Probabilistic Numerics for Scientific Machine Learning

Jonathan Wenger

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

imprs-is

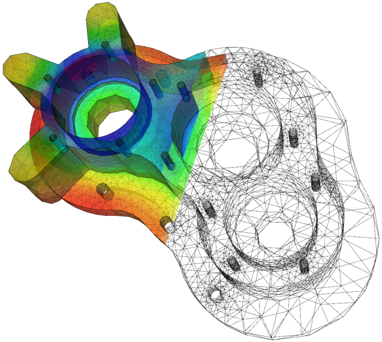
Mechanistic Models

Describing the laws of nature mathematically.

Scientific computing relies on mechanistic models.

Example: Physical processes modeled by linear PDEs

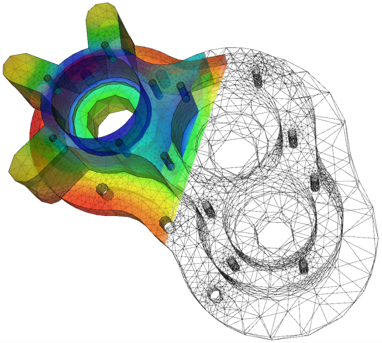
- thermal conduction (heat equation)
- electromagnetism (Maxwell's equations)
- wave mechanics (wave equation)



Mechanistic Models

Describing the laws of nature mathematically.

Scientific computing relies on mechanistic models.



©User A1, Wikimedia Commons, CC BY-SA 3.0

Strengths

- Interpretable / causal relationships
- Experimentally validated

Mechanistic Models

Describing the laws of nature mathematically.

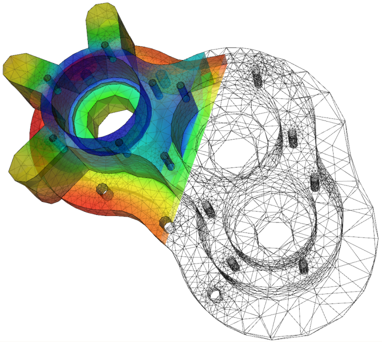
Scientific computing relies on mechanistic models.

Strengths

- Interpretable / causal relationships
- Experimentally validated

Weaknesses

- Unknown parameters
- Computationally expensive



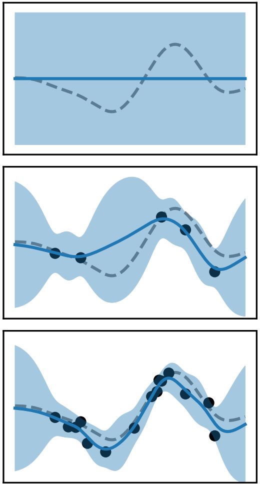
Machine Learning Models

Learning to predict from data.

Machine learning relies on statistical models.

Example: Supervised Learning

- parametric models (linear regression)
- hierarchical models (neural networks)
- probabilistic models (Gaussian processes)



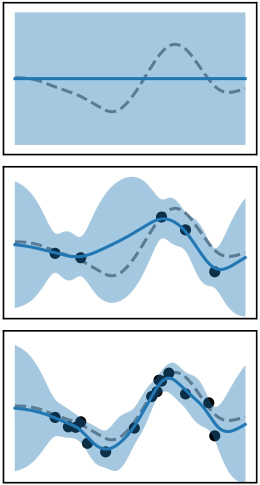
Machine Learning Models

Learning to predict from data.

Machine learning relies on statistical models.

Strengths

- Learn relationships from *unstructured* data
- Representation of uncertainty $\longrightarrow$ decision-making



Machine Learning Models

Learning to predict from data.

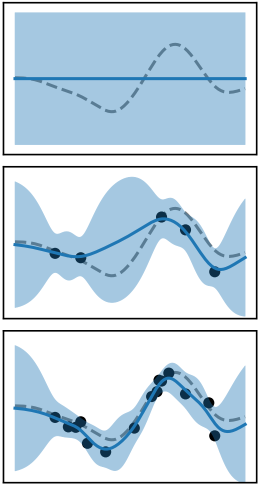
Machine learning relies on statistical models.

Strengths

- Learn relationships from *unstructured* data
- Representation of uncertainty $\longrightarrow$ decision-making

Weaknesses

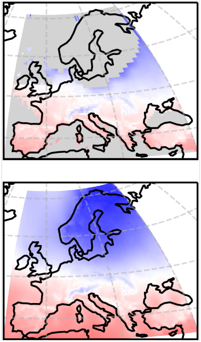
- Lack of guarantees
- Unclear or implicit assumptions



Combining Mechanistic and Statistical Models

...while retaining the benefits of both?

Modern science necessitates combining mechanistic and statistical models.



Axen et al. [2022]

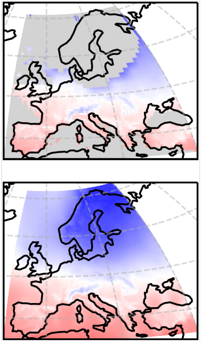
Combining Mechanistic and Statistical Models

…while retaining the benefits of both?

Modern science necessitates combining mechanistic and statistical models.

Why?

- Experiments produce volumes of unstructured, multi-modal data



Axen et al. [2022]

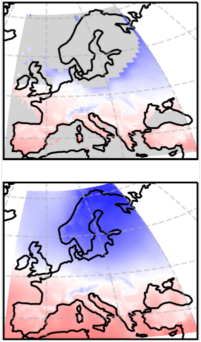
Combining Mechanistic and Statistical Models

…while retaining the benefits of both?

Modern science necessitates combining mechanistic and statistical models.

Why?

- Experiments produce volumes of unstructured, multi-modal data
- Mechanistic model parameters are only approximately known



Axen et al. [2022]

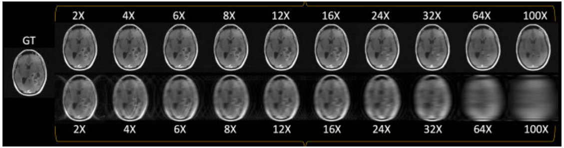
Combining Mechanistic and Statistical Models

…while retaining the benefits of both?

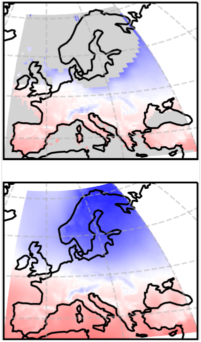
Modern science necessitates combining mechanistic and statistical models.

Why?

- Experiments produce volumes of unstructured, multi-modal data
- Mechanistic model parameters are only approximately known
- Critical decisions under uncertainty are made using scientific models



Radmanesh et al. [2022]



Axen et al. [2022]

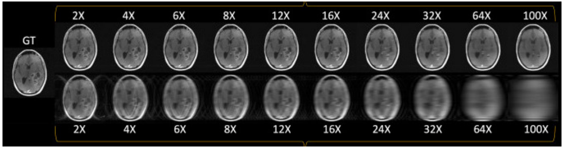
Combining Mechanistic and Statistical Models

…while retaining the benefits of both?

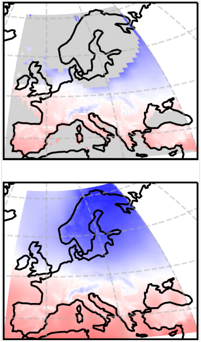
Modern science necessitates combining mechanistic and statistical models.

Why?

- Experiments produce volumes of unstructured, multi-modal data
- Mechanistic model parameters are only approximately known
- Critical decisions under uncertainty are made using scientific models



Radmanesh et al. [2022]

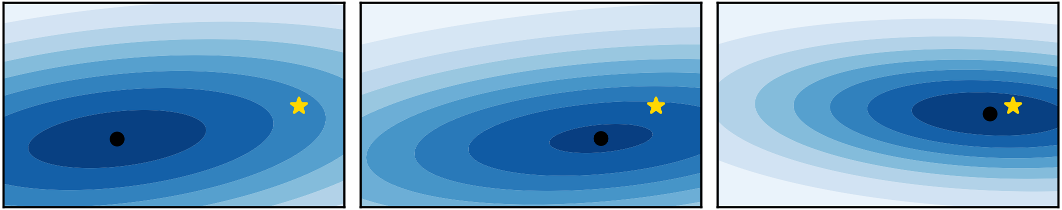


Axen et al. [2022]

Sources of error / uncertainty: Limited computation and limited data.

Core Insights

- The solution to any numerical problem is fundamentally uncertain.



★ Solution $\boldsymbol{x}_*$ ● Estimate $\boldsymbol{x}_i = \mathbb{E}(\boldsymbol{x}_*)$ ▬ Belief $p(\boldsymbol{x}_*)$

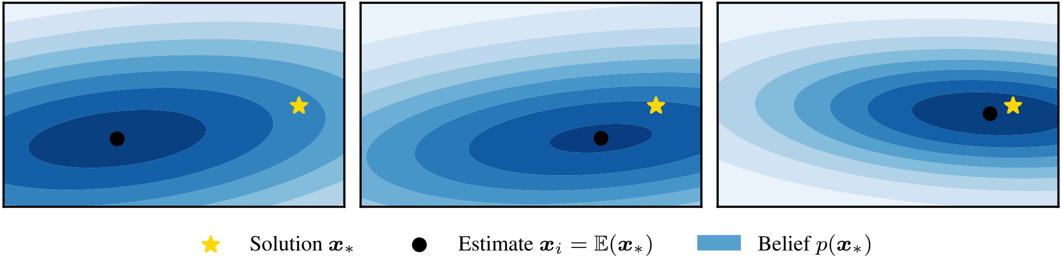
Probabilistic Numerics

Interpreting problems from numerical analysis as statistical inference.

Hennig et al. [2015], Cockayne et al. [2019]

Core Insights

- The solution to any numerical problem is fundamentally uncertain.
- Numerical algorithms are learning agents, which actively collect data and make predictions.



Linear Partial Differential Equations

Mechanistic models for thermal conduction, electromagnetism, wave mechanics, …

We look for a function $\boldsymbol{u} : \mathbb{D} \to \mathbb{R}^{d'}$ which solves the equation

$$\mathcal{D}_{\boldsymbol{\theta}}[\boldsymbol{u}] = f$$

on an open and bounded domain $\mathbb{D} \subset \mathbb{R}^d$, where $\mathcal{D}_{\boldsymbol{\theta}}$ is a linear differential operator and $f : \mathbb{D} \to \mathbb{R}$.

Typically, we require $\boldsymbol{u} \in \mathbb{U}$ and $f \in \mathbb{V}$ for Banach spaces $\mathbb{U}, \mathbb{V}$.

Linear Partial Differential Equations

Mechanistic models for thermal conduction, electromagnetism, wave mechanics, …

We look for a function $\boldsymbol{u} : \mathbb{D} \to \mathbb{R}^{d'}$ which solves the equation

$$\mathcal{D}_{\boldsymbol{\theta}}[u] = f$$

on an open and bounded domain $\mathbb{D} \subset \mathbb{R}^d$, where $\mathcal{D}_{\boldsymbol{\theta}}$ is a linear differential operator and $f : \mathbb{D} \to \mathbb{R}$.

Typically, we require $\boldsymbol{u} \in \mathbb{U}$ and $f \in \mathbb{V}$ for Banach spaces $\mathbb{U}, \mathbb{V}$.

Problems

- Usually no analytic solution $\Rightarrow$ numerical solvers necessary $\Rightarrow$ discretization error
- Parameters of the PDE (diffop parameters, right-hand side, etc.) are usually not known exactly

Physics-Informed Gaussian Process Regression

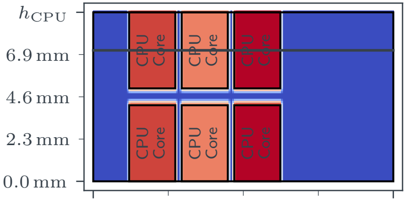
Case Study: The Heat Distribution in a CPU

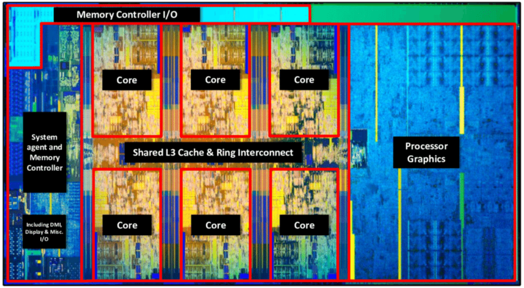
A Computer Scientist's Linear PDE

The Heat Distribution in a CPU

Spatial Domain: $\mathbb{D}_{\mathrm{CPU}} = [0, l_{\mathrm{CPU}}] \times [0, w_{\mathrm{CPU}}] \times [0, d_{\mathrm{CPU}}] \subset \mathbb{R}^3$

Spatial Domain: $\mathbb{D}_{\text{CPU,2D}} = [0, l_{\text{CPU}}] \times [0, w_{\text{CPU}}] \subset \mathbb{R}^2$



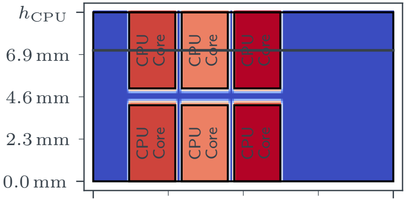


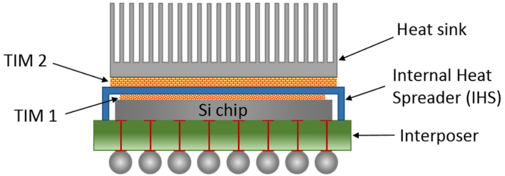
from Hebbar [2018]

A Computer Scientist's Linear PDE

The Heat Distribution in a CPU

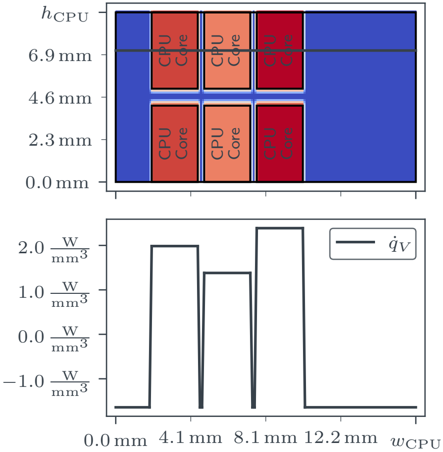
Spatial Domain: $\mathbb{D}_{\text{CPU,2D}} = [0, l_{\text{CPU}}] \times [0, w_{\text{CPU}}] \subset \mathbb{R}^2$

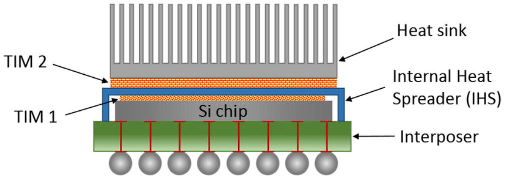




from Nylander [2018]

Spatial Domain: $\mathbb{D}_{\text{CPU,1D}} = [0, l_{\text{CPU}}] \subset \mathbb{R}$





from Nylander [2018]

Heat Equation

$$c_p \rho \frac{\partial u}{\partial t} - \kappa \Delta u = \dot{q}_V,$$

where

- $u \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ temperature
- c_p, ρ, κ material parameters
- $\dot{q}_V \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ heat source

A Computer Scientist's Linear PDE II

The Heat Distribution in a CPU

Lienhard and Lienhard [2020]

Heat Equation

$$c_p \rho \frac{\partial u}{\partial t} - \kappa \Delta u = \dot{q}_V,$$

where

- $u \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ temperature
- c_p, ρ, κ material parameters
- $\dot{q}_V \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ heat source

Stationary Heat Equation

$$-\kappa \Delta u = \dot{q}_V$$

where

- $u \colon \mathbb{D} \to \mathbb{R}$ temperature
- κ thermal conductivity
- $\dot{q}_V \colon \mathbb{D} \to \mathbb{R}$ heat source

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Lienhard and Lienhard [2020]

Heat Equation

$$c_p \rho \frac{\partial u}{\partial t} - \kappa \Delta u = \dot{q}_V,$$

where

- $u \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ temperature
- c_p, ρ, κ material parameters
- $\dot{q}_V \colon [0, T] \times \mathbb{D} \to \mathbb{R}$ heat source

Stationary Heat Equation

$$-\kappa \Delta u = \dot{q}_V$$

where

- $u \colon \mathbb{D} \to \mathbb{R}$ temperature
- κ thermal conductivity
- $\dot{q}_V \colon \mathbb{D} \to \mathbb{R}$ heat source

How can we phrase this as a learning problem?

(Linear) PDEs are Indirect Observations of Their Solution

Conservation Laws and Information Operators

Classic Notion of an Observation

$$u(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) \iff \delta_{\boldsymbol{x}_i}[u] - f(\boldsymbol{x}_i) = 0$$

- Observations are point evaluations.
- Interpret as applying evaluation functional.

(Linear) PDEs are Indirect Observations of Their Solution

Conservation Laws and Information Operators

Classic Notion of an Observation

$$u(\mathbf{x}_i) = f(\mathbf{x}_i) \iff \delta_{\mathbf{x}_i}[u] - f(\mathbf{x}_i) = 0$$

- Observations are point evaluations.
- Interpret as applying evaluation functional.

Generalized "Observation"

$$-\kappa \Delta u = \dot{q}_V \iff \mathcal{D}[u] - f = 0$$

- Heat equation is a conservation law.
- A conservation law is an observation of the behavior of the system u!

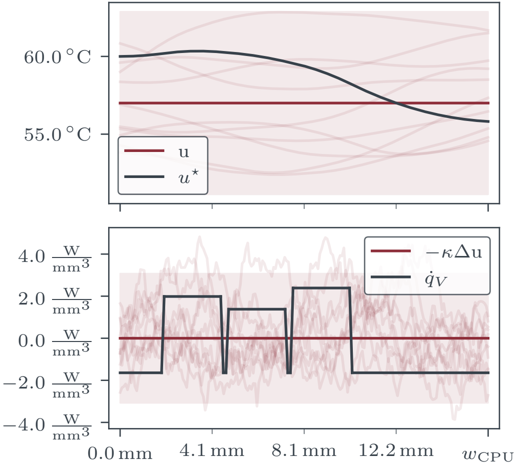
Idea: Relax notion of an observation to an *information operator* $\mathcal{I}[\mathbf{u}] := \mathcal{D}[\mathbf{u}] - f = 0$.

Prior

$$u \sim \mathcal{GP}(m, k)$$

Observations / Information Operators

$$\mathcal{I}_{\text{PDE}}[u] := -\kappa \Delta u(X_{\text{PDE}}) - \dot{q}_V(X_{\text{PDE}}) = \mathbf{0}$$



Prior

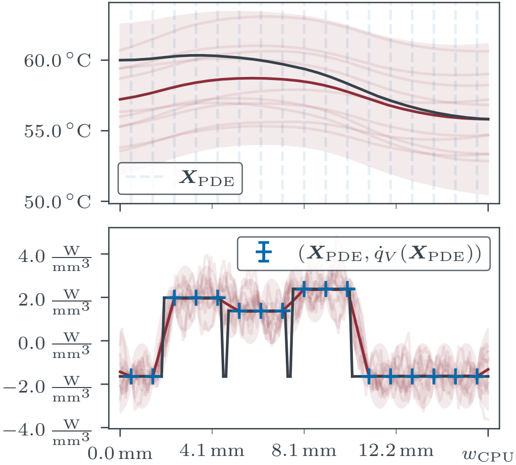
$$\mathrm{u} \sim \mathcal{GP}(m, k)$$

Observations / Information Operators

$$\mathcal{I}_{\mathrm{PDE}}[\mathrm{u}] := -\kappa \Delta \mathrm{u}(X_{\mathrm{PDE}}) - \dot{q}_V(X_{\mathrm{PDE}}) = \mathbf{0}$$

Posterior

$$\left(\mathrm{u} \mid \mathcal{I}_{\mathrm{PDE}}[\mathrm{u}] = \mathbf{0}\right) \Big| \mathcal{I}_{\mathrm{DBC}}[\mathrm{u}] = \mathbf{0} \sim \mathcal{GP}$$



Prior

$$\mathsf{u} \sim \mathcal{GP}(m, k)$$

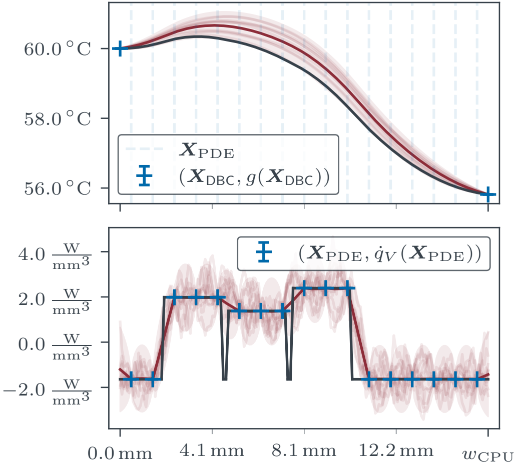
Observations / Information Operators

$$\mathcal{I}_{\mathsf{PDE}}[\mathsf{u}] := -\kappa \Delta \mathsf{u}(X_{\mathsf{PDE}}) - \dot{q}_V(X_{\mathsf{PDE}}) = \mathbf{0}$$

$$\mathcal{I}_{\mathsf{DBC}}[\mathsf{u}] := \mathsf{u}(X_{\mathsf{BC}}) - u^\star(X_{\mathsf{BC}}) = \mathbf{0}$$

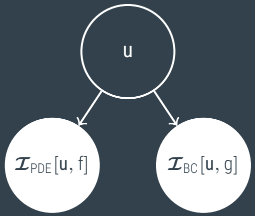
Posterior

$$\mathsf{u} \mid \mathcal{I}_{\mathsf{PDE}}[\mathsf{u}] = \mathbf{0} \sim \mathcal{GP}$$



We have seen that GP inference can produce

- an approximate solution of the BVP, and
- an estimate of the approximation error.

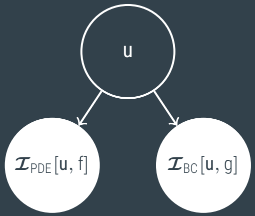


We have seen that GP inference can produce

- an approximate solution of the BVP, and
- an estimate of the approximation error.

Unfortunately,

- the boundary values are unknown in deployment, and
- the values of the heat source distribution are uncertain.

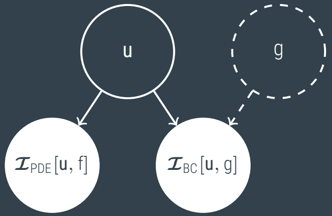


We have seen that GP inference can produce

- an approximate solution of the BVP, and
- an estimate of the approximation error.

Unfortunately,

- the boundary values are unknown in deployment, and
- the values of the heat source distribution are uncertain.

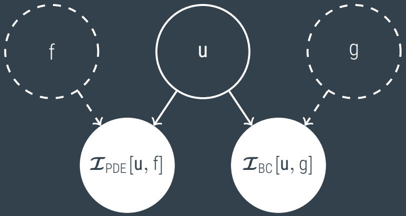


We have seen that GP inference can produce

- an approximate solution of the BVP, and
- an estimate of the approximation error.

Unfortunately,

- the boundary values are unknown in deployment, and
- the values of the heat source distribution are uncertain.



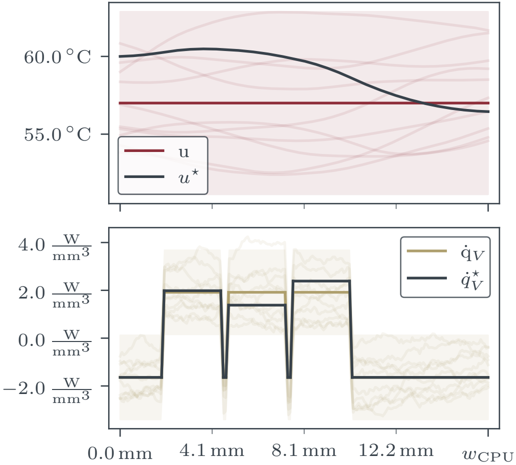
Epistemic Parameter Uncertainty and Measured Data

Uncertain Right-Hand Side

Prior

$$u \sim \mathcal{GP}(m_u, k_u)$$
$$\dot{q}_V \sim \mathcal{GP}(m_{\dot{q}_V}, k_{\dot{q}_V})$$

Information Operators



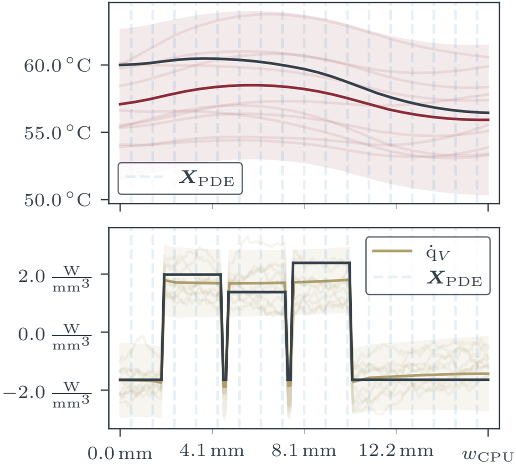
Prior

$$\mathrm{u} \sim \mathcal{GP}(m_{\mathrm{u}}, k_{\mathrm{u}})$$

$$\dot{q}_V \sim \mathcal{GP}(m_{\dot{q}_V}, k_{\dot{q}_V})$$

Information Operators

$$\mathcal{I}_{\mathrm{PDE}}[\mathrm{u}, \dot{q}_V] = -\kappa \Delta \mathrm{u}(X_{\mathrm{PDE}}) - \dot{q}_V(X_{\mathrm{PDE}}) = \mathbf{0}$$



Epistemic Parameter Uncertainty and Measured Data

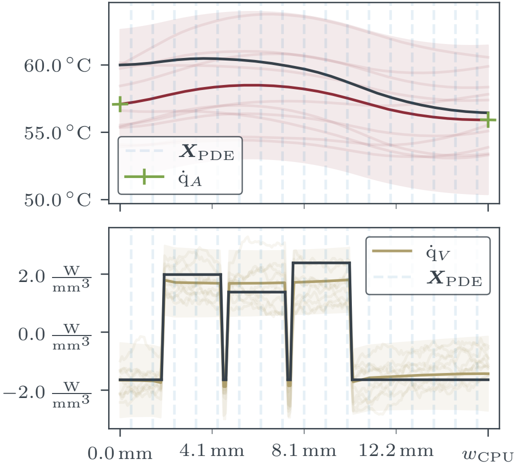
Uncertain Neumann Boundary Conditions

Prior

$$\mathrm{u} \sim \mathcal{GP}\left(m_{\mathrm{u}}, k_{\mathrm{u}}\right)$$
$$\dot{\mathrm{q}}_V \sim \mathcal{GP}\left(m_{\dot{\mathrm{q}}_V}, k_{\dot{\mathrm{q}}_V}\right)$$
$$\dot{\mathrm{q}}_A \sim \mathcal{GP}\left(m_{\dot{\mathrm{q}}_A}, k_{\dot{\mathrm{q}}_A}\right)$$

Information Operators

$$\mathcal{I}_{\mathrm{PDE}}[\mathrm{u}, \dot{\mathrm{q}}_V] = -\kappa \Delta \mathrm{u}\left(X_{\mathrm{PDE}}\right) - \dot{\mathrm{q}}_V\left(X_{\mathrm{PDE}}\right) = \mathbf{0}$$
$$\mathcal{I}_{\mathrm{NBC}}[\mathrm{u}, \dot{\mathrm{q}}_A] = -\kappa \partial_{\nu(X_{\mathrm{BC}})} \mathrm{u}\left(X_{\mathrm{BC}}\right) - \dot{\mathrm{q}}_A\left(X_{\mathrm{BC}}\right) = \mathbf{0}$$



Epistemic Parameter Uncertainty and Measured Data

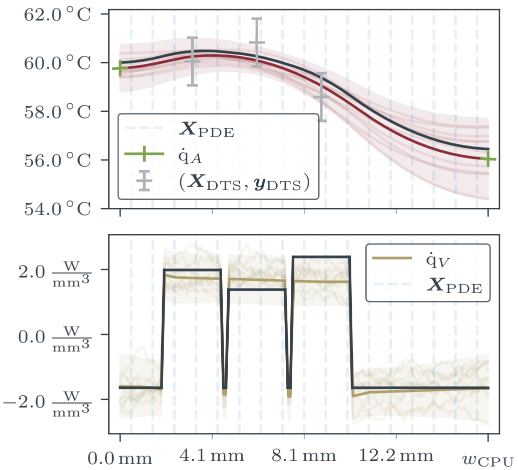
Noisy Sensor Data

Prior

$$u \sim \mathcal{GP}(m_u, k_u)$$
$$\dot{q}_V \sim \mathcal{GP}(m_{\dot{q}_V}, k_{\dot{q}_V})$$
$$\dot{q}_A \sim \mathcal{GP}(m_{\dot{q}_A}, k_{\dot{q}_A})$$
$$\epsilon_{\mathrm{DTS}} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathrm{DTS}})$$

Information Operators

$$\mathcal{I}_{\mathrm{PDE}}[u, \dot{q}_V] = -\kappa \Delta u(X_{\mathrm{PDE}}) - \dot{q}_V(X_{\mathrm{PDE}}) = \mathbf{0}$$
$$\mathcal{I}_{\mathrm{NBC}}[u, \dot{q}_A] = -\kappa \partial_{\nu(X_{\mathrm{BC}})} u(X_{\mathrm{BC}}) - \dot{q}_A(X_{\mathrm{BC}}) = \mathbf{0}$$
$$\mathcal{I}_{\mathrm{DTS}}[u, \epsilon_{\mathrm{DTS}}] = u(X_{\mathrm{DTS}}) + \epsilon_{\mathrm{DTS}} = y_{\mathrm{DTS}}$$



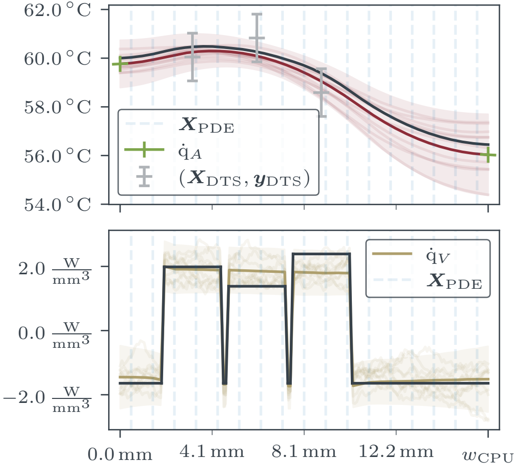
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Prior

$$\mathrm{u} \sim \mathcal{GP}(m_{\mathrm{u}}, k_{\mathrm{u}})$$
$$\dot{\mathrm{q}}_V \sim \mathcal{GP}(m_{\dot{\mathrm{q}}_V}, k_{\dot{\mathrm{q}}_V})$$
$$\dot{\mathrm{q}}_A \sim \mathcal{GP}(m_{\dot{\mathrm{q}}_A}, k_{\dot{\mathrm{q}}_A})$$
$$\epsilon_{\mathrm{DTS}} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathrm{DTS}})$$

Information Operators

$$\mathcal{I}_{\mathrm{PDE}}[\mathrm{u}, \dot{\mathrm{q}}_V] = -\kappa \Delta \mathrm{u}(X_{\mathrm{PDE}}) - \dot{\mathrm{q}}_V(X_{\mathrm{PDE}}) = \mathbf{0}$$
$$\mathcal{I}_{\mathrm{NBC}}[\mathrm{u}, \dot{\mathrm{q}}_A] = -\kappa \partial_{\nu(X_{\mathrm{BC}})} \mathrm{u}(X_{\mathrm{BC}}) - \dot{\mathrm{q}}_A(X_{\mathrm{BC}}) = \mathbf{0}$$
$$\mathcal{I}_{\mathrm{DTS}}[\mathrm{u}, \boldsymbol{\epsilon}_{\mathrm{DTS}}] = \mathrm{u}(X_{\mathrm{DTS}}) + \boldsymbol{\epsilon}_{\mathrm{DTS}} = y_{\mathrm{DTS}}$$
$$\mathcal{I}_{\mathrm{STAT}}[\dot{\mathrm{q}}_V, \dot{\mathrm{q}}_A] = d_{\mathrm{CPU}} \int_{\mathbb{D}} \dot{\mathrm{q}}_V \, dx - \int_{\partial \mathbb{D}} \dot{\mathrm{q}}_A \, dS = 0$$

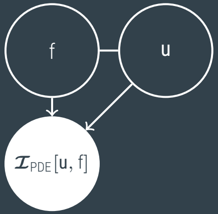


The GP approach integrates enables seamless integration of

- prior knowledge about the solution,

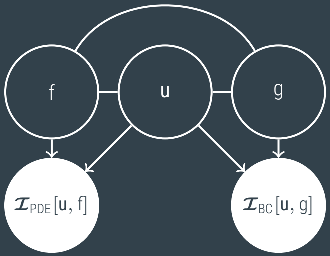
The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,



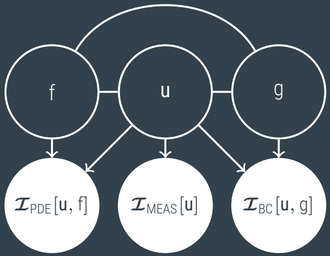
The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and



The GP approach integrates enables seamless integration of

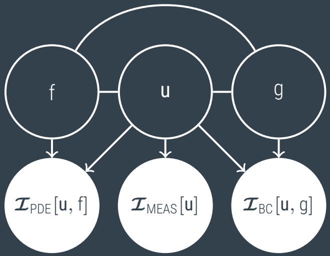
- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and
- noisy empirical measurements,



The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and
- noisy empirical measurements,

all while providing

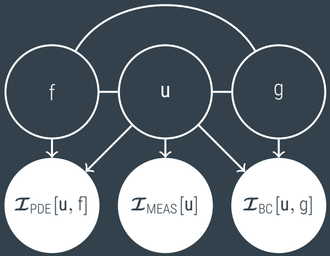


The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and
- noisy empirical measurements,

all while providing

- quantification of approximation error,

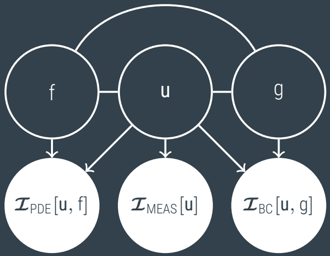


The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and
- noisy empirical measurements,

all while providing

- quantification of approximation error,
- error propagation from uncertain system parameters, and

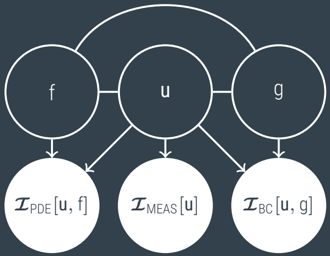


The GP approach integrates enables seamless integration of

- prior knowledge about the solution,
- mechanistic knowledge in the form of linear PDEs with uncertain right-hand sides,
- uncertain boundary conditions, and
- noisy empirical measurements,

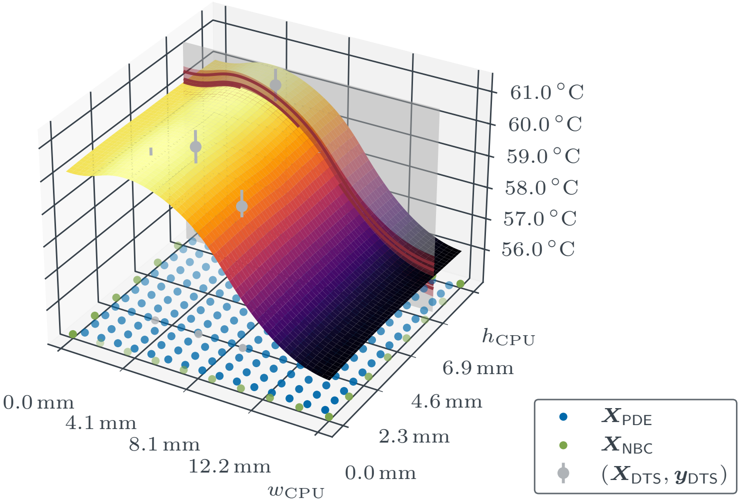
all while providing

- quantification of approximation error,
- error propagation from uncertain system parameters, and
- a Bayesian solution to the inverse problem of estimating the right-hand side and boundary function from data.



All this is only possible because we give up on trying to identify a single unique solution in favor of a probability measure over infinitely many solution candidates.

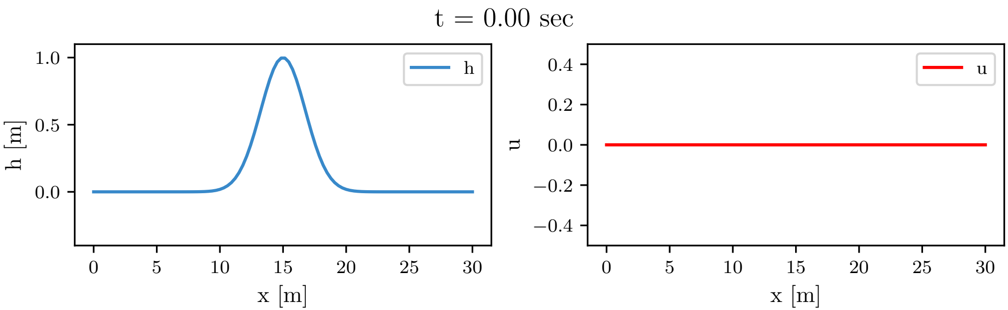
2D Version of the CPU Simulation



Example: Systems of Time-Dependent PDEs

The (Linearized) Shallow Water / Saint-Venant Equations

Credit: Tim Weiland



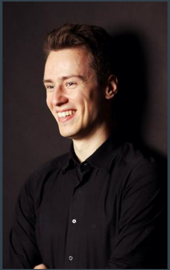
Physics-Informed Gaussian Process Regression Generalizes Linear PDE Solvers

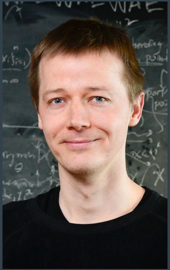
Marvin Pförtner, Ingo Steinwart, Philipp Hennig, Jonathan Wenger

- PDEs can be solved via GP inference $\Rightarrow$ structured uncertainty
- GPs provide a rigorous framework for probabilistic inference of unknown functions from heterogeneous information sources provided by affine information operators
- A vast class of classical PDE solvers (methods of weighted residuals) can be recovered in the mean of a GP posterior
- Proof of GP inference theorem with bounded linear operator observations in separable Banach path spaces

Paper / 2212.12474

Code / marvinpfoertner / linpde-gp





Generalized Gaussian Process Inference

A well-known conjecture...

Prior $\mathrm{u} \sim \mathcal{GP}(m, k)$ with paths in $\mathbb{U} \subset \mathbb{R}^{\mathbb{D}}$

Generalized Gaussian Process Inference

A well-known conjecture...

Prior $\mathrm{u} \sim \mathcal{GP}(m, k)$ with paths in $\mathbb{U} \subset \mathbb{R}^{\mathbb{D}}$

Observations $\mathbf{y} = \mathcal{L}[\mathrm{u}] + \boldsymbol{\epsilon}$, where

- $\mathcal{L} : \mathbb{U} \to \mathbb{R}^n$ linear (e.g. $\mathcal{L}_i = \mathcal{D}[\cdot](\boldsymbol{x}_i)$)
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\epsilon} \perp\!\!\!\perp \mathrm{u}$

Generalized Gaussian Process Inference

A well-known conjecture...

Prior $\mathrm{u} \sim \mathcal{GP}(m, k)$ with paths in $\mathbb{U} \subset \mathbb{R}^{\mathbb{D}}$

Observations $\mathbf{y} = \boldsymbol{\mathcal{L}}[\mathrm{u}] + \boldsymbol{\epsilon}$, where

- $\boldsymbol{\mathcal{L}} : \mathbb{U} \to \mathbb{R}^n$ linear (e.g. $\mathcal{L}_i = \mathcal{D}[\cdot](\boldsymbol{x}_i)$)
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\epsilon} \perp\!\!\!\perp \mathrm{u}$

Predictive $\boldsymbol{\mathcal{L}}[\mathrm{u}] + \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{\mathcal{L}}[m], \boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}' + \boldsymbol{\Sigma}\right)$, where

$$(\boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}')_{ij} := \mathcal{L}_i[x \mapsto \mathcal{L}_j[k(x, \cdot)]]$$

Generalized Gaussian Process Inference

A well-known conjecture...

Prior $\mathrm{u} \sim \mathcal{GP}(m, k)$ with paths in $\mathbb{U} \subset \mathbb{R}^{\mathbb{D}}$

Observations $\boldsymbol{y} = \boldsymbol{\mathcal{L}}[\mathrm{u}] + \boldsymbol{\epsilon}$, where

- $\boldsymbol{\mathcal{L}} : \mathbb{U} \to \mathbb{R}^n$ linear (e.g. $\mathcal{L}_i = \mathcal{D}[\cdot](\boldsymbol{x}_i)$)
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\epsilon} \perp\!\!\!\perp \mathrm{u}$

Predictive $\boldsymbol{\mathcal{L}}[\mathrm{u}] + \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{\mathcal{L}}[m], \boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}' + \boldsymbol{\Sigma}\right)$, where

$$(\boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}')_{ij} := \mathcal{L}_i[x \mapsto \mathcal{L}_j[k(x, \cdot)]]$$

Posterior $\mathrm{u} \mid \boldsymbol{\mathcal{L}}[\mathrm{u}] + \boldsymbol{\epsilon} = \boldsymbol{y} \sim \mathcal{GP}\left(m^{\mathrm{u}|\boldsymbol{y}}, k^{\mathrm{u}|\boldsymbol{y}}\right)$, where

$$m^{\mathrm{u}|\boldsymbol{y}}(\boldsymbol{x}) := m(\boldsymbol{x}) + \boldsymbol{\mathcal{L}}[k(\cdot, \boldsymbol{x})]^\top \left(\boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}' + \boldsymbol{\Sigma}\right)^\dagger (\boldsymbol{y} - \boldsymbol{\mathcal{L}}[m])$$

$$k^{\mathrm{u}|\boldsymbol{y}}(\boldsymbol{x}_1, \boldsymbol{x}_2) := k(\boldsymbol{x}_1, \boldsymbol{x}_2) - \boldsymbol{\mathcal{L}}[k(\cdot, \boldsymbol{x}_1)]^\top \left(\boldsymbol{\mathcal{L}} k \boldsymbol{\mathcal{L}}' + \boldsymbol{\Sigma}\right)^\dagger \boldsymbol{\mathcal{L}}[k(\cdot, \boldsymbol{x}_2)]$$

Connections to Classical Methods

MWR Information Operators

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\mathrm{PDE}}[\mathrm{u}, \mathrm{f}] = \mathcal{D}[\mathrm{u}](X_{\mathrm{PDE}}) - \mathrm{f}(X_{\mathrm{PDE}})$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathrm{u}) \subset \mathbb{U}$ and paths $(\mathrm{f}) \subset \mathbb{V}$ (or continuously embedded)
- $\mathcal{D} \colon \mathbb{U} \to \mathbb{V}$ linear and bounded

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}_i^{\text{PDE}}[\mathbf{u}, \mathrm{f}] = \delta_{\boldsymbol{x}_{\text{PDE}}^{(i)}} \left[\mathcal{D}\left[\quad \mathbf{u}\right] - \mathrm{f}\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathbf{u}) \subset \mathbb{U}$ and paths $(\mathrm{f}) \subset \mathbb{V}$ (or continuously embedded)
- $\mathcal{D} \colon \mathbb{U} \to \mathbb{V}$ linear and bounded

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\text{PDE}}_{\ell^{(i)},\text{id}_{\mathbb{U}}}[\mathsf{u}, \mathsf{f}] = \ell^{(i)}\left[\mathcal{D}\left[\quad \mathsf{u}\right] - \mathsf{f}\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathsf{u}) \subset \mathbb{U}$ and paths $(\mathsf{f}) \subset \mathbb{V}$ (or continuously embedded)
- $\mathcal{D} \colon \mathbb{U} \to \mathbb{V}$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}'$

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\mathrm{PDE}}_{\ell^{(i)},\mathcal{P}_{\hat{\mathbb{U}}}}[\mathsf{u}, \mathsf{f}] = \ell^{(i)}\left[\mathcal{D}\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathsf{u}]\right] - \mathsf{f}\right]$$

- $\mathbb{U}$, $\mathbb{V}$ (separable) Banach spaces
- paths $(\mathsf{u}) \subset \mathbb{U}$ and paths $(\mathsf{f}) \subset \mathbb{V}$ (or continuously embedded)
- $\mathcal{D} \colon \mathbb{U} \to \mathbb{V}$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}'$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} \colon \mathbb{U} \to \mathbb{U}$ bounded projection with $\operatorname{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\text{PDE}}_{\ell^{(i)}, \mathcal{P}_{\hat{\mathbb{U}}}}[\mathbf{u}, \mathrm{f}^{W}] = \ell^{(i)}\left[\mathcal{D}^{W}\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathbf{u}]\right] - \mathrm{f}^{W}\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathbf{u}) \subset \mathbb{U}$ and paths $(\mathrm{f}^{W}) \subset \mathbb{V}'$ (or continuously embedded)
- $\mathcal{D}^{W} : \mathbb{U} \to \mathbb{V}'$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}''$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} : \mathbb{U} \to \mathbb{U}$ bounded projection with $\mathrm{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$
- applicable to weak formulations

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\text{PDE}}_{\ell^{(i)}, \mathcal{P}_{\hat{\mathbb{U}}}}[\mathsf{u}, \mathsf{f}^W] = \ell^{(i)}\left[\mathcal{D}^W\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathsf{u}]\right] - \mathsf{f}^W\right]$$

- $\mathbb{U}$, $\mathbb{V}$ (separable) Banach spaces
- paths $(\mathsf{u}) \subset \mathbb{U}$ and paths $(\mathsf{f}^W) \subset \mathbb{V}'$ (or continuously embedded)
- $\mathcal{D}^W : \mathbb{U} \to \mathbb{V}'$ linear and bounded
- test functionals: $\ell^{(1)}, \ldots, \ell^{(n)} \in \mathbb{V}''$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} : \mathbb{U} \to \mathbb{U}$ bounded projection with $\operatorname{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$
- applicable to weak formulations

Example (Ritz-Galerkin Method)

- $\mathbb{U} = H^1(\mathbb{D})$, $\mathbb{V} = H^1_0(\mathbb{D})$
- $\mathcal{D}^W[u](v) = \int_{\mathbb{D}} \langle \kappa \nabla u, \nabla v \rangle \, \mathrm{d}x$
- $\mathsf{f}^W[v] = \langle \mathsf{f}, v \rangle_{L_2}$, where $\mathsf{f} \in L_2(\mathbb{D})$

$$\mathcal{I}^{\mathrm{PDE}}_{\ell^{(i)}, \mathcal{P}_{\hat{\mathbb{U}}}}[\mathbf{u}, \mathrm{f}^{W}] = \ell^{(i)}\left[\mathcal{D}^{W}\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathbf{u}]\right] - \mathrm{f}^{W}\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathbf{u}) \subset \mathbb{U}$ and paths $(\mathrm{f}^{W}) \subset \mathbb{V}'$ (or continuously embedded)
- $\mathcal{D}^{W} : \mathbb{U} \to \mathbb{V}'$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}''$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} : \mathbb{U} \to \mathbb{U}$ bounded projection with $\mathrm{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$
- applicable to weak formulations

Example (Ritz-Galerkin Method)

- $\mathbb{U} = H^{1}(\mathbb{D}), \mathbb{V} = H^{1}_{0}(\mathbb{D})$
- $\mathcal{D}^{W}[u](v) = \int_{\mathbb{D}} \langle \kappa \nabla u, \nabla v \rangle \, dx$
- $\mathrm{f}^{W}[v] = \langle \mathrm{f}, v \rangle_{L_2}$, where $\mathrm{f} \in L_2(\mathbb{D})$
- $\ell^{(i)}$ induced by test functions $\psi^{(i)} \in \mathbb{V} \hookrightarrow \mathbb{V}''$

  $\Rightarrow \ell^{(i)}[\mathcal{D}^{W}[u]] = \mathcal{D}^{W}[u](\psi^{(i)})$

  $\Rightarrow \ell^{(i)}[\mathrm{f}^{W}] = \langle \mathrm{f}, \psi^{(i)} \rangle_{L_2}$

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\mathrm{PDE}}_{\ell^{(i)}, \mathcal{P}_{\hat{\mathbb{U}}}}[\mathsf{u}, \mathsf{f}^W] = \ell^{(i)}\left[\mathcal{D}^W\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathsf{u}]\right] - \mathsf{f}^W\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathsf{u}) \subset \mathbb{U}$ and paths $(\mathsf{f}^W) \subset \mathbb{V}'$ (or continuously embedded)
- $\mathcal{D}^W : \mathbb{U} \to \mathbb{V}'$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}''$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} : \mathbb{U} \to \mathbb{U}$ bounded projection with $\mathrm{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$
- applicable to weak formulations

Example (Ritz-Galerkin Method)

- $\mathbb{U} = H^1(\mathbb{D}), \mathbb{V} = H^1_0(\mathbb{D})$
- $\mathcal{D}^W[u](v) = \int_{\mathbb{D}} \langle \kappa \nabla u, \nabla v \rangle \,\mathrm{d}x$
- $\mathsf{f}^W[v] = \langle \mathsf{f}, v \rangle_{L_2}$, where $\mathsf{f} \in L_2(\mathbb{D})$
- $\ell^{(i)}$ induced by test functions $\psi^{(i)} \in \mathbb{V} \hookrightarrow \mathbb{V}''$
- $\hat{\mathbb{U}} = \mathrm{span}\left(\phi^{(1)}, \dots, \phi^{(n)}\right)$ with trial functions $\phi^{(i)} = \psi^{(i)}$

MWR Information Operators

From Dirac to Galerkin

[Pförtner et al., 2022, Section 3.3]

$$\mathcal{I}^{\text{PDE}}_{\ell^{(i)},\mathcal{P}_{\hat{\mathbb{U}}}}[\mathsf{u}, \mathsf{f}^W] = \ell^{(i)}\left[\mathcal{D}^W\left[\mathcal{P}_{\hat{\mathbb{U}}}[\mathsf{u}]\right] - \mathsf{f}^W\right]$$

- $\mathbb{U}, \mathbb{V}$ (separable) Banach spaces
- paths $(\mathsf{u}) \subset \mathbb{U}$ and paths $(\mathsf{f}^W) \subset \mathbb{V}'$ (or continuously embedded)
- $\mathcal{D}^W : \mathbb{U} \to \mathbb{V}'$ linear and bounded
- test functionals: $\ell^{(1)}, \dots, \ell^{(n)} \in \mathbb{V}''$
- trial projection: $\mathcal{P}_{\hat{\mathbb{U}}} : \mathbb{U} \to \mathbb{U}$ bounded projection with $\operatorname{ran}(\mathcal{P}_{\hat{\mathbb{U}}}) = \hat{\mathbb{U}} \subset \mathbb{U}$
- applicable to weak formulations

Example (Ritz-Galerkin Method)

- $\mathbb{U} = H^1(\mathbb{D}), \mathbb{V} = H^1_0(\mathbb{D})$
- $\mathcal{D}^W[u](v) = \int_{\mathbb{D}} \langle \kappa \nabla u, \nabla v \rangle \, \mathrm{d}x$
- $\mathsf{f}^W[v] = \langle \mathsf{f}, v \rangle_{L_2}$, where $\mathsf{f} \in L_2(\mathbb{D})$
- $\ell^{(i)}$ induced by test functions $\psi^{(i)} \in \mathbb{V} \hookrightarrow \mathbb{V}''$
- $\hat{\mathbb{U}} = \operatorname{span}\left(\phi^{(1)}, \dots, \phi^{(n)}\right)$ with trial functions $\phi^{(i)} = \psi^{(i)}$
- choose $\mathcal{P}_{\hat{\mathbb{U}}}$ e.g. as L_2 projection onto $\hat{\mathbb{U}}$

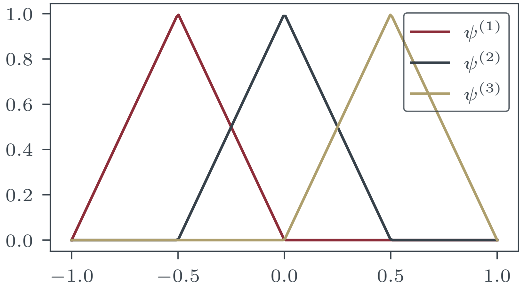
$$\mathcal{P}_{\hat{\mathbb{U}}}[\mathsf{u}] = \sum_{i=1}^{n} \phi^{(i)} \sum_{j=1}^{n} (P^{-1})_{ij} \langle \phi^{(j)}, \mathsf{u} \rangle_{L_2},$$

where $P_{ij} = \langle \phi^{(i)}, \phi^{(j)} \rangle_{L_2}$

Example: The Finite Element Method for the 1D Poisson Equation

A Ritz-Galerkin Method with Locally Supported Trial Functions

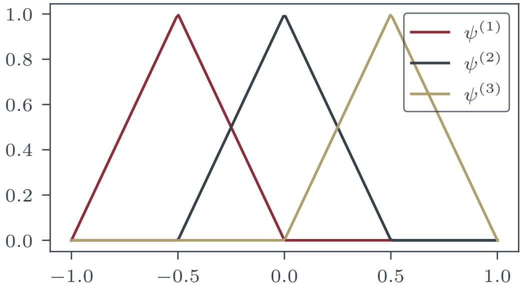
Test Functions: Linear Lagrange Elements



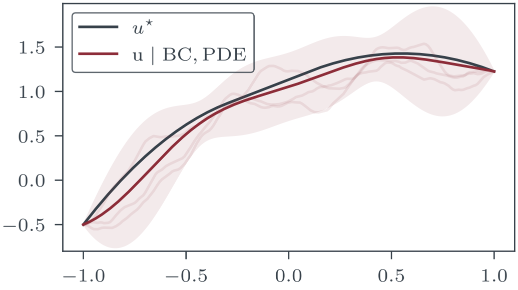
Example: The Finite Element Method for the 1D Poisson Equation

A Ritz-Galerkin Method with Locally Supported Trial Functions

Test Functions: Linear Lagrange Elements

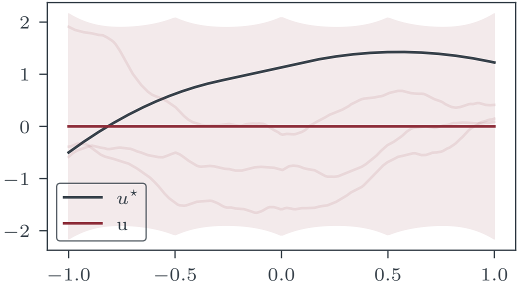


GP Posterior



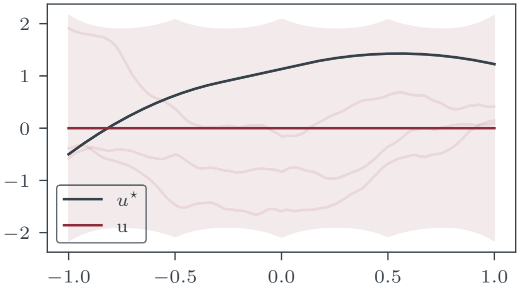
Matérn-$\frac{3}{2}$ Prior Covariance $\Rightarrow$ paths $(\mathrm{u}) \subset H^1(\mathbb{D})$

MWR Recovery Prior



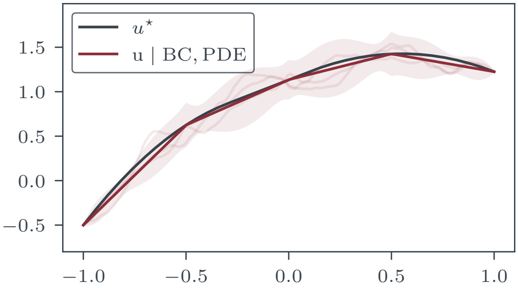
$$
\begin{aligned}
k^{\mathrm{MWR}} &:= \mathcal{P}_{\hat{\mathbb{U}}} k \mathcal{P}'_{\hat{\mathbb{U}}} + \mathcal{P}_{\ker(\mathcal{P}_{\hat{\mathbb{U}}})} k \mathcal{P}'_{\ker(\mathcal{P}_{\hat{\mathbb{U}}})} \\
&= k - \mathcal{P}_{\hat{\mathbb{U}}} k - k \mathcal{P}'_{\hat{\mathbb{U}}} + 2 \mathcal{P}_{\hat{\mathbb{U}}} k \mathcal{P}'_{\hat{\mathbb{U}}}
\end{aligned}
$$

MWR Recovery Prior



$$
\begin{aligned}
k^{\text{MWR}} &:= \mathcal{P}_{\hat{\mathbb{U}}} k \mathcal{P}'_{\hat{\mathbb{U}}} + \mathcal{P}_{\ker(\mathcal{P}_{\hat{\mathbb{U}}})} k \mathcal{P}'_{\ker(\mathcal{P}_{\hat{\mathbb{U}}})} \\
&= k - \mathcal{P}_{\hat{\mathbb{U}}} k - k \mathcal{P}'_{\hat{\mathbb{U}}} + 2 \mathcal{P}_{\hat{\mathbb{U}}} k \mathcal{P}'_{\hat{\mathbb{U}}}
\end{aligned}
$$

Posterior



Connections to Classical Methods

Weighted Residual Methods Through the Lens of GP Inference

[Pförtner et al., 2022, Section 3.3]

- we show that all weighted residual methods [Fletcher, 1984] can be realized as posterior means corresponding to an MWR recovery prior

Connections to Classical Methods

Weighted Residual Methods Through the Lens of GP Inference

[Pförtner et al., 2022, Section 3.3]

- we show that all weighted residual methods [Fletcher, 1984] can be realized as posterior means corresponding to an MWR recovery prior
  - parametric and nonparametric collocation methods
  - finite-volume methods
  - pseudospectral methods
  - (Petrov-)Galerkin methods
    - finite-element methods
    - spectral methods

Connections to Classical Methods

Weighted Residual Methods Through the Lens of GP Inference

[Pförtner et al., 2022, Section 3.3]

- we show that all weighted residual methods [Fletcher, 1984] can be realized as posterior means corresponding to an MWR recovery prior
  - parametric and nonparametric collocation methods
  - finite-volume methods
  - pseudospectral methods
  - (Petrov-)Galerkin methods
    - finite-element methods
    - spectral methods
- the remaining uncertainty lies in the kernel of the trial projection $\mathcal{P}_{\hat{\mathbb{U}}}$ $\Rightarrow$ probabilistic Galerkin orthogonality

Connections to Classical Methods

Weighted Residual Methods Through the Lens of GP Inference

[Pförtner et al., 2022, Section 3.3]

- we show that all weighted residual methods [Fletcher, 1984] can be realized as posterior means corresponding to an MWR recovery prior
  - parametric and nonparametric collocation methods
  - finite-volume methods
  - pseudospectral methods
  - (Petrov-)Galerkin methods
    - finite-element methods
    - spectral methods
- the remaining uncertainty lies in the kernel of the trial projection $\mathcal{P}_{\hat{\mathbb{U}}} \Rightarrow$ probabilistic Galerkin orthogonality

$\Rightarrow$ GP-based approaches as uncertainty-aware drop-in replacements for classical methods

Theoretical Backbone

Gaussian Process Regression with Linear Operator Observations

Definition (Gaussian Process)

A Gaussian process is a family of random variables $\{\omega \mapsto \mathrm{f}(x, \omega)\}_{x \in \mathbb{X}}$ on a common Borel probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P})$ such that every finite combination $f(x_1, \cdot), \dots, f(x_n, \cdot)$ of the random variables follows a multivariate normal distribution.

Gaussian Process Regression with Linear Operator Observations I

Why all the fuss?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Definition (Gaussian Process)

A Gaussian process is a family of random variables $\{\omega \mapsto \mathrm{f}(x, \omega)\}_{x \in \mathbb{X}}$ on a common Borel probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P})$ such that every finite combination $f(x_1, \cdot), \ldots, f(x_n, \cdot)$ of the random variables follows a multivariate normal distribution.

- for a generic GP, we can only reason about finitely many evaluations

Gaussian Process Regression with Linear Operator Observations I

Why all the fuss?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Definition (Gaussian Process)

A Gaussian process is a family of random variables $\{\omega \mapsto \mathrm{f}(x, \omega)\}_{x \in \mathbb{X}}$ on a common Borel probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P})$ such that every finite combination $f(x_1, \cdot), \dots, f(x_n, \cdot)$ of the random variables follows a multivariate normal distribution.

- for a generic GP, we can only reason about finitely many evaluations
- some of the observation operators we care about (partial derivatives and integrals) implicitly operate on an infinite set of evaluations

Gaussian Process Regression with Linear Operator Observations I

Why all the fuss?

Definition (Gaussian Process)

A Gaussian process is a family of random variables $\{\omega \mapsto \mathrm{f}(\boldsymbol{x}, \omega)\}_{\boldsymbol{x} \in \mathbb{X}}$ on a common Borel probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P})$ such that every finite combination $f(\boldsymbol{x}_1, \cdot), \dots, f(\boldsymbol{x}_n, \cdot)$ of the random variables follows a multivariate normal distribution.

- for a generic GP, we can only reason about finitely many evaluations
- some of the observation operators we care about (partial derivatives and integrals) implicitly operate on an infinite set of evaluations
- to apply existing results, we need $\omega \mapsto (\mathrm{f}(\boldsymbol{X}, \omega), \boldsymbol{\mathcal{L}}[\mathrm{f}(\cdot, \omega)])$ to be a Gaussian random variable, but it is unclear if this is even measurable

Gaussian Process Regression with Linear Operator Observations I

Why all the fuss?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Definition (Gaussian Process)

A Gaussian process is a family of random variables $\{\omega \mapsto \mathrm{f}(\boldsymbol{x}, \omega)\}_{\boldsymbol{x} \in \mathbb{X}}$ on a common Borel probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P})$ such that every finite combination $f(\boldsymbol{x}_1, \cdot), \dots, f(\boldsymbol{x}_n, \cdot)$ of the random variables follows a multivariate normal distribution.

- for a generic GP, we can only reason about finitely many evaluations
- some of the observation operators we care about (partial derivatives and integrals) implicitly operate on an infinite set of evaluations
- to apply existing results, we need $\omega \mapsto (\mathrm{f}(\boldsymbol{X}, \omega), \boldsymbol{\mathcal{L}}[\mathrm{f}(\cdot, \omega)])$ to be a Gaussian random variable, but it is unclear if this is even measurable
- theoretical results should be easily applicable to GPs specified via their mean and covariance functions (as opposed to projections of Gaussian measures in functions spaces)

Gaussian Process Regression with Linear Operator Observations II

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Theorem (Pförtner et al. 2022, Theorem 1)

Let $\mathrm{f} \sim \mathcal{GP}(m, k)$ *be a Gaussian process prior with index set* $\mathbb{X}$ *on the probability space* $(\Omega, \mathcal{F}, \mathrm{P})$*, whose paths lie in a real separable reproducing kernel Banach space (RKBS)* $\mathbb{B} \subset \mathbb{R}^{\mathbb{X}}$ *such that* $\omega \mapsto \mathrm{f}(\cdot, \omega)$ *is a* $\mathbb{B}$*-valued Gaussian random variable. Let* $\mathcal{L} \colon \mathbb{B} \to \mathbb{R}^n$ *be a bounded linear operator. Then*

$$\mathcal{L}[\mathrm{f}] \sim \mathcal{N}\left(\mathcal{L}[m], \mathcal{L}k\mathcal{L}'\right).$$

Let $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *be an* $\mathbb{R}^n$*-valued Gaussian random vector with* $\boldsymbol{\epsilon} \perp\!\!\!\perp \mathrm{f}$*. Then, for any* $\boldsymbol{y} \in \mathbb{R}^n$*,*

$$\mathrm{f} \mid \mathcal{L}[\mathrm{f}] + \boldsymbol{\epsilon} = \boldsymbol{y} \sim \mathcal{GP}\left(m^{\mathrm{f}|\boldsymbol{y}}, k^{\mathrm{f}|\boldsymbol{y}}\right),$$

with conditional mean and covariance function given by

$$m^{\mathrm{f}|\boldsymbol{y}}(\boldsymbol{x}) = m(\boldsymbol{x}) + \mathcal{L}[k(\boldsymbol{x}, \cdot)]^\top \left(\mathcal{L}k\mathcal{L}' + \boldsymbol{\Sigma}\right)^\dagger (\boldsymbol{y} - (\mathcal{L}[m] + \boldsymbol{\mu})), \quad \text{and}$$

$$k^{\mathrm{f}|\boldsymbol{y}}(\boldsymbol{x}_1, \boldsymbol{x}_2) = k(\boldsymbol{x}_1, \boldsymbol{x}_2) - \mathcal{L}[k(\boldsymbol{x}_1, \cdot)]^\top \left(\mathcal{L}k\mathcal{L}' + \boldsymbol{\Sigma}\right)^\dagger \mathcal{L}[k(\cdot, \boldsymbol{x}_2)].$$

- We show that the assumptions of the theorem are fulfilled for Gaussian processes with
  - paths in any separable reproducing kernel Hilbert space $\mathbb{H}$
    $\Rightarrow \mathbb{B} = \mathbb{H}$
    $\Rightarrow$ Sobolev spaces [see Steinwart, 2019, Kanagawa et al., 2018]

Gaussian Process Regression with Linear Operator Observations III

On Prior Selection

[Pförtner et al., 2022, Sections B.2 and B.4]

- We show that the assumptions of the theorem are fulfilled for Gaussian processes with
  - paths in any separable reproducing kernel Hilbert space $\mathbb{H}$
    $\Rightarrow \mathbb{B} = \mathbb{H}$
    $\Rightarrow$ Sobolev spaces [see Steinwart, 2019, Kanagawa et al., 2018]
  - continuous paths on a compact domain $\Rightarrow \mathbb{B} = C(\mathbb{D})$

- We show that the assumptions of the theorem are fulfilled for Gaussian processes with
  - paths in any separable reproducing kernel Hilbert space $\mathbb{H}$
    $\Rightarrow \mathbb{B} = \mathbb{H}$
    $\Rightarrow$ Sobolev spaces [see Steinwart, 2019, Kanagawa et al., 2018]
  - continuous paths on a compact domain $\Rightarrow \mathbb{B} = C(\mathbb{D})$
  - continuously differentiable paths with bounded and uniformly continuous partial derivatives
    $\Rightarrow \mathbb{B} = C^k(\overline{\mathbb{D}})$ or $\mathbb{B} = C^{\boldsymbol{\alpha}}(\overline{\mathbb{D}})$
    $\Rightarrow$ Hölder spaces

- We show that the assumptions of the theorem are fulfilled for Gaussian processes with
  - paths in any separable reproducing kernel Hilbert space $\mathbb{H}$
    $\Rightarrow \mathbb{B} = \mathbb{H}$
    $\Rightarrow$ Sobolev spaces [see Steinwart, 2019, Kanagawa et al., 2018]
  - continuous paths on a compact domain $\Rightarrow \mathbb{B} = C(\mathbb{D})$
  - continuously differentiable paths with bounded and uniformly continuous partial derivatives
    $\Rightarrow \mathbb{B} = C^k(\overline{\mathbb{D}})$ or $\mathbb{B} = C^{\boldsymbol{\alpha}}(\overline{\mathbb{D}})$
    $\Rightarrow$ Hölder spaces
- in these spaces, the most important observation operators (point evaluated partial derivatives and integrals) are bounded

- We show that the assumptions of the theorem are fulfilled for Gaussian processes with
  - paths in any separable reproducing kernel Hilbert space $\mathbb{H}$
    $\Rightarrow \mathbb{B} = \mathbb{H}$
    $\Rightarrow$ Sobolev spaces [see Steinwart, 2019, Kanagawa et al., 2018]
  - continuous paths on a compact domain $\Rightarrow \mathbb{B} = C(\mathbb{D})$
  - continuously differentiable paths with bounded and uniformly continuous partial derivatives
    $\Rightarrow \mathbb{B} = C^k(\overline{\mathbb{D}})$ or $\mathbb{B} = C^{\boldsymbol{\alpha}}(\overline{\mathbb{D}})$
    $\Rightarrow$ Hölder spaces
- in these spaces, the most important observation operators (point evaluated partial derivatives and integrals) are bounded
- path properties can be verified from properties of the covariance function [see e.g. Adler and Taylor, 2007]

- a GP whose covariance function is a tensor product of 1D Matérn-$(p_i + \frac{1}{2})$ kernels has paths in $\mathbb{B} = C^{(p_1, \dots, p_d)}(\overline{\mathbb{D}})$ [Wang et al., 2021]

- a GP whose covariance function is a tensor product of 1D Matérn-$(p_i + \frac{1}{2})$ kernels has paths in $\mathbb{B} = C^{(p_1, \dots, p_d)}(\overline{\mathbb{D}})$ [Wang et al., 2021]
- a GP with Gaussian covariance function has smooth paths (i.e. $\mathbb{B} = C^k(\overline{\mathbb{D}})$ for any $k \geq 0$)

- a GP whose covariance function is a tensor product of 1D Matérn-($p_i + \frac{1}{2}$) kernels has paths in $\mathbb{B} = C^{(p_1, \dots, p_d)}(\overline{\mathbb{D}})$ [Wang et al., 2021]
- a GP with Gaussian covariance function has smooth paths (i.e. $\mathbb{B} = C^k(\overline{\mathbb{D}})$ for any $k \geq 0$)
- a GP with Matérn-($p + \frac{1}{2}$) covariance function has paths in an RKHS which is norm-equivalent to the Sobolev space $H^p(\mathbb{D})$ (under mild assumptions on the domain $\mathbb{D}$, see Kanagawa et al. 2018)

References I

Robert J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer, New York, NY, first edition, 2007. doi:10.1007/978-0-387-48116-6.

Seth D. Axen, Alexandra Gessner, Christian Sommer, Nils Weitzel, and Álvaro Tejero-Cantero. Spatiotemporal modeling of European paleoclimate using doubly sparse Gaussian processes. In *Workshop on Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems (NeurIPS)*, 2022. doi:10.48550/arXiv.2211.08160.

Jon Cockayne, Chris Oates, Tim Sullivan, and Mark Girolami. Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. In Geert Verdoolaege, editor, *Proceedings of the 36th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1853 of *AIP Conference Proceedings*, pages 060001–1 – 060001–8, 2017. doi:10.1063/1.4985359.

Jon Cockayne, Chris J. Oates, T. J. Sullivan, and Mark Girolami. Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789, 2019. doi:10.1137/17M1139357.

C. A. J. Fletcher. *Computational Galerkin Methods*. Scientific Computation. Springer, Berlin, Heidelberg, first edition, 1984. doi:10.1007/978-3-642-85949-6.

References II

Ranjan Hebbar. *SPEC CPU2017: PERFORMANCE, ENERGY AND EVENT CHARACTERIZATION ON MODERN PROCESSORS*. PhD thesis, The University of Alabama in Huntsville, 06 2018.

Philipp Hennig, Michael A. Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A*, 471(2179), 2015. doi:10.1098/rspa.2015.0142.

Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K. Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

John H. Lienhard, IV and John H. Lienhard, V. *A Heat Transfer Textbook*. Phlogiston Press, Cambridge, MA, fifth edition, 2020.

Andreas Nylander. *Fabrication and Characterisation of Carbon Nanotube Array Thermal Interface Materials*. PhD thesis, Chalmers University of Technology, 11 2018.

Marvin Pförtner, Ingo Steinwart, Philipp Hennig, and Jonathan Wenger. Physics-informed Gaussian process regression generalizes linear PDE solvers, 2022. URL `https://arxiv.org/abs/2212.12474`.

References III

Alireza Radmanesh, Matthew J Muckley, Tullie Murrell, Emma Lindsey, Anuroop Sriram, Florian Knoll, Daniel K Sodickson, and Yvonne W Lui. Exploring the acceleration limits of deep learning variational network–based two-dimensional brain mri. *Radiology: Artificial Intelligence*, 4(6):e210313, 2022.

Ingo Steinwart. Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *Potential Analysis*, 51:361–395, 2019. doi:10.1007/s11118-018-9715-5.

Junyang Wang, Jon Cockayne, Oksana Chkrebtii, Tim J. Sullivan, and Chris J. Oates. Bayesian numerical methods for nonlinear partial differential equations. *Statistics and Computing*, 31(55), 2021. doi:10.1007/s11222-021-10030-w.