Gaussian Processes and Kernel Methods

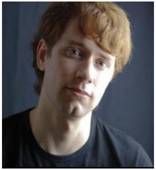
for Solving Nonlinear PDEs and Inverse Problems

Yifan Chen

Applied and Computational Math, Caltech

USNCCM, July 2023

Collaborators

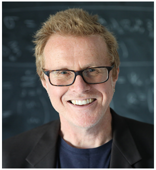


Pau Batlle
Caltech

Bamdad Hosseini
Univ. of Washington

Houman Owhadi
Caltech

Florian Schäfer
Georgia Tech



Andrew Stuart
Caltech

Solving PDEs/Inverse Problems

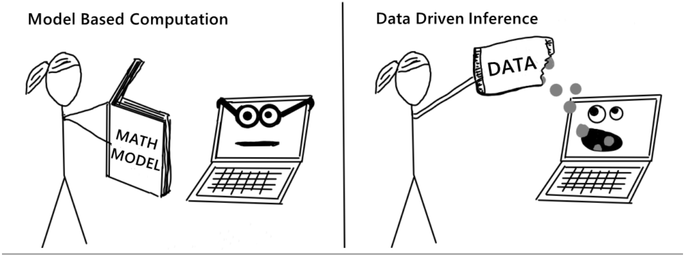
Traditional numerical methods designed by experts

- Finite difference/element/volume, spectral methods, ...
- Adjoint methods, ...

Solving PDEs/Inverse Problems

Traditional numerical methods designed by experts

- Finite difference/element/volume, spectral methods, ...
- Adjoint methods, ...



Machine learning methods aiming for automation

- Physics informed neural networks, ...
- Operator learning, ...

Our Focus: Gaussian Processes and Kernel Methods

Advantages

- Interpretable, amenable to analysis, and built-in UQ
- Connect to traditional meshless methods
- Connect to neural network methods in the infinite-width limit

Many related works in the literature

- [Poincaré 1896], [Palasti, Renyi 1956], [Sul'din 1959], [Sard 1963], [Kimeldorf, Wahba 1970], [Larkin 1972], [Traub, Wasilkowski, Woźniakowski 1988], [Diaconis 1988], [Schaback, Wendland 2006], [Stuart 2010], [Owhadi 2015], [Hennig, Osborne, Girolami 2015], [Cockayne, Oates, Sullivan, Girolami 2017], [Raissi, Perdikaris, Karniadakis 2017], ...

This talk

- A **rigorous, scalable** computational framework for solving **nonlinear** PDEs and inverse problems

Outline

Outline

The Methodology

A nonlinear elliptic PDE example for demonstration

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

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

The Methodology for A Nonlinear Elliptic PDE

1. Choose a kernel $K : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$ (Choose the prior $\mathcal{GP}(0, K)$)
   - Corresponding RKHS $\mathcal{U}$ with norm $\|\cdot\|_K$
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")

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

- Convergence guarantee when solution is in $\mathcal{U}$
- Uncertainty quantification can also be done

How to Solve: Introducing Slack Variables

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

$$\Updownarrow$$

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

How to Solve: Inner optimization

A linear inner problem

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

- Notations for kernel vectors and matrices

$$K(\mathbf{x}, \boldsymbol{\phi}) = \Big(K(\mathbf{x}, X^{\mathsf{bd}}), K(\mathbf{x}, X^{\mathsf{int}}), \Delta_{\mathbf{y}} K(\mathbf{x}, X^{\mathsf{int}})\Big) \in \mathbb{R}^N$$

$$K(\boldsymbol{\phi}, \boldsymbol{\phi}) =$$

$$\begin{pmatrix} K(X^{\mathsf{bd}}, X^{\mathsf{bd}}) & K(X^{\mathsf{bd}}, X^{\mathsf{int}}) & \Delta_{\mathbf{y}} K(X^{\mathsf{bd}}, X^{\mathsf{int}}) \\ K(X^{\mathsf{int}}, X^{\mathsf{bd}}) & K(X^{\mathsf{int}}, X^{\mathsf{int}}) & \Delta_{\mathbf{y}} K(X^{\mathsf{int}}, X^{\mathsf{int}}) \\ \Delta_{\mathbf{x}} K(X^{\mathsf{int}}, X^{\mathsf{bd}}) & \Delta_{\mathbf{x}} K(X^{\mathsf{int}}, X^{\mathsf{int}}) & \Delta_{\mathbf{x}} \Delta_{\mathbf{y}} K(X^{\mathsf{int}}, X^{\mathsf{int}}) \end{pmatrix}$$

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

How to Solve: Finite Dimensional Representation

Representer Theorem

Every minimizer $u^\dagger$ can be represented as

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

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

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

- $F : \mathbb{R}^N \to \mathbb{R}^M$ encodes PDE on collocation points
- $\mathbf{y}$ encondes PDE boundary and RHS data
- We can solve the optimization by sequential quadratic programming (equivalent to Gauss-Newton)

Outline

Outline

Numerical Experiments: Elliptic PDEs

- Equation with $\tau(u) = u^3$, $d = 2$

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

- Kernel: $K(\mathbf{x}, \mathbf{y}; \sigma) = \exp\left(-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma^2}\right), \sigma = 0.2$

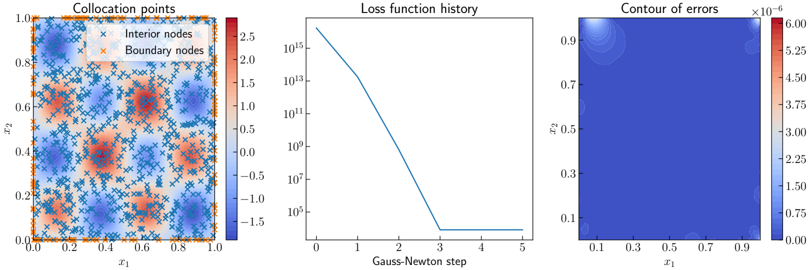


Figure: $N_{\text{domain}} = 900, N_{\text{boundary}} = 124$

Convergence Study

- For $\tau(u) = 0, u^3$, use Gaussian kernel with lengthscale σ
- L^2, L^∞ accuracy, compared with Finite Difference (FD)

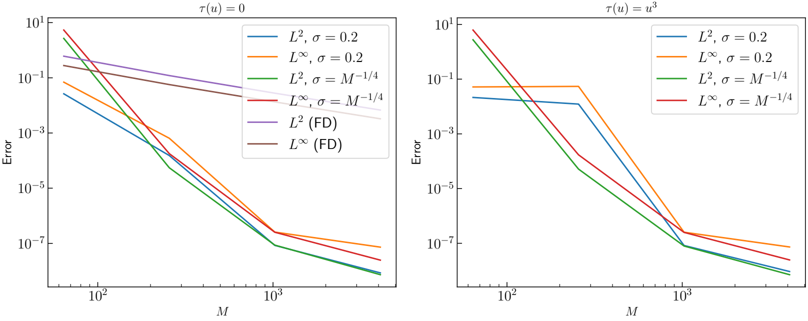


Figure: Fast convergence, since the solution is smooth

Outline

Darcy Flow Example

Darcy Flow inverse problems

- Equation: $-\nabla \cdot (\exp(a)\nabla u) = 1$ in Ω, and $u = 0$ on $\partial\Omega$
- Unknown functions a, u
- Measurement data $u(\mathbf{x}_j^{\mathrm{data}}) = o_j + \mathcal{N}(0, \gamma^2), 1 \leq j \leq N_{\mathrm{data}}$

$$\begin{aligned} \underset{u,a}{\text{minimize}} \quad & \|u\|_K^2 + \|a\|_K^2 + \frac{1}{\gamma^2}\sum_{j=1}^{N_{\mathsf{data}}} |u(\mathbf{x}_j^{\mathsf{data}}) - o_j|^2 \\ \text{constraint} \quad & -\nabla \cdot (\exp(a)\nabla u)(\mathbf{x}_m^{\mathrm{int}}) = 1 \text{ for some } \mathbf{x}_m^{\mathrm{int}} \in (0,1)^2 \\ & u(\mathbf{x}_m^{\mathrm{bd}}) = 0 \text{ for some } \mathbf{x}_m^{\mathrm{bd}} \in \partial(0,1)^2 \end{aligned}$$

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

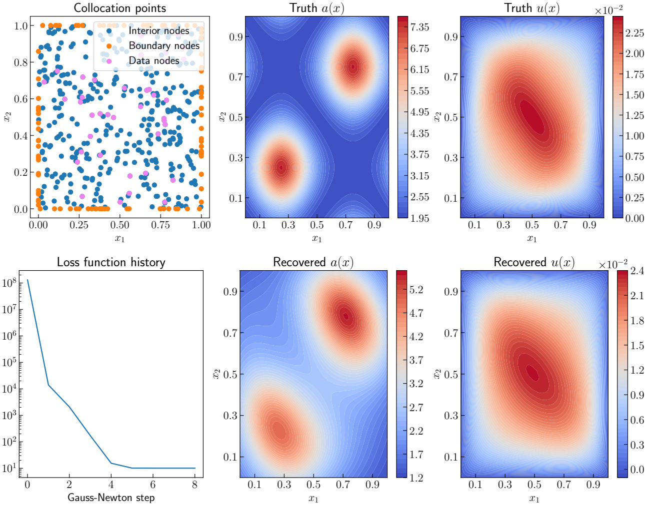


Figure: $N_{\text{domain}} = 400, N_{\text{boundary}} = 100, N_{\text{data}} = 50$

Other Examples of Nonlinear and Parametric PDEs

Reported in [Chen, Hosseni, Owhadi, Stuart 2021], [Batlle, Chen, Hosseni, Owhadi, Stuart 2023], [Chen, Owhadi, Schäfer 2023]

- Burgers' equations: $u_t + uu_x = \nu u_{xx}$
- Regularized Eikonal equations: $|\nabla u|^2 = f^2 + \epsilon \Delta u$
- Hamilton-Jacobi equations: $(\partial_t + \Delta)V(x,t) - |\nabla V(x,t)|^2 = 0$
- Parametric elliptic equations: $\nabla_x \cdot (a(x,\theta)\nabla_x u(x,\theta)) = f(x)$
- Monge-Amperè equations: $\det(D^2 u) = f$

Overall observations:

- The method is fast and achieves high accuracy with $10^3 - 10^4$ collocation points, if the solution is relatively smooth and Matérn/Gaussian kernels are chosen
- For more challenging cases, kernel learning can be used to adapt the kernel to the solution. Sparse Cholesky factorization algorithms can be applied to address $> 10^6$ collocation points

Thank You

Gaussian processes and kernel methods for

- **Solving nonlinear PDEs and inverse problems**
  - General computational framework for both
  - Convergence guarantee when kernel selected properly
  - Fast convergence using sequential quadratic programming

Relevant papers

- Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. JCP, 2021.
- Yifan Chen, Houman Owhadi, Florian Schaefer. Sparse Cholesky Factorization for Solving Nonlinear PDEs via Gaussian Processes. arxiv: 2304.01294, 2023.
- Pau Batlle, Yifan Chen, Bamdad Hosseini, Houman Owhadi, Andrew M. Stuart. Error Analysis of Kernel/GP Methods for Nonlinear and Parametric PDEs. arxiv: 2305.04962, 2023.

Back Up Slides

Summary

Gaussian processes and kernel methods

- **Solving PDEs and inverse problems**
  - General computational framework for both
  - Convergence guarantee when kernel selected properly
  - Fast convergence using sequential quadratic programming

- **Kernel learning and sparse Cholesky factorization**
  - Adapt the kernel to the solution
  - Scale to massive collocation points
  - Future works: adaptive sampling of the points

Convergence Theory for Solving PDEs

Convergence of the minimizer $u^\dagger$ to the truth $u^\star$

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

Asymptotic convergence [Chen, Hosseni, Owhadi, Stuart 2021]

Assumptions:

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

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

- Convergence rates obtained when stability of the PDE is further assumed **[Batlle, Chen, Hosseni, Owhadi, Stuart 2023]**

Burgers' Equation

- $\partial_t u + u\partial_x u - 0.001\partial_x^2 u = 0, \quad \forall (x,t) \in (-1,1) \times (0,1]$
- $\Delta t = 0.02, \rho = 4$, solve to $t = 1$

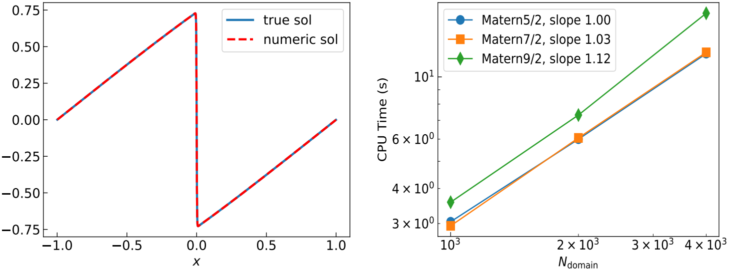


Figure: Run 2 linearization steps at each time step

Monge-Ampère Equation

- Equation: $\det(D^2u) = f$ in $(0,1)^2$
- Truth $u(\mathbf{x}) = \exp\big(0.5((x_1 - 0.5)^2 + (x_2 - 0,5)^2)\big)$
- Matérn kernel with $\nu = 5/2$, lengthscale 0.3

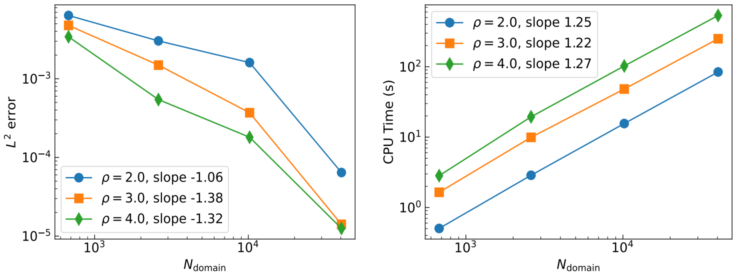


Figure: Run 3 linearization steps with initial guess $1/2\|\mathbf{x}\|^2$. Accuracy floor due to finite ρ