The shape of the aggregate distribution $p(u)$ is unknown *a priori*. It depends on
  - the electronic structures of each of the N molecules in the dataset
  - the quadrature formula
- This problem is well-fitted for GP regression, and most easily so with the weight-space view

GP Regression – function space viewpoint

- Build a prior MVN on observations, then condition it on the data
- Given a prior Gaussian distribution on function $f(x)$,

$$f(x) \sim GP(\mu(x), k(x, x'))$$

- With $y = f(x) + \sigma_{\mathrm{d}}\epsilon$, we have the prior on $\boldsymbol{y}$ at $\boldsymbol{x}$, and $\boldsymbol{f}^*$ at $\boldsymbol{x}^*$

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K_{\boldsymbol{x}\boldsymbol{x}} + \sigma_{\mathrm{d}}^2 I & K_{\boldsymbol{x}\boldsymbol{x}^*} \\ K_{\boldsymbol{x}^*\boldsymbol{x}} & K_{\boldsymbol{x}^*\boldsymbol{x}^*} \end{bmatrix}\right)$$

- Conditioning this MVN on the data $(\boldsymbol{x}, \mathbf{y})$ gives Rasmussen & Williams '06

$$\boldsymbol{f}^* | \mathbf{y} \sim \mathcal{N}\left(K_{\boldsymbol{x}^*\boldsymbol{x}}(K_{\boldsymbol{x}\boldsymbol{x}} + \sigma_{\mathrm{d}}^2 I)^{-1}\mathbf{y},\; K_{\boldsymbol{x}^*\boldsymbol{x}^*} - K_{\boldsymbol{x}^*\boldsymbol{x}}(K_{\boldsymbol{x}\boldsymbol{x}} + \sigma_{\mathrm{d}}^2 I)^{-1} K_{\boldsymbol{x}\boldsymbol{x}^*}\right)$$

- Function view provides no path through a posterior on $\boldsymbol{w}$
- Applicability to models with nonlinear dependence on the GP requires further development, e.g. warped GPs, etc

GP Regression – weight space viewpoint

- With $x \sim p(x)$, with covariance $k(x, x')$ eigenvalues/functions, have

$$k(x, x') = \sum_{k=0}^{\infty} \lambda_k \phi_k(x) \phi_k(x')$$

$$\int \phi_i(x) \phi_j(x) p(x) \mathrm{d}x = \delta_{ij}$$

and can write an expansion for $f()$ as $f(x; \boldsymbol{w}) = \sum_{k=1}^{\infty} w_k \phi_k(x)$

- Gaussian prior on $\boldsymbol{w}$, $p(\boldsymbol{w}) = \mathcal{N}(0, \Sigma_p)$, with $\Sigma_p = \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$
- Applying Bayes rule gives the $\boldsymbol{w}$ and $\boldsymbol{f}$ posteriors

$$\begin{aligned} p(\boldsymbol{w}|\boldsymbol{y}) &= \mathcal{N}(\sigma_{\mathrm{d}}^{-2} \Sigma_{\boldsymbol{w}} \Phi(\boldsymbol{x}) \boldsymbol{y},\ (\sigma_{\mathrm{d}}^{-2} \Phi(\boldsymbol{x}) \Phi(\boldsymbol{x})^T + \Sigma_p^{-1})^{-1}) \\ p(\boldsymbol{f}^*|\boldsymbol{y}) &= \mathcal{N}(\Phi(\boldsymbol{x}^*)^T \sigma_{\mathrm{d}}^{-2} \Sigma_{\boldsymbol{w}} \Phi(\boldsymbol{x}) \boldsymbol{y},\ \Phi(\boldsymbol{x}^*)^T \Sigma_{\boldsymbol{w}} \Phi(\boldsymbol{x}^*)) \end{aligned}$$

which, for high-D $\boldsymbol{w}$, approaches the function space predictive density

- However, we now have the posterior on $\boldsymbol{w}$
- With MCMC this can extend to forward models that are nonlinear in $\boldsymbol{w}$

Select challenges of interest in practice

Choice of GP kernel and its parameters

- Can be informed by prior knowledge
- Parameters can be optimized (max marg. likelihood) or inferred

Specific for the weight view

- Choice of order of the representation for improved accuracy
- Choice of kernel/basis functions to minimize $p(\boldsymbol{w}|D)$-correlation

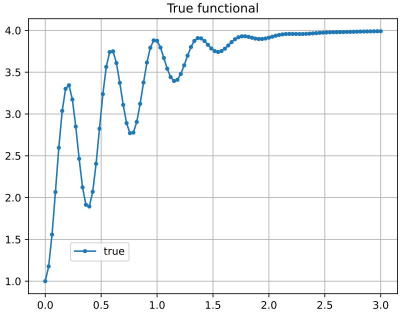
General to both contexts

- Application of constraints on the GP
- For GP-based model error constructions:
  - Disambiguation of GP and physical model

DFT Inverse problem setup with synthetic data

Define synthetic data with N "structures" for demonstration purposes

- Each structure defines a spatial (electron density) dependence on $\boldsymbol{r} \in \mathbb{R}^3$ as a Gaussian mixture with a number of 3D-MVN components with randomized means and covariances.



- This provides the functions

$$u_i(\boldsymbol{r}),\ g_i(\boldsymbol{r}), \quad i = 1, \ldots, N$$

and the data points

$$y_i = \int_{\mathbb{R}^3} g_i(\boldsymbol{r}) f(u_i(\boldsymbol{r}))\mathrm{d}\boldsymbol{r} + \epsilon_{\mathrm{d}}$$

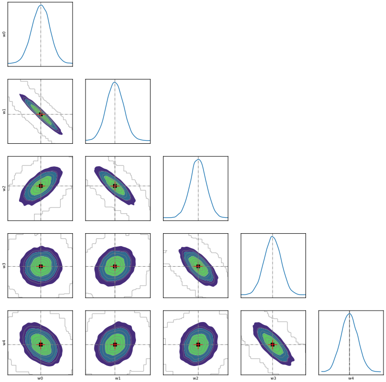
using a "true" functional*

$$f_{\text{true}}(u) := 2 - \cos(16u)\exp(-u^2) + 2\tanh(u)$$

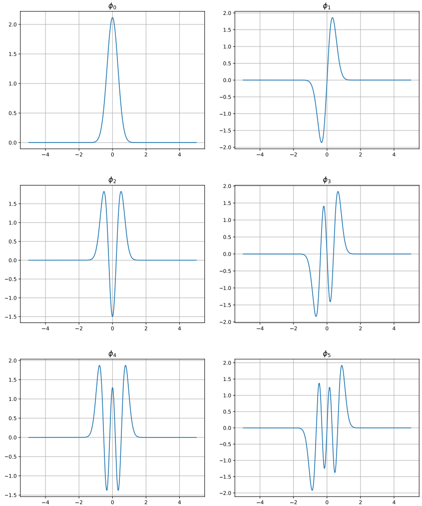
* This is a fictional functional for purposes of illustration.

Solution with $m = 20$ modes — Sq. Exp. kernel, $\ell = 0.1$

- $\mathcal{N}(\boldsymbol{w}|0, \Sigma_p)$ prior
  $\Sigma_p = \text{diag}(\lambda_0, \dots, \lambda_{m-1})$



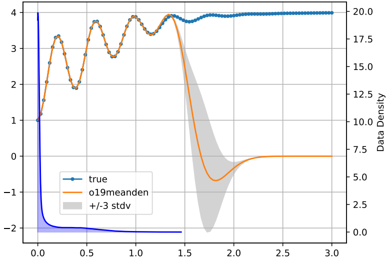
First 5 posterior dimensions



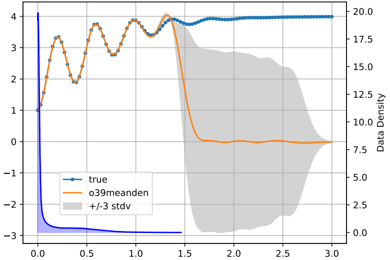
First 6 basis functions

Functional/GP posterior prediction and y-data fit

- Given GP prior, for extrapolative prediction far from data, expect
  - $\mathcal{N}(f|0,1)$ marginal density
- Decay in extrapolative predictive uncertainty is due to m truncation
- Increasing m extends the range of agreement with the function prior in the extrapolation region
- Increased dimensionality can be a concern



$m = 20$



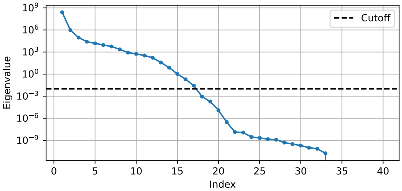
$m = 40$

Choice of the order of the expansion (m)

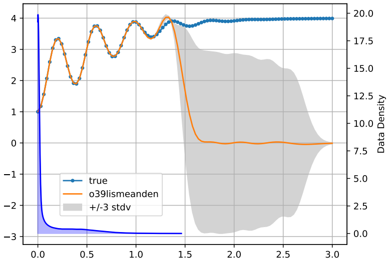
- Irrespective of the basis choice, increasing order, thus w-dimension, increases the cost of the Bayesian solution
- For the information-content in a given data set, there will be minimal change from prior to posterior in w components beyond a given order
- The likelihood informed subspace (LIS) method identifies a lower dimensional space in which to learn from the data, while pushing forward prior information in the remaining dimensions Spantini 2015
- This is most effective for linear systems
- Nonlinear model extension is also available Cui 2014

Case with $m = 40$ with reduced dimensionality using LIS

- Identify parameter subspace sufficiently informed by the data
  - Cutoff threshold at mode 18
- Estimate the posterior in this 18-dimensional subspace
- Propagate priors in uninformed 22-dimensional subspace



LIS spectrum



Choice of basis functions

- In practice, the choice of basis functions as the eigenfunctions of a guessed-at kernel will not be optimal
  - we may not have enough data to determine $p(x)$ well
  - we do not have data on $f(x, \boldsymbol{w})$ but rather on $\mathcal{L}(f(x, \boldsymbol{w}))$
- Consider the DFT Bayesian fitting problem with $\epsilon_d \sim \mathcal{N}(0, \Sigma_d)$

$$\begin{aligned} \boldsymbol{y} &= \sum_{k=0}^{m} w_k \mathcal{L}(\phi_k(\boldsymbol{x})) + \epsilon_d = A\boldsymbol{w} + \epsilon_d \\ A &= [A_0, \dots, A_{m-1}] \in \mathbb{R}^{N\times m}, \quad A_k = \mathcal{L}(\boldsymbol{\phi}_k(\boldsymbol{x})) \end{aligned}$$

- With a prior $\mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}, \Sigma_p)$, the $\boldsymbol{w}$-posterior covariance is

$$\Sigma_{\boldsymbol{w}} = (\Sigma_p^{-1} + A^\top \Sigma_d^{-1} A)^{-1}$$

- We will start from some basis functions, then update them to orthonormalize the columns of A

Orthonormalization of the basis functions - 1

- Presume a squared exponential kernel
  - eigenvalues/functions $\lambda_k, \phi_k()$, $k = 0, \dots, m-1$
- Map the basis functions:

$$\Phi(x) = (\phi_0(x), \dots, \phi_{m-1}(x)) \qquad \boldsymbol{y} = A\boldsymbol{w} + \epsilon_d, \; A_k = \mathcal{L}(\phi_k(\boldsymbol{x}))$$

to new basis functions:

$$\Theta(x) = (\theta_0(x), \dots, \theta_{m-1}(x)) \qquad \boldsymbol{y} = A^\theta \boldsymbol{w}^\theta + \epsilon_d, \; A_k^\theta = \mathcal{L}(\theta_k(\boldsymbol{x}))$$

so that the columns of A^θ are orthonormal.

- Solve for the lower triangular matrix $P \in \mathbb{R}^{m\times m}$ where

$$\Theta(x) = P\Phi(x) \qquad \text{s.t.} \qquad (A^\theta)^\top A^\theta = I.$$

where $A^\theta = AP^\top$, $\boldsymbol{w}^\theta = P^{-\top}\boldsymbol{w}$

Orthonormalization of the basis functions - 2

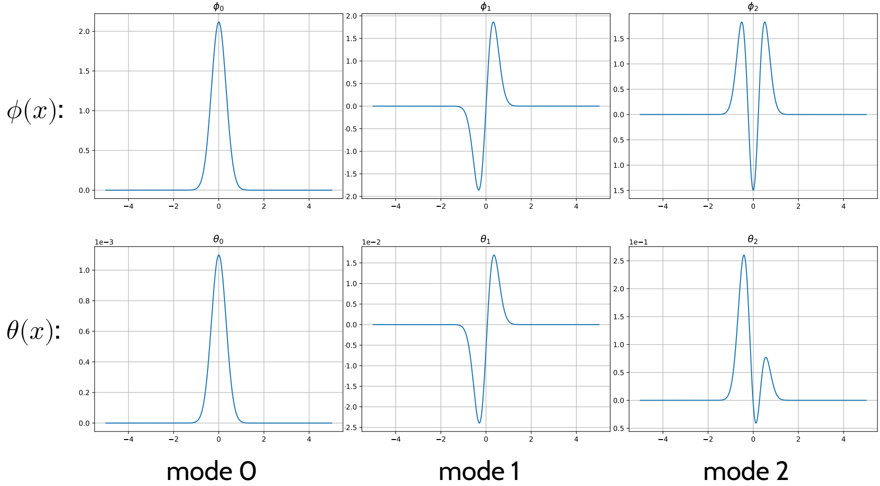
- This can be done with the QR decomposition. With $A \in \mathbb{R}^{N\times m}$

$$A = QR$$

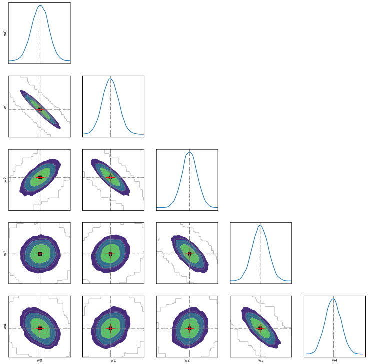
where $Q \in \mathbb{R}^{N\times m}$ is orthonormal, and $R \in \mathbb{R}^{m\times m}$ is upper triangular

- As we are seeking $A = A^\theta P^{-\top}$, we have $A^\theta \equiv Q$, and $P \equiv R^{-T}$
- This can also be done with Gram-Schmidt (GS) orthonormalization
  - Given well-known instability of GS, use *modified* GS
  - Also apply in multiple stages with repeated application of MGS
- Both implementations are in PyTUQ (github.com/sandialabs/pytuq)
  - For a general linear problem $\boldsymbol{y} = \mathcal{L}(\Phi(\boldsymbol{x})^T \boldsymbol{w})$

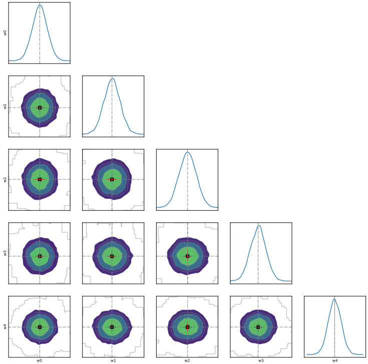
Basis functions $\phi(x)$ and $\theta(x)$ – modes 0,1,2



w Posteriors in both sets of basis functions



$\boldsymbol{w}$



$\boldsymbol{w}^\theta$

- With a diagonal prior covariance in both cases
- Thus ignoring off-diagonal terms in $\Sigma_{p,\boldsymbol{w}^\theta} = P^{-\top}\Sigma_{p,\boldsymbol{w}}P^{-1}$
  - to illustrate elimination of likelihood contribution to posterior Σ

Constrained GP

- It is often useful to apply user-defined constraints on the learned GP
- For example, if $f(x) = \boldsymbol{\phi}^T(x)\boldsymbol{w} \sim \mathcal{GP}(\boldsymbol{\mu}, \Sigma)$ is to be constrained s.t.

$$a \leq f(x) \leq b \qquad \text{or} \qquad x_i \leq x_{i+1} \Rightarrow f(x_i) \leq f(x_{i+1})$$

- Let's update the posterior $p(\boldsymbol{w}|D)$ s.t. $f(x)$ satisfies given constraints
- We will use a projection based method — Astfalck 2024
  - See also — Lin & Dunson 2014, Agrell 2019, Chakraborty & Ghosal 2021
  - Prefer projection *vs.* rejection sampling in high-dimension
    - constrained space can have low measure in original distribution
  - Rejection sampling fails entirely when the constrained space has zero measure in the original distribution
    - e.g. constrain 2D distribution onto a 1D space

Constrained PDF

Astfalck, Sen, Patra, Cripps, & Dunson (2024) arxiv:1812.05741

- Given:
  - Random vector $\theta \in \Theta \subset \mathbb{R}^n$, with $\theta \sim \pi(\theta)$, $\Sigma_\theta = \text{cov}(\theta, \theta')$
  - Constraints $C(\theta)$ define constrained space $\tilde{\Theta} := \{\theta : C(\theta)\} \subset \Theta$
  - Update $\pi(\theta)$ to satisfy the constraints

Procedure:

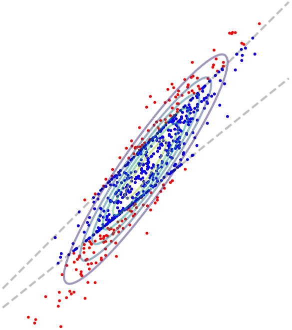
- Generate random samples $(\theta_1, \dots, \theta_N)$ from $\pi(\theta)$
- For $i = 1, \dots, N$: project θ_i onto the constrained space:

$$\begin{array}{rcl} \tilde{\theta}_i & = & \underset{\theta\in\Theta}{\text{argmin}} \|\theta_i - \theta\| \\ \text{s.t.} & & C(\theta) \end{array}$$

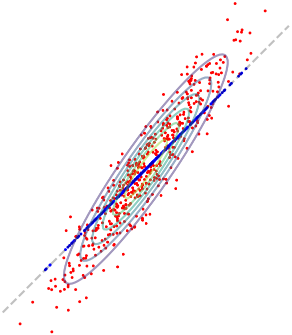
where $\|\theta - \theta'\|^2 := \|\theta - \theta'\|^2_{\Sigma_\theta} = (\theta - \theta')^T \Sigma_\theta^{-1} (\theta - \theta')$

- $\forall i$, $\tilde{\theta}_i \in \tilde{\Theta}$ are samples from the constrained density $\tilde{\pi}(\theta)$
- On $\tilde{\Theta}$, $\tilde{\pi}(\theta)$ minimizes the Wasserstein distance from $\pi(\theta)$

θ Projection examples with projection constraints



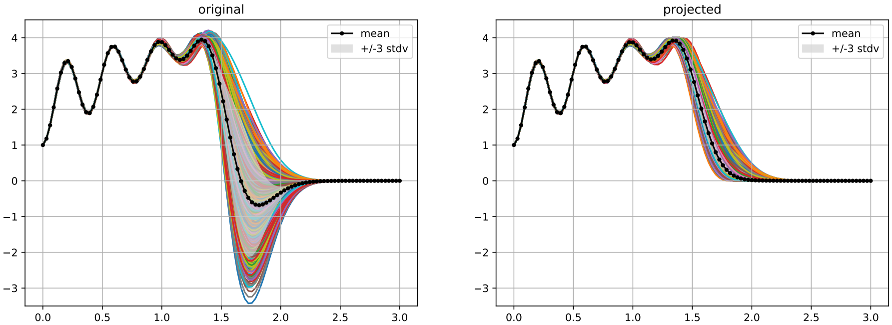
$$C(\theta):\ L_1(\theta_1) \leq \theta_2 \leq L_2(\theta_1)$$



$$C(\theta):\ \theta_2 = L(\theta_1)$$

- All samples are used, including projection onto lower dimensions
- Projected sample distribution can be mixed, with dirac delta measures

w-posterior Projection of GP samples

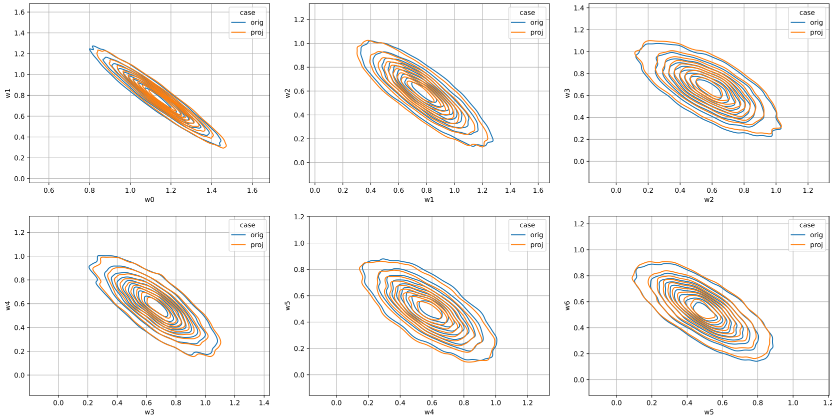


- Project w posterior samples to constrain functional to satisfy:

$$0 \leq f(u) \leq 4$$

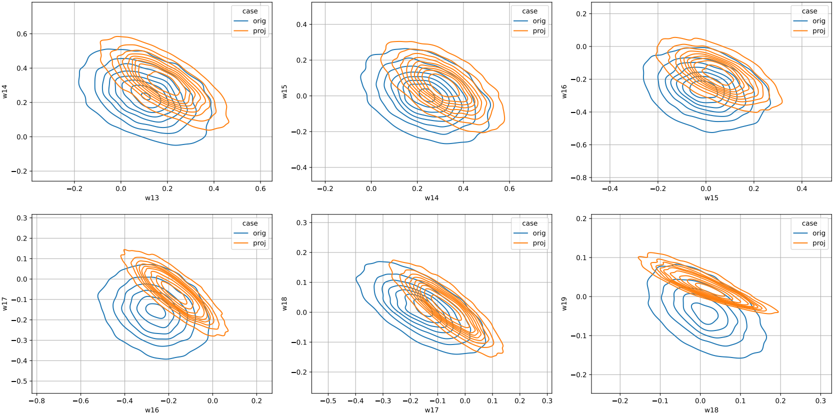
- Resulting functional samples satisfy both constraints

Original/Constrained w Posteriors – low orders



- Pair-marginal densities for low-order w terms exhibit minor change
- They remain nearly Gaussian

Original/Constrained w Posteriors – high orders



- Densities for high-order w terms exhibit more significant changes

Model Error Embedding — Kuppa *et al.* arxiv:2602.17923

- Risk of overconfident Bayesian predictions when using a wrong model
- Need to account for model error/inadequacy in Bayesian model fitting
- Modeling error in observable model predictions vs data using GPs

$$y = f(x;\lambda) + \delta_w(x) + \epsilon_d$$

Kennedy & O'Hagan '01

- Model error embedding, key relevance in physical models

$$y = f(x;\lambda,\delta_w(x)) + \epsilon_d$$

  - Typically use random variable embedding — Sargsyan '15, '19; Morrison '18, Portone '19, White '26
  - Some work on model error 2-DOF GP embedding — Fan '25

- The weight-space embedded GP provides a diagnostic tool, by estimating the missing functional dependence in the model

Challenges with Model Error GP Embedding

- Increased dimensionality, and limited data — LIS can help
- Disambiguation between model and GP Arendt '12

$$y = \begin{cases} f(x;\lambda) + \delta_w(x) + \epsilon_d \\ \tilde{f}(x;\lambda,\delta_w(x)) \quad + \epsilon_d \end{cases} = f_{\text{true}}(x) + \epsilon_d$$

- Unless otherwise constrained, a GP can dominate the data fit
  - This is a concern when parameters have physical meaning
- From an application perspective it is of interest to:
  - Let $f(x,\lambda)$ best-fit the data, and fit the GP to residual discrepancy
- This can been done using staged fitting procedures
  - LS fit: $\lambda^* = \text{argmin}_\lambda \|y - f(x;\lambda)\|$
  - LS fit: $w^* = \text{argmin}_w \|y - \tilde{f}(x;\lambda^*,\delta_w)\|$
  - Bayesian inference for λ: $y = \tilde{f}(x;\lambda,\delta_{w^*}) + \epsilon_d \;\Rightarrow\; p(\lambda|w^*,D)$
- Prefer joint Bayesian inference that minimizes ambiguity

Orthogonal Gaussian Process (OGP) – Plumlee, 2017

- Given data
$$y = f_t(x) + \epsilon$$
- With model $f(x, \lambda)$, $\lambda \in \mathbb{R}^p$, let
$$\lambda^* := \underset{\lambda}{\text{argmin}} \underbrace{\|f_t(x) - f(x, \lambda)\|_2^2}_{L_\lambda}$$
- Then, with
$$f_t(x) \approx f(x, \lambda^*) + \delta_w(x)$$
- we have, at $\lambda = \lambda^*$,
$$\nabla_\lambda \underbrace{\|f(x, \lambda^*) + \delta_w(x) - f(x, \lambda)\|_2^2}_{L_\lambda} = 0$$
- Get p linear constraints on $\delta_w(x)$, resulting in a modified GP covariance ensuring that $\delta_w(x)$ is orthogonal to $\nabla_\lambda f(x, \lambda)|_{\lambda^*}$
- Eigendecomposition of this covariance enables using the weight view

Orthogonal GP (OGP) regression – a weight space view

- Consider a linear model with an additive GP

$$\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\lambda}) + \delta_{\boldsymbol{w}}(\boldsymbol{x}) + \boldsymbol{\epsilon} = \begin{bmatrix} A & \boldsymbol{\phi}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} & \boldsymbol{w} \end{bmatrix}^\top + \boldsymbol{\epsilon}$$

then with $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{\mathsf{d}})$, $\Sigma_{\mathsf{d}} \propto I$, and the MVN prior

$$\pi(\boldsymbol{\lambda}, \boldsymbol{w}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\lambda}} \\ \boldsymbol{\mu}_{\boldsymbol{w}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\boldsymbol{\lambda}} & \boldsymbol{0} \\ \boldsymbol{0} & \Sigma_{\boldsymbol{w}} \end{bmatrix} \right)$$

we have the posterior precision matrix

$$\Sigma_{\mathsf{pos}}^{-1} = \begin{bmatrix} A^\top \Sigma_{\mathsf{d}}^{-1} A + \Sigma_{\boldsymbol{\lambda}}^{-1} & A^\top \Sigma_{\mathsf{d}}^{-1} \boldsymbol{\phi}^\top \\ \boldsymbol{\phi} \Sigma_{\mathsf{d}}^{-1} A & \boldsymbol{\phi} \Sigma_{\mathsf{d}}^{-1} \boldsymbol{\phi}^\top + \Sigma_{\boldsymbol{w}}^{-1} \end{bmatrix}$$

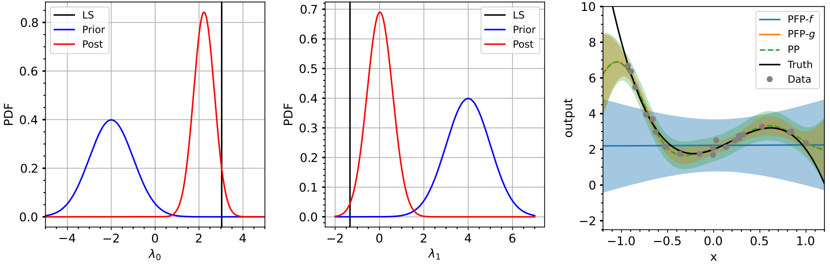
- OGP targets modification of the GP covariance such that

$$A^\top \Sigma_{\mathsf{d}}^{-1} \boldsymbol{\phi}^\top \approx \boldsymbol{0}$$

thereby minimizing posterior correlation between $\boldsymbol{\lambda}$ and $\boldsymbol{w}$

Linear Model Example – KOH with regular GP

Truth:	$f_t(x)$	$=$	$2 + 2x + 3x^2 - 5x^3$
Model:	$f(x, \lambda)$	$=$	$\lambda_0 + \lambda_1 x$
GP-augmented:	$\tilde{f}(x, \lambda, \delta_w(x))$	$=$	$f(x, \lambda) + \delta_w(x)$

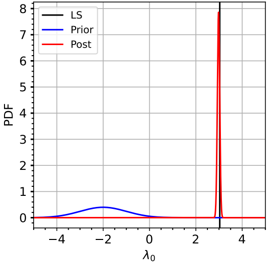


$p(\lambda_0|D)$ $p(\lambda_1|D)$ Push-forward posterior

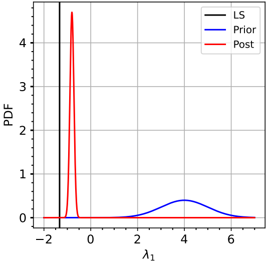
- Without orthogonalization, there is minimal learning of parameters λ
- The model on its own does not fit the data in the mean
- The GP carries the dominant weight in fitting the data

Linear Model Example – KOH with OGP

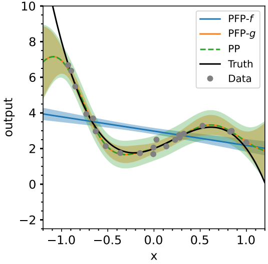
Truth:	$f_t(x)$	$=$	$2 + 2x + 3x^2 - 5x^3$
Model:	$f(x, \lambda)$	$=$	$\lambda_0 + \lambda_1 x$
GP-augmented:	$\tilde{f}(x, \lambda, \delta_w(x))$	$=$	$f(x, \lambda) + \delta_w(x)$



$p(\lambda_0|D)$



$p(\lambda_1|D)$



Push-forward posterior

- With orthogonalization, there is much improved learning of λ
- The model on its own fit the data well in the mean
- The GP does not dominate the game in fitting the data

Nonlinear ModErr OGP Embedding

GP-augmented Fit Model: $y = \tilde{f}(x, \lambda, \delta_w(x)) + \epsilon_d$

Linearization approach: (LOGP) $\Rightarrow$ Linear Bayesian solution

- Linearize $\tilde{f}|_{(\lambda^*,\mathbf{0})}$, and enforce $\nabla_\lambda L_\lambda|_{\lambda^*} = 0$, with the linearized loss

$$L_\lambda = \|f(x, \lambda^*) + \partial_{\delta_w(x)} \tilde{f}|_{(\lambda^*,\mathbf{0})} \delta_w(x) - f(x, \lambda)\|_2^2$$

Regularization approach: (ROGP) $\Rightarrow$ MCMC based solution

- Enforce $\nabla_\lambda L_\lambda|_{(\lambda^*,\mathbf{0})} = 0$ on the loss

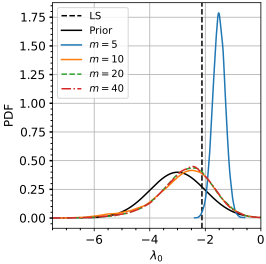
$$L_\lambda = \|\tilde{f}(x, \lambda^*, \delta_w(x)) - f(x, \lambda)\|_2^2$$

- Resulting in p nonlinear constraints: $\mathfrak{R}_k(\lambda^*, \delta_w(x)) = 0$ applied via the regularization log prior increment

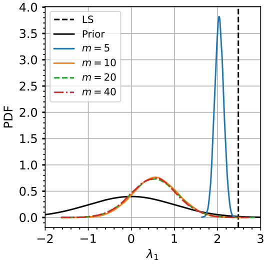
$$-\sum_{k=1}^{p} \alpha_k \mathfrak{R}_k(\lambda^*, \delta_w(x))^2$$

Nonlinear Unconstrained GP Embedding Example

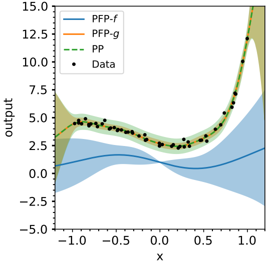
Truth:	$f_t(x)$	$=$	$e^{(1-0.5x+x^2+x^3)}$
Model:	$f(x,\lambda)$	$=$	$\sin(\lambda_0 x) + e^{\lambda_1 x}$
GP-augmented:	$\tilde{f}(x,\lambda,\delta_w(x))$	$=$	$\sin(\lambda_0 x) + e^{(\lambda_1 x + \delta_w(x))}$



$p(\lambda_0|D)$



$p(\lambda_1|D)$



PFP, $m = 40$

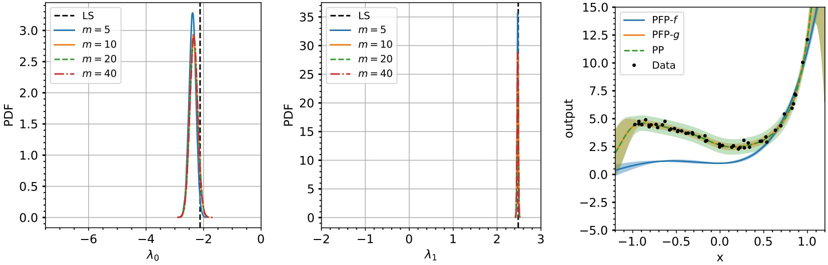
- The true model behind the data is exponential
- No additive embedding is useful in the sine function
- Examine utility of embedding in the exponential (“submodel”)
  - Same observations as in the linear case,

Nonlinear ROGP Embedding Example

Truth: $f_t(x) = e^{(1-0.5x+x^2+x^3)}$

Model: $f(x,\lambda) = \sin(\lambda_0 x) + e^{\lambda_1 x}$

GP-augmented: $\tilde{f}(x,\lambda,\delta_w(x)) = \sin(\lambda_0 x) + e^{(\lambda_1 x + \delta_w(x))}$

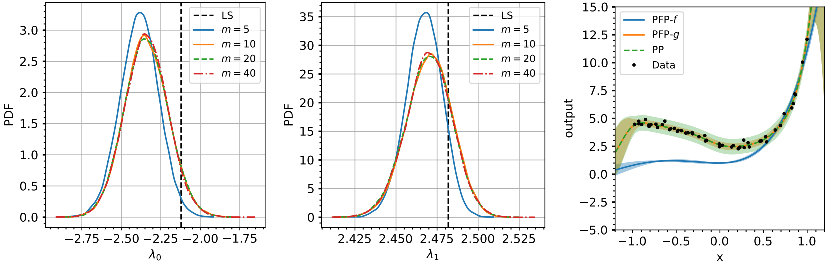


$p(\lambda_0|D)$ $p(\lambda_1|D)$ PFP, $m=40$

- With ROGP, there is improved learning of λ_0, and more so for λ_1
- The model alone: fits data better in the mean, especially at large x
- The GP does not dominate where the model can actually fit the data

Nonlinear ROGP Embedding Example

Truth:	$f_t(x)$	$=$	$e^{(1-0.5x+x^2+x^3)}$
Model:	$f(x,\lambda)$	$=$	$\sin(\lambda_0 x) + e^{\lambda_1 x}$
GP-augmented:	$\tilde{f}(x,\lambda,\delta_w(x))$	$=$	$\sin(\lambda_0 x) + e^{(\lambda_1 x+\delta_w(x))}$



$p(\lambda_0|D)$ $p(\lambda_1|D)$ PFP, $m = 40$

- With ROGP, there is improved learning of λ_0, and more so for λ_1
- The model alone: fits data better in the mean, especially at large x
- The GP does not dominate where the model can actually fit the data

Closure

- Examined utility of GPs in weight-space view for Bayesian inference
- Synthetic model problem: DFT-functional fitting
- Likelihood informed subspace facilitates handling of high-order terms
- Data-specific basis orthonormalization minimizes posterior correlation
- Projection of posterior weight samples enforces constraints on GP
- Weight-view facilitates model error GP embedding in general models
- Adapted orthogonal GP construction to the weight view, targeting model error embedding in both linear and nonlinear models
- Illustrated decoupling/disambiguation of GP and model fitting in both linear and linear examples

Select Papers

Liklihood informed subspace:

- A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk, Optimal low-rank approximations of bayesian linear inverse problems, *SIAM Journal on Scientific Computing*, 37, pp. A2451–A2487, 2015.
- T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini, Likelihood informed dimension reduction for nonlinear inverse problems, *Inverse Problems*, 30, p. 114015, 2014.

Projection constrained distributions:

- L. Astfalck, D. Sen, S. Patra, E. Cripps, and D. Dunson, *Posterior Projection for Inference in Constrained Spaces*, arXiv: 1812.05741, 2024.

Orthogonal Gaussian processes and model error:

- M. Plumlee, Bayesian calibration of inexact computer models, *Journal of the American Statistical Association*, 112, pp. 1274–1285, 2017.
- M. Kuppa, K. Sargsyan, M. Panesi, and H. N. Najm, Model Error Embedding with Orthogonal Gaussian Processes, arXiv: 2602.17923, 2026.
  - Upcoming talk by Mridula Kuppa at SIAM UQ, Minneapolis, March 22-25, 2026.