

Solving and Learning Nonlinear PDEs with Gaussian Processes

Yifan Chen, Caltech

Joint work with

Bamdad Hosseini, Houman Owhadi and Andrew Stuart

SIAM Uncertainty Quantification, 2022

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Numerical Approximation and Inference

- **Partial Differential Equations:** infinite degrees of freedom (DOF)

$$\mathcal{F}(x, t, u, \partial_t u, \nabla_x u, \nabla_x^2 u, a, \xi, \dots) = 0$$

- Stationary PDEs, dynamical systems, inverse problems, UQ, ...
- **Numerical Approximation** (finite DOF) designed by experts
 - Finite difference/element/volume
 - Spectral methods
 - Boundary integral methods
 - Meshless methods, collocation methods
 - Multiscale methods, numerical homogenization
- **Inference and ML** to automate the finite \leftrightarrow infinite process
 - Physics informed ML (Deep Ritz methods, PINNs, SDEs...)
 - Operator learning techniques (Neural Operators, DeepONets...)
 - Bayes probabilistic numerics, Gaussian processes and kernel methods
 - ...

This talk^{*}

Our Goal

A general GP framework for solving and learning nonlinear PDEs

- Interpretable, convergent and amenable to numerical analysis^{*}
- Near-linear time and space complexity implementation^{*}
- Hierarchical parameter learning in the GP, or kernel learning

¹Yifan Chen, Houman Owhadi, and Andrew Stuart. “Consistency of empirical Bayes and kernel flow for hierarchical parameter estimation”. In: *Mathematics of Computation* (2021).

²Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M Stuart. “Solving and learning nonlinear pdes with gaussian processes”. In: *Journal of Computational Physics* (2021).

³Yifan Chen, Florian Schaefer, and Houman Owhadi. “Sparse Cholesky Factorization for Solving Nonlinear PDEs via Gaussian Processes”. In preparation.

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

A Nonlinear Elliptic PDE Example

- Consider the stationary elliptic PDE

$$\begin{cases} -\Delta u(\mathbf{x}) + u(\mathbf{x})^3 = f(\mathbf{x}), & \forall \mathbf{x} \in \Omega, \\ u(\mathbf{x}) = g(\mathbf{x}), & \forall \mathbf{x} \in \partial\Omega. \end{cases}$$

- Domain $\Omega \subset \mathbb{R}^d$.
- PDE data $f, g : \Omega \rightarrow \mathbb{R}$.
- PDE has a unique strong/classical solution u^* .

A Nonlinear Elliptic PDE: The Methodology

- 1 Choose a kernel $K : \overline{\Omega} \times \overline{\Omega} \rightarrow \mathbb{R}$
 - Corresponding RKHS \mathcal{U} with norm $\|\cdot\|$

- 2 Choose some collocation points
 - $X^{\text{int}} = \{\mathbf{x}_1^{\text{int}}, \dots, \mathbf{x}_{M^{\text{int}}}^{\text{int}}\} \subset \Omega$
 - $X^{\text{bd}} = \{\mathbf{x}_1^{\text{bd}}, \dots, \mathbf{x}_{M^{\text{bd}}}^{\text{bd}}\} \subset \partial\Omega$

- 3 Solve the optimization problem

$$\begin{cases} \text{minimize}_{u \in \mathcal{U}} \|u\| \\ \text{s.t.} & -\Delta u(\mathbf{x}_m) + u(\mathbf{x}_m)^3 = f(\mathbf{x}_m), \quad \text{for } \mathbf{x}_m \in X^{\text{int}} \\ & u(\mathbf{x}_n) = g(\mathbf{x}_n), \quad \text{for } \mathbf{x}_n \in X^{\text{bd}} \end{cases}$$

Bayes Inference Interpretation of the Methodology

- 1 Choose a kernel $K : \overline{\Omega} \times \overline{\Omega} \rightarrow \mathbb{R}$ (Choose the prior $\mathcal{GP}(0, K)$)
 - Corresponding RKHS \mathcal{U} with norm $\|\cdot\|$
- 2 Choose some collocation points (Choose the data/likelihood)
 - $X^{\text{int}} = \{\mathbf{x}_1^{\text{int}}, \dots, \mathbf{x}_{M^{\text{int}}}^{\text{int}}\} \subset \Omega$
 - $X^{\text{bd}} = \{\mathbf{x}_1^{\text{bd}}, \dots, \mathbf{x}_{M^{\text{bd}}}^{\text{bd}}\} \subset \partial\Omega$
- 3 Solve the optimization problem (Find the “MAP”)

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} & \|u\| \\ \text{s.t.} & -\Delta u(\mathbf{x}_m) + u(\mathbf{x}_m)^3 = f(\mathbf{x}_m), \quad \text{for } \mathbf{x}_m \in X^{\text{int}} \\ & u(\mathbf{x}_n) = g(\mathbf{x}_n), \quad \text{for } \mathbf{x}_n \in X^{\text{bd}} \end{cases}$$

⁴Jon Cockayne, Chris J Oates, Timothy John Sullivan, and Mark Girolami. “Bayesian probabilistic numerical methods”. In: *SIAM Review* 61.4 (2019), pp. 756–789.

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - **Representer Theorem**
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Representation of the Minimizer

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} \quad \|u\| \\ \text{s.t.} \quad -\Delta u(\mathbf{x}_m) + u(\mathbf{x}_m)^3 = f(\mathbf{x}_m), & \text{for } \mathbf{x}_m \in X^{\text{int}} \\ \quad \quad \quad u(\mathbf{x}_n) = g(\mathbf{x}_n), & \text{for } \mathbf{x}_n \in X^{\text{bd}} \end{cases}$$

$$\Updownarrow (N = M^{\text{bd}} + 2M^{\text{int}})$$

$$\begin{cases} \underset{\mathbf{z}=(\mathbf{z}^{\text{bd}}, \mathbf{z}^{\text{int}}, \mathbf{z}_{\Delta}^{\text{int}}) \in \mathbb{R}^N}{\text{minimize}} \quad \begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} \quad \|u\| \\ \text{s.t.} \quad u(X^{\text{bd}}) = \mathbf{z}^{\text{bd}} \\ \quad \quad \quad u(X^{\text{int}}) = \mathbf{z}^{\text{int}} \\ \quad \quad \quad \Delta u(X^{\text{int}}) = \mathbf{z}_{\Delta}^{\text{int}} \end{cases} \\ \text{s.t.} \quad -\mathbf{z}^{\text{int}} + (\mathbf{z}_{\Delta}^{\text{int}})^3 = f(X^{\text{int}}) \\ \quad \quad \mathbf{z}^{\text{bd}} = g(X^{\text{bd}}) \end{cases}$$

Inner optimization

$$\begin{aligned} & \underset{u \in \mathcal{U}}{\text{minimize}} \quad \|u\| \\ & \text{s.t.} \quad u(X^{\text{bd}}) = \mathbf{z}^{\text{bd}}, u(X^{\text{int}}) = \mathbf{z}^{\text{int}}, \Delta u(X^{\text{int}}) = \mathbf{z}_{\Delta}^{\text{int}} \end{aligned}$$

- Bounded linear functionals: $\delta_{\mathbf{x}_m} \circ \Delta, \delta_{\mathbf{x}_m}, \delta_{\mathbf{x}_n} \in \mathcal{U}^*$
- Vector $\phi := (\delta_{X^{\text{bd}}}, \delta_{X^{\text{int}}}, \delta_{X^{\text{int}}} \circ \Delta) \in (\mathcal{U}^*)^{\otimes N}$
- Kernel vector and matrix:

$$\begin{aligned} K(\mathbf{x}, \phi)_j &:= \phi_j^{\mathbf{x}'} K(\mathbf{x}, \cdot) \\ K(\phi, \phi)_{i,j} &:= \phi_i^{\mathbf{x}} (\phi_j^{\mathbf{x}'} K(\cdot, \cdot)) \end{aligned}$$

Entry examples: $K(\mathbf{x}_i, \mathbf{x}_j), \Delta^{\mathbf{x}} K(\mathbf{x}_i, \mathbf{x}_j), \Delta^{\mathbf{x}} \Delta^{\mathbf{x}'} K(\mathbf{x}_i, \mathbf{x}_j)$

- Minimizer $u(\mathbf{x}) = K(\mathbf{x}, \phi) K(\phi, \phi)^{-1} \mathbf{z}$

Representation of the Minimizer

Combine the two level optimization:

Representer theorem

Every minimizer u^\dagger can be represented as

$$u^\dagger(\mathbf{x}) = K(\mathbf{x}, \phi) K(\phi, \phi)^{-1} \mathbf{z}^\dagger,$$

where the vector $\mathbf{z}^\dagger \in \mathbb{R}^N$ is a minimizer of

$$\begin{cases} \min_{\mathbf{z} \in \mathbb{R}^N} & \mathbf{z}^T K(\phi, \phi)^{-1} \mathbf{z} \\ \text{s.t.} & F(\mathbf{z}) = \mathbf{y}. \end{cases}$$

- Function $F : \mathbb{R}^N \rightarrow \mathbb{R}^M$ depends on PDE collocation constraints
- \mathbf{y} contains PDE boundary and RHS data

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Towards A Practical Algorithm

Quadratic optimization with nonlinear constraints

- A linearization algorithm $\mathbf{z}^k \rightarrow \mathbf{z}^{k+1}$

$$\begin{cases} \min_{\mathbf{z} \in \mathbb{R}^N} & \mathbf{z}^T K(\phi, \phi)^{-1} \mathbf{z} \\ \text{s.t.} & F(\mathbf{z}^k) + F'(\mathbf{z}^k)(\mathbf{z} - \mathbf{z}^k) = \mathbf{y}. \end{cases}$$

“Newton’s iteration for the nonlinear PDE”

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Numerical Experiments: Stationary Problems

■ Nonlinear Elliptic Equation

$$\begin{cases} -\Delta u(\mathbf{x}) + u(\mathbf{x})^3 = f(\mathbf{x}), & \forall \mathbf{x} \in \Omega, \\ u(\mathbf{x}) = g(\mathbf{x}), & \forall \mathbf{x} \in \partial\Omega. \end{cases}$$

■ Truth: $d = 2$, $u^*(\mathbf{x}) = \sin(\pi x_1) \sin(\pi x_2) + 4 \sin(4\pi x_1) \sin(4\pi x_2)$

■ Kernel: $K(\mathbf{x}, \mathbf{y}; \sigma) = \exp(-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma^2})$

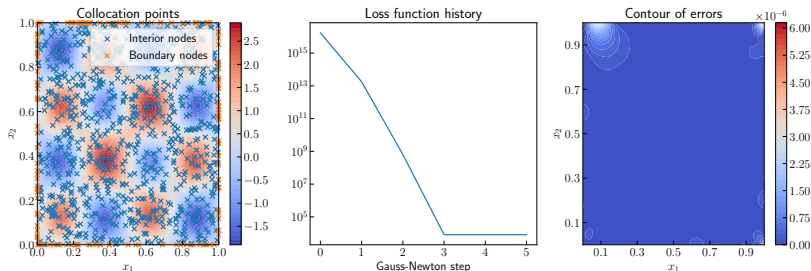


Figure: $N_{\text{domain}} = 900$, $N_{\text{boundary}} = 124$

Taming the Dense Kernel Matrix Numerically

Dense kernel matrix $K(\phi, \phi)$

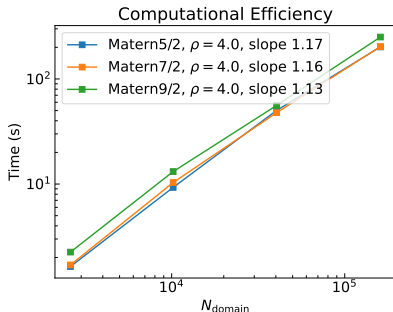
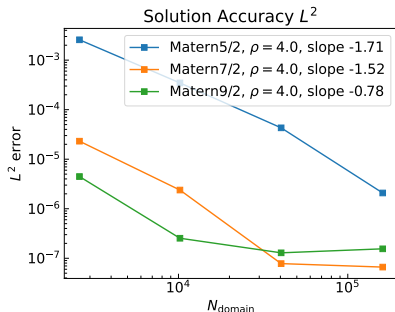
- Poor conditioning, and scale imbalance between blocks
Adding **scale-aware** nugget term $K(\phi, \phi) + \lambda \text{diag}(K(\phi, \phi))$
- Sparse Cholesky factorization under “coarse to fine” ordering
Thanks to **screening effects** hold for PDE-type measurements

⁵Michael L Stein. “The screening effect in kriging”. In: *Annals of statistics* 30.1 (2002), pp. 298–323.

⁶Florian Schäfer, Matthias Katzfuss, and Houthan Owhadi. “Sparse Cholesky Factorization by Kullback–Leibler Minimization”. In: *SIAM Journal on Scientific Computing* 43.3 (2021), A2019–A2046.

Near Linear Complexity by Sparse Cholesky

- Sparse Cholesky parameter $\rho = 4.0$
- Matérn kernel regularity parameter $\nu = 5/2, 7/2, 9/2$



- Accuracy floor due to finite ρ and nugget terms

⁷Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Numerical Experiments: Time Dependent Problems

Viscous Burgers' Equation

- Viscosity $\nu = 0.02$

$$\begin{cases} \partial_t u + u \partial_s u - \nu \partial_s^2 u = 0, & \forall (s, t) \in (-1, 1) \times (0, 1]. \\ u(s, 0) = -\sin(\pi s), \\ u(-1, t) = u(1, t) = 0. \end{cases}$$

- Shock when $\nu = 0$. Problem harder for smaller ν
- Choose an anisotropic spatio-temporal GP

Numerical Experiments: Viscous Burgers' Equation

- Kernel: $K((s, t), (s', t')) = \exp(-20^2 |s - s'|^2 - 3^2 |t - t'|^2)$

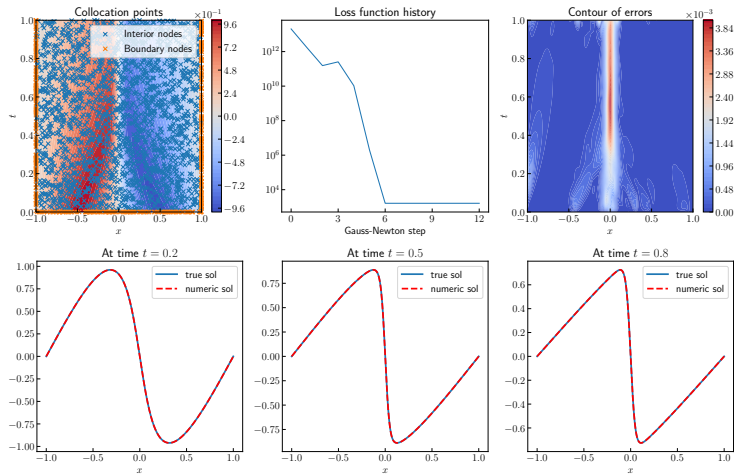


Figure: $N_{\text{domain}} = 2000, N_{\text{boundary}} = 400$

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Numerical Experiments: Inverse Problems

Darcy Flow inverse problems

$$\left\{ \begin{array}{ll} \min_{u,a} \|u\|_K^2 + \|a\|_\Gamma^2 + \frac{1}{\gamma^2} \sum_{j=1}^I |u(\mathbf{x}_j) - o_j|^2, \\ \text{s.t.} \quad \begin{array}{ll} -\text{div}(\exp(a)\nabla u)(\mathbf{x}_m) = 1, & \forall \mathbf{x}_m \in (0,1)^2 \\ u(\mathbf{x}_m) = 0, & \forall \mathbf{x}_m \in \partial(0,1)^2. \end{array} \end{array} \right.$$

- Recover a from pointwise measurements of u
- Model (u, a) as independent GPs
- Impose PDE constraints and formulate Bayesian inverse problem

Numerical Experiments: Darcy Flow

- Kernel $K(\mathbf{x}, \mathbf{x}'; \sigma) = \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma^2}\right)$ for both u and a

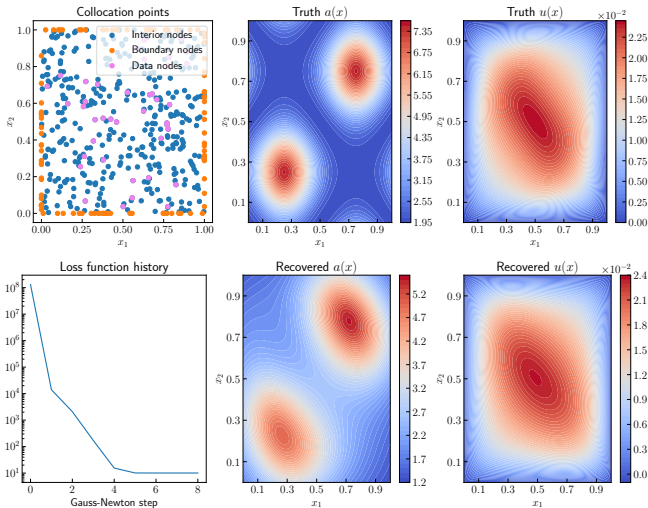


Figure: $N_{\text{domain}} = 400, N_{\text{boundary}} = 100, N_{\text{observation}} = 50$

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Theoretical Foundation: Consistency

Consistency of the minimizer

$$\begin{cases} \min_{u \in \mathcal{U}} & \|u\| \\ \text{s.t.} & \text{PDE constraints at } \{\mathbf{x}_1, \dots, \mathbf{x}_M\} \in \overline{\Omega}. \end{cases}$$

Convergence theory

- K is chosen so that
 - $\mathcal{U} \subseteq H^s(\Omega)$ for some $s > s^*$ where $s^* = d/2 + \text{order of PDE}$.
 - $u^* \in \mathcal{U}$.
- Fill distance of $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \rightarrow 0$ as $M \rightarrow \infty$.

Then as $M \rightarrow \infty$, $u^\dagger \rightarrow u^*$ pointwise in Ω and in $H^t(\Omega)$ for $t \in (s^*, s)$.

Roadmap

- 1 Motivation
 - Numerical Computation via Inference
- 2 The Methodology
 - Formulation
 - Representer Theorem
 - Algorithm
- 3 Numerical Examples
 - Elliptic PDEs
 - Viscous Burgers' Equation
 - Darcy Flow
- 4 Theoretical Foundation
 - Consistency
- 5 Summary
 - Take-aways

Solving and Learning Nonlinear PDEs with Gaussian Processes

Algorithm

- A simple framework for solving and learning PDEs, via GPs
- Near-linear complexity treatment of the dense kernel matrices
- Experiments: stationary PDEs, time dependent, inverse problems

Theory

- Consistency as fill-in distance goes to 0
- Consistency of kernel learning: Kernel Flow and Empirical Bayes

Thank you!