On Multiscale and Statistical Numerical Methods

for PDEs and Inverse Problems

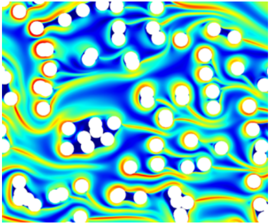
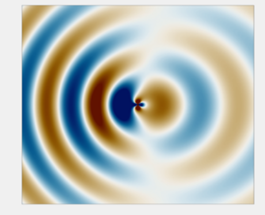
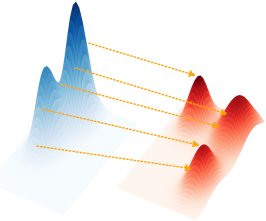
Yifan Chen

Applied and Computational Mathematics, Caltech

Thesis Defense, May 2023

Partial Differential Equations (PDEs)

PDEs widely employed in scientific computing and scientific ML

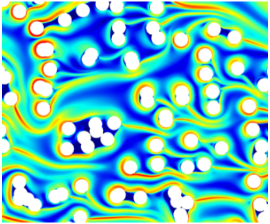
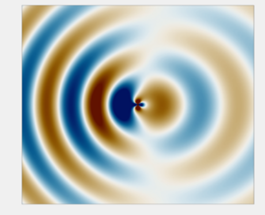
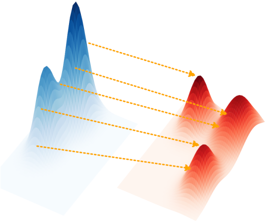




e.g., flows, waves, transport of data and uncertainty, ...

Figure credited to Google online search

Partial Differential Equations (PDEs)

PDEs widely employed in scientific computing and scientific ML

e.g., flows, waves, transport of data and uncertainty, ...

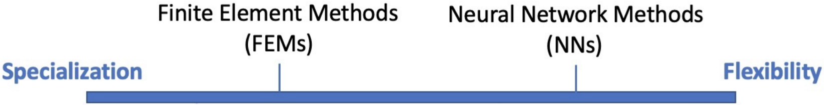
Problem of focus: Numerical methods for PDEs/inverse problems

Figure credited to Google online search

Numerical Methods for PDEs and Inverse Problems

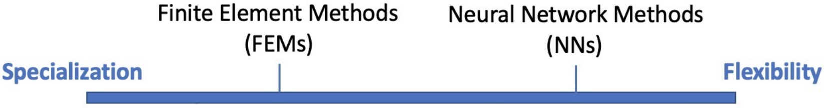
Key step: Construct finite dimensional numerical approximations

Key step: Construct finite dimensional numerical approximations



Numerical Methods for PDEs and Inverse Problems

Key step: Construct finite dimensional numerical approximations

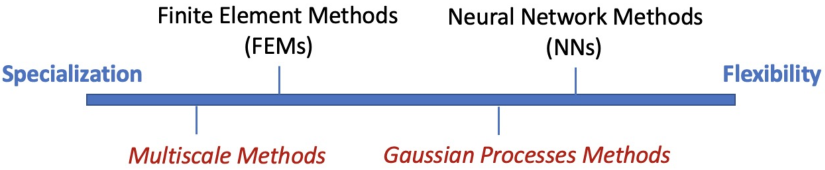


Challenges:

- FEMs not specialized enough to solve multiscale PDEs
- NNs flexible but may sometimes be too complicated to analyze

Numerical Methods for PDEs and Inverse Problems

Key step: Construct finite dimensional numerical approximations



Challenges:

- FEMs not specialized enough to solve multiscale PDEs
- NNs flexible but may sometimes be too complicated to analyze

Our focus: Addressing the above challenges by advancing *multiscale and Gaussian processes methods*

Outline

Outline

Multiscale Problems

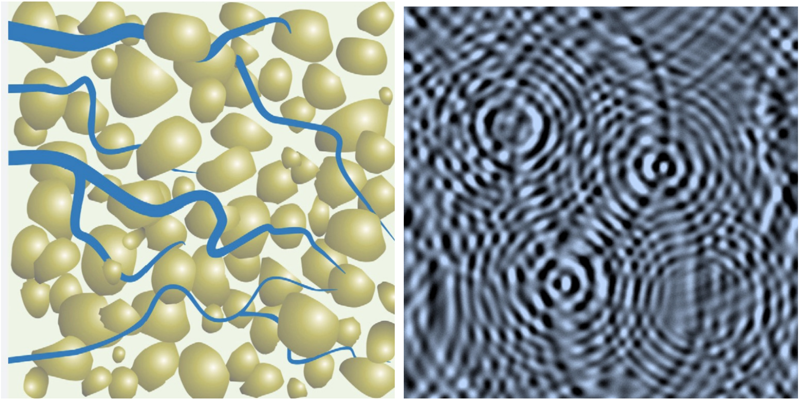


Figure: Heterogeneity and high frequency

Figure credited to Google online search

Mathematical Setup

Model problem: Heterogeneous Helmholtz's equation

$$-\nabla \cdot (A\nabla u) - k^2 u = f, \text{ in } \Omega, \quad \text{w/ boundary conditions}$$

(subsurface flows, diffusions, elasticity, waves)

Mathematical Setup

Model problem: Heterogeneous Helmholtz's equation

$$-\nabla \cdot (A\nabla u) - k^2 u = f, \text{ in } \Omega, \quad \text{w/ boundary conditions}$$

(subsurface flows, diffusions, elasticity, waves)

Mathematical conditions for multiscale phenomenon:

- Heterogeneity (i.e., A varies a lot spatially):

$$A \in L^\infty(\Omega), \text{ and } 0 < A_{\min} \le A(x) \le A_{\max} < \infty$$

- High frequency: k^2 is large

Mathematical Setup

Model problem: Heterogeneous Helmholtz's equation

$$-\nabla \cdot (A\nabla u) - k^2 u = f, \text{ in } \Omega, \quad \text{w/ boundary conditions}$$

(subsurface flows, diffusions, elasticity, waves)

Mathematical conditions for multiscale phenomenon:

- Heterogeneity (i.e., A varies a lot spatially):

$$A \in L^\infty(\Omega), \text{ and } 0 < A_{\min} \le A(x) \le A_{\max} < \infty$$

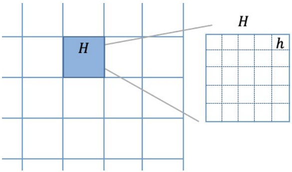
- High frequency: k^2 is large

Challenges of FEMs: Need very small grid size h for accuracy

- $h \le h^\star \ll 1$ to resolve the heterogeneity
- $h = O(1/k^2)$ to handle the indefiniteness (known as pollution effects) [Babuška, Osborn 2000], [Babuška, Sauter, 1997]

Set-up: $\Omega = [0,1]^d$

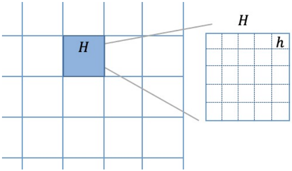
- Fine grid size $h \ll 1$ small enough to resolve the physics
- Coarse grid size H



Typical Ingredients of Multiscale Methods "Divide and Conquer"

Set-up: $\Omega = [0,1]^d$

- Fine grid size $h \ll 1$ small enough to resolve the physics
- Coarse grid size H

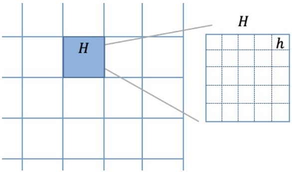


Offline: given A, solve some local problems of domain size $O(H)$ to get local basis functions. Assembly the stiffness matrix

- Local computation, parallelizable

Set-up: $\Omega = [0,1]^d$

- Fine grid size $h \ll 1$ small enough to resolve the physics
- Coarse grid size H



Offline: given A, solve some local problems of domain size $O(H)$ to get local basis functions. Assembly the stiffness matrix

- Local computation, parallelizable

Online: for any source term f, solve a global linear system with the stiffness matrix to get coefficients of the basis functions

- Computation involves a linear system of size $=$ the number of local basis functions

Compared to the complexity of fine-grid FEM

- Computation involves a global linear system of size $\sim 1/h^d$

Many Multiscale Methods for Constructing Local Basis Functions

Handling rough coefficients $A \in L^\infty(\Omega)$:

- Harmonic coordinates [Owhadi, Zhang 2007]
- Multiscale spectral generalized FEMs [Babuška, Lipton 2011]
- Generalized Multiscale FEM [Efendiev, Galvis, Hou 2013]
- Rough polyharmonic splines [Owhadi, Zhang, Berlyand 2014]
- Local orthogonal decomposition [Målqvist, Peterseim 2014]
- Gamblets [Owhadi 2017]
- ...

Handling large k:

- hp-FEM: [Melenk, Sauter 2010, 2011]
- Local orthogonal decomposition: [Peterseim, et al 2017]
- Wavelet-based edge multiscale FEM [Fu, Li, Craster, Guenneau 2021]
- ...

What Constitutes An Ideal Multiscale Method?

High level parameters of a multiscale method

- H: size of coarse grid
- lH: size of local domains for computing basis functions
- m: number of basis functions in each local domain

- Let e be the error of the solution obtained by the multiscale method
- Ideally, for a fixed H, we want small m, l and e

Our Contributions **[Chen, Hou, Wang 2021,2021,2022]**

Exponentially convergent multiscale FEM (ExpMsFEM)

A multiscale framework for heterogeneous Helmholtz's equations

- Require $H = O(1/k)$ (standard in the literature)
- Error $e \leq C_\epsilon \exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)(\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)})$
- Framework based on non-overlapped domain decomposition

Our Contributions [Chen, Hou, Wang 2021,2021,2022]

Exponentially convergent multiscale FEM (ExpMsFEM)

A multiscale framework for heterogeneous Helmholtz's equations

- Require $H = O(1/k)$ (standard in the literature)
- Error $e \leq C_\epsilon \exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)(\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)})$
- Framework based on non-overlapped domain decomposition

- Pre-existing[1] methods for heterogeneous Helmholtz's equations are at most algebraic convergence, or have accuracy floor $O(H)$, e.g. $e = O(H), m = 1, l = O(\log(1/H)\log k)$ [Peterseim, et al 2017]

[1] There is a contemporary work achieving exponential convergence based on overlapped domain decomposition [Ma, Alber, Scheichl 2021]

Our Contributions **[Chen, Hou, Wang 2021,2021,2022]**

Exponentially convergent multiscale FEM (ExpMsFEM)

A multiscale framework for heterogeneous Helmholtz's equations

- Require $H = O(1/k)$ (standard in the literature)
- Error $e \le C_\epsilon \exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)(\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)})$
- Framework based on non-overlapped domain decomposition

- Pre-existing[1] methods for heterogeneous Helmholtz's equations are at most algebraic convergence, or have accuracy floor $O(H)$, e.g. $e = O(H), m = 1, l = O(\log(1/H)\log k)$ [Peterseim, et al 2017]

- Non-overlapped domain decomposition leads to basis functions with smaller support and less overlapping (l is smaller). Pre-existing methods rely on overlapped domain decomposition [Babuška, Lipton 2011].

[1] There is a contemporary work achieving exponential convergence based on overlapped domain decomposition [Ma, Alber, Scheichl 2021]

How Does ExpMsFEM Work?

Local Structure $+$ **Global Decomposition**

How Does ExpMsFEM Work?

Local Structure + Global Decomposition

Local Structure: Helmholtz-harmonic functions

- $U_k(D) := \{v \in H^1(D), -\nabla \cdot (A\nabla v) - k^2 v = 0 \text{ in } D\}/\mathbb{R}$
- Energy norm: $\|v\|^2_{\mathcal{H}(D)} := \|A^{1/2}\nabla v\|^2_{L^2(D)} + \|kv\|^2_{L^2(D)}$

Local Structure: Helmholtz-harmonic functions

- $U_k(D) := \{v \in H^1(D), -\nabla \cdot (A\nabla v) - k^2 v = 0 \text{ in } D\}/\mathbb{R}$
- Energy norm: $\|v\|^2_{\mathcal{H}(D)} := \|A^{1/2}\nabla v\|^2_{L^2(D)} + \|kv\|^2_{L^2(D)}$

Theorem [Chen, Hou, Wang 2021]

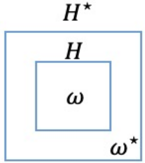
Let $H^\star = O(1/k)$. Consider the restriction operator

$$R : (U_k(\omega^*), \|\cdot\|_{\mathcal{H}(\omega^*)}) \to (U_k(\omega), \|\cdot\|_{\mathcal{H}(\omega)})$$

such that $Rv = v|_\omega$. Then, its singular values $\sigma_m(R)$ decays nearly exponentially fast:

$$\sigma_m(R) \leq C_\epsilon \exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)$$

for some C_ϵ independent of k, H and m



- Pre-existing result for A-harmonic functions [Babuška, Lipton 2011]
- Key of analysis: $H^\star = O(1/k) \Rightarrow$ the Helmholtz operator is locally positive definite and elliptic techniques can apply

Consequence: if $H^\star = O(1/k)$, then

- For any $u \in U_k(\omega^*)$, there are m functions $v_j, 1 \le j \le m$, s.t.

$$\inf_{c_j} \|u - \sum_{j=1}^{m} c_j v_j\|_{\mathcal{H}(\omega)} \le C_\epsilon \exp\Big(-m^{\frac{1}{d+1}-\epsilon}\Big)\|u\|_{\mathcal{H}(\omega^*)}$$

- v_j are left singular vectors of the restriction operator R

Sanity check

1. SVD: $R = \sum_j \sigma_j v_j \otimes w_j$ where $v_j \in U_k(\omega)$ and $w_j \in U_k(\omega^\star)$
2. $Ru = \sum_j \sigma_j v_j \langle u, w_j\rangle$
3. $Ru - \sum_{j=1}^m \sigma_j v_j \langle u, w_j\rangle = \sum_{j>M} \sigma_j v_j \langle u, w_j\rangle$
4. $\|u - \sum_{j=1}^m \sigma_j v_j \langle u, w_j\rangle\|_{\mathcal{H}(\omega)} \le \sigma_{m+1}\|u\|_{\mathcal{H}(\omega^\star)}$

Consequence: if $H^\star = O(1/k)$, then

- For any $u \in U_k(\omega^*)$, there are m functions $v_j, 1 \le j \le m$, s.t.

$$\inf_{c_j} \|u - \sum_{j=1}^{m} c_j v_j\|_{\mathcal{H}(\omega)} \le C_\epsilon \exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)\|u\|_{\mathcal{H}(\omega^*)}$$

- v_j are left singular vectors of the restriction operator R

Sanity check

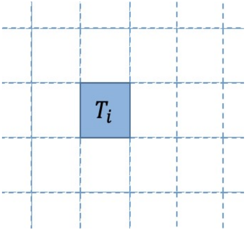
1. SVD: $R = \sum_j \sigma_j v_j \otimes w_j$ where $v_j \in U_k(\omega)$ and $w_j \in U_k(\omega^\star)$
2. $Ru = \sum_j \sigma_j v_j \langle u, w_j \rangle$
3. $Ru - \sum_{j=1}^{m} \sigma_j v_j \langle u, w_j \rangle = \sum_{j>M} \sigma_j v_j \langle u, w_j \rangle$
4. $\|u - \sum_{j=1}^{m} \sigma_j v_j \langle u, w_j \rangle\|_{\mathcal{H}(\omega)} \le \sigma_{m+1} \|u\|_{\mathcal{H}(\omega^\star)}$

Summarize the property: Restrictions of Helmholtz-harmonic functions are of low approximation complexity

How Does ExpMsFEM Work?

Local Structure $+$ **Global Decomposition**

1. **Decomposition using indicator funcs**

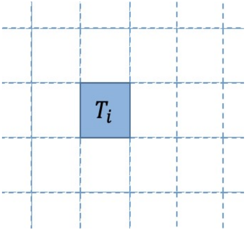


$$u = \sum_i \mathbb{1}_{T_i} u$$
$$= \sum_i \mathbb{1}_{T_i} u^{\mathsf{h}}_{T_i} + \underbrace{\sum_i \mathbb{1}_{T_i} u^{\mathsf{b}}_{T_i}}_{\text{small and locally computable}}$$

$$\begin{cases} -\nabla \cdot (A \nabla u^{\mathsf{h}}_{T_i}) - k^2 u^{\mathsf{h}}_{T_i} = 0, & \text{in } T_i \\ u^{\mathsf{h}}_{T_i} = u, & \text{on } \partial T_i \end{cases}$$
$$\begin{cases} -\nabla \cdot (A \nabla u^{\mathsf{b}}_{T_i}) - k^2 u^{\mathsf{b}}_{T_i} = f, & \text{in } T_i \\ u^{\mathsf{b}}_{T_i} = 0, & \text{on } \partial T_i \end{cases}$$

Global Decomposition in 2D

1. Decomposition using indicator funcs



$$u = \sum_i \mathbb{1}_{T_i} u$$

$$= \color{red}{\sum_i \mathbb{1}_{T_i} u^{\mathsf{h}}_{T_i}} + \underbrace{\sum_i \mathbb{1}_{T_i} u^{\mathsf{b}}_{T_i}}_{\text{small and locally computable}}$$

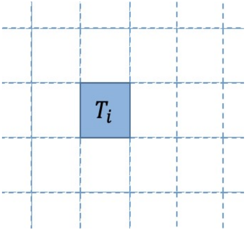
$$\begin{cases} -\nabla \cdot (A \nabla u^{\mathsf{h}}_{T_i}) - k^2 u^{\mathsf{h}}_{T_i} = 0, & \text{in } T_i \\ u^{\mathsf{h}}_{T_i} = u, & \text{on } \partial T_i \end{cases}$$

$$\begin{cases} -\nabla \cdot (A \nabla u^{\mathsf{b}}_{T_i}) - k^2 u^{\mathsf{b}}_{T_i} = f, & \text{in } T_i \\ u^{\mathsf{b}}_{T_i} = 0, & \text{on } \partial T_i \end{cases}$$

Goal: write $\color{red}{\sum_i \mathbb{1}_{T_i} u^{\mathsf{h}}_{T_i}}$ as local restrictions of Helmholtz-harmonic functions

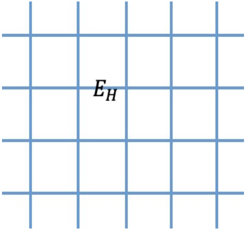
Global Decomposition in 2D

1. Decomposition using indicator funcs



$$u = \sum_i \mathbb{1}_{T_i} u$$
$$= \sum_i \mathbb{1}_{T_i} u^{\mathsf{h}}_{T_i} + \underbrace{\sum_i \mathbb{1}_{T_i} u^{\mathsf{b}}_{T_i}}_{\text{small and locally computable}}$$

2. Focus on edge functions



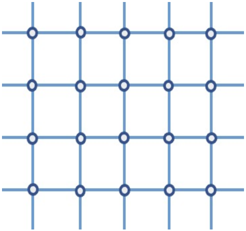
$$\sum_i \mathbb{1}_{T_i} u^{\mathsf{h}}_{T_i} = Q\tilde{u}^{\mathsf{h}}$$

- where $Q : H^{1/2}(E_H) \to H^1(\Omega)$ is the Helmholtz-harmonic extension operator

3. Edge localization

$$\tilde{u}^{\mathsf{h}} = \underbrace{I_H \tilde{u}^{\mathsf{h}}}_{\text{Nodal interp.}} + \underbrace{(\tilde{u}^{\mathsf{h}} - I_H \tilde{u}^{\mathsf{h}})}_{\text{Decoupled to each edges}}$$

- $I_H \tilde{u}^{\mathsf{h}} = \sum_n u(x_n)\psi_n$ spanned by nodal basis funcs

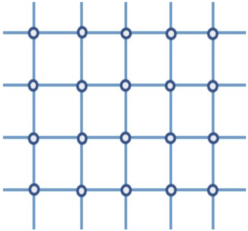


Global Decomposition

3. Edge localization

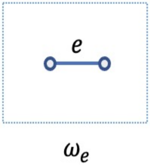
$$\tilde{u}^{\mathsf{h}} = \underbrace{I_H \tilde{u}^{\mathsf{h}}}_{\text{Nodal interp.}} + \underbrace{(\tilde{u}^{\mathsf{h}} - I_H \tilde{u}^{\mathsf{h}})}_{\text{Decoupled to each edges}}$$

- $I_H \tilde{u}^{\mathsf{h}} = \sum_n u(x_n)\psi_n$ spanned by nodal basis funcs



4. Oversampling

$$(\tilde{u}^{\mathsf{h}} - I_H \tilde{u}^{\mathsf{h}})|_e = (u - I_H u)|_e$$
$$= \sum_{j=1}^{m} c_{j,e} \tilde{v}_{j,e} + O\left(\exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)\right)$$
$$+ \underbrace{(u^{\mathsf{b}}_{\omega_e} - I_H u^{\mathsf{b}}_{\omega_e})|_e}_{\text{small and locally computable}}$$



Oversampling domain

ExpMsFEM in 2D

Theorem [Chen, Hou, Wang 2021]

The following holds for the solution u of Helmholtz's equation

$$
\begin{aligned}
u = &\left(\sum_n b_n \psi_n + \sum_e \sum_{j=1}^{m} c_{j,e} v_{j,e}\right) \\
&+ \left(\sum_i \mathbb{1}_{T_i} u^{\mathsf{b}}_{T_i} + \sum_e Q(u^{\mathsf{b}}_{\omega_e} - I_H u^{\mathsf{b}}_{\omega_e})|_e\right) \\
&\quad + O\left(\exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)(\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)})\right)
\end{aligned}
$$

ExpMsFEM in 2D

Theorem [Chen, Hou, Wang 2021]

The following holds for the solution u of Helmholtz's equation

$$
\begin{aligned}
u = &\left(\sum_n b_n \psi_n + \sum_e \sum_{j=1}^{m} c_{j,e} v_{j,e} \right) \\
&+ \left(\sum_i \mathbb{1}_{T_i} u_{T_i}^{\mathsf{b}} + \sum_e Q(u_{\omega_e}^{\mathsf{b}} - I_H u_{\omega_e}^{\mathsf{b}})|_e \right) \\
&\quad + O\left(\exp\left(-m^{\frac{1}{d+1} - \epsilon} \right) (\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)}) \right)
\end{aligned}
$$

- Line 1 consists nodal, edge basis functions
  - $O(m/H^2)$ number of local basis functions, obtained by solving local spectral problems
  - $b_n, c_{j,e}$ can be computed by Galerkin's methods

ExpMsFEM in 2D

Theorem [Chen, Hou, Wang 2021]

The following holds for the solution u of Helmholtz's equation

$$u = \left(\sum_n b_n\psi_n + \sum_e \sum_{j=1}^{m} c_{j,e}v_{j,e}\right) + \left(\sum_i \mathbb{1}_{T_i} u^{\mathsf{b}}_{T_i} + \sum_e Q(u^{\mathsf{b}}_{\omega_e} - I_H u^{\mathsf{b}}_{\omega_e})|_e\right) + O\left(\exp\left(-m^{\frac{1}{d+1}-\epsilon}\right)(\|u\|_{\mathcal{H}(\Omega)} + \|f\|_{L^2(\Omega)})\right)$$

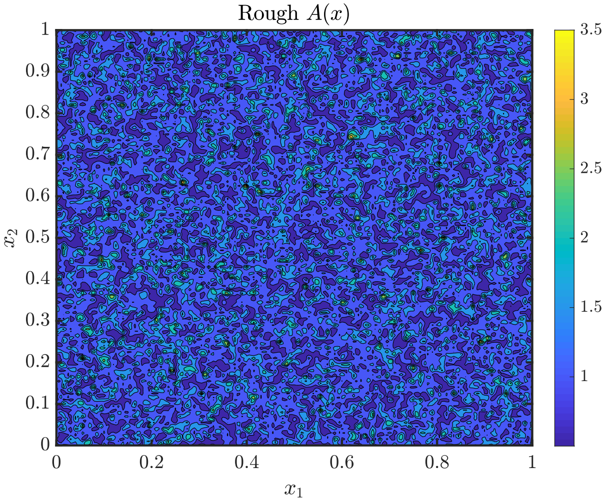
- Line 1 consists nodal, edge basis functions
  - $O(m/H^2)$ number of local basis functions, obtained by solving local spectral problems
  - $b_n, c_{j,e}$ can be computed by Galerkin's methods
- Line 2 consists fine scale bubble terms are locally computable
  - Obtained by solving local linear systems

Heterogeneous Helmholtz's equation:

$$-\nabla \cdot (A\nabla u) - k^2 u = f, \text{ in } \Omega = [0,1]^2$$

- Wavenumber $k = 2^5$
- $A(x) = |\xi(x)| + 0.5$ where $\xi(x)$ is piecewise linear functions
  - nodal values drawn from unit Gaussian random variable
  - piecewise scale: 2^{-7}
- Source term $f(x_1, x_2) = x_1^4 - x_2^3 + 1$
- Boundary condition: mixed
  - one side Dirichlet, one side Neumann, two sides Robin

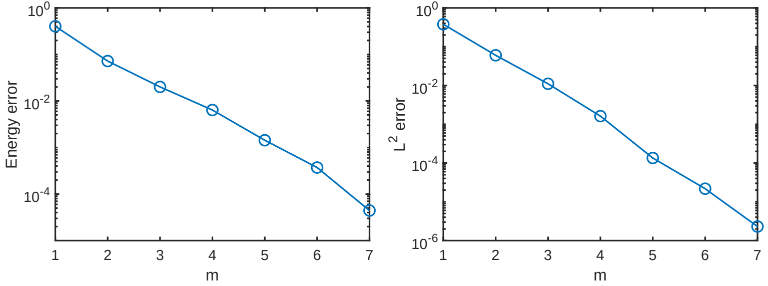
Visualization of the Field



Quadrilateral mesh

- Fine mesh size $h = 2^{-10}$, coarse mesh size $H = 2^{-5}$

Accuracy of ExpMsFEM's solution

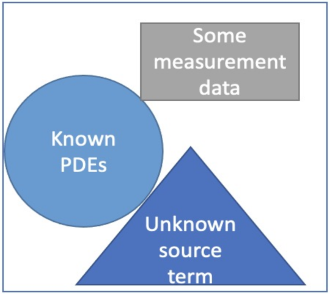


Note: $(2m+1)/H^2$ number of local basis functions are used. The accuracy is calculated by comparing to the fine mesh FEM solution.

Outline

Specialization and Flexibility of Solvers

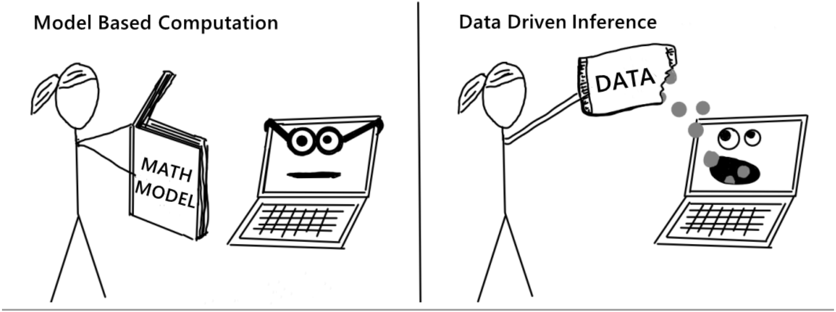
- Specialized solvers effective for their targeted class of problems
  - Design very accurate basis functions for approximation
- Real world problems are more fruitful and complicated



- Inverse problems
- Material design
- Multi-physics
- Data assimilation
- ...

Flexible numerical framework for many applications?

Model based v.s. data driven methods

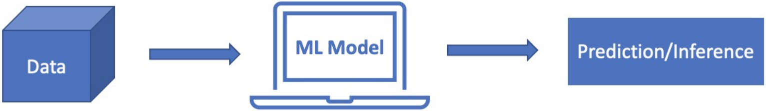


"Apply machine learning and statistical inference to automate scientific computing"

(PINNs, operator learning, ...)

Figure from Yiping Lu's slides, with some new edits by the presenter

Typical Ingredients of ML Based Methods for PDEs



"Data":

- PDE information: e.g., $-\Delta u(x_i) = f(x_i)$
- Physical measurements: e.g., $u(x_i) = y_i$ in inverse problems

ML model:

- Neural networks
- Gaussian processes and kernel methods
- Tensor format
- ...

Our Focus: Gaussian Processes for PDEs and Inverse Problems

Advantages:

- Interpretable, amenable to analysis, and built-in UQ
- Connect to radial basis funcs methods in numerical analysis
- 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], ...

What's new?

- A rigorous mathematical framework for nonlinear PDEs **[Chen, Hosseni, Owhadi, Stuart 2021]**

An Optimization Problem as the MAP Estimator of Gaussian Processes

The "maximum a posterior" (MAP) estimator

$$\underset{u\in\mathcal{U}}{\text{minimize}} \quad \|u\|_K$$

$$\text{constraint} \quad P(x,u,\Delta u,...) = 0 \text{ at some } x_1,...,x_M$$

- **Constraint:** "Data" (any PDE or measurement of u)
  - P can be nonlinear
  - e.g., $P(x,u,\Delta u,...) = -\Delta u + u^3$ or $P(x,u,\Delta u,...) = u$ or combination of both

- **Notation:** kernel function $K : \Omega \times \Omega \to \mathbb{R}$
  - Corresponding RKHS $\mathcal{U}$ with norm $\|\cdot\|_K$
  - Formally, $\|u\|_K^2 = [u, \mathcal{K}^{-1}u]_{L^2}$ where $\mathcal{K}v = \int K(\cdot,y)v(y)\mathrm{d}y$

- **In which sense it is MAP?**
  - Formally, density of $u \sim \mathcal{GP}(0,K)$ is $\propto \exp\left(-\frac{1}{2}\|u\|_K^2\right)$
  - Formally, $-\log\rho(u) = \frac{1}{2}\|u\|_K^2 + C$

An Optimization Problem as the MAP Estimator of Gaussian Processes

The "maximum a posterior" (MAP) estimator

$$\underset{u \in \mathcal{U}}{\text{minimize}} \quad \|u\|_K$$

$$\text{constraint} \quad P(x, u, \Delta u, ...) = 0 \text{ at some } x_1, ..., x_M$$

Equivalent finite dimensional optimization problem

- Nonlinear representer theorem:

$$\text{The optimizer } u^\dagger(x) \in \text{span}\{K(x, x_m), \Delta_{x_m} K(x, x_m), ... \text{ for } 1 \le m \le M\}$$

An Optimization Problem as the MAP Estimator of Gaussian Processes

The "maximum a posterior" (MAP) estimator

$$\begin{aligned} &\underset{u\in\mathcal{U}}{\text{minimize}} && \|u\|_K \\ &\text{constraint} && P(x,u,\Delta u,\ldots)=0 \text{ at some } x_1,\ldots,x_M \end{aligned}$$

Equivalent finite dimensional optimization problem

- Nonlinear representer theorem:

  The optimizer $u^\dagger(x) \in \mathrm{span}\{K(x,x_m), \Delta_{x_m}K(x,x_m), \ldots$ for $1 \le m \le M\}$

- Substitute these basis functions into the problem to get

$$\begin{cases} \underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} & \mathbf{z}^T\Theta^{-1}\mathbf{z} \\ \text{constraint} & F(\mathbf{z})=0 \end{cases}$$

- Θ dense kernel matrix with entries $K(x_m,x_n), \Delta_{x_m}K(x_m,x_n)$, ...
- F encodes the corresponding finite-dim constraint

An Optimization Problem as the MAP Estimator of Gaussian Processes

The "maximum a posterior" (MAP) estimator

$$\underset{u \in \mathcal{U}}{\text{minimize}} \quad \|u\|_K$$

$$\text{constraint} \quad P(x, u, \Delta u, ...) = 0 \text{ at some } x_1, ..., x_M$$

Equivalent finite dimensional optimization problem

- Nonlinear representer theorem:

  The optimizer $u^\dagger(x) \in \text{span}\{K(x, x_m), \Delta_{x_m} K(x, x_m), \ldots$
  $\text{for } 1 \leq m \leq M\}$

- Substitute these basis functions into the problem to get

$$\begin{cases} \underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} & \mathbf{z}^T \Theta^{-1} \mathbf{z} \\ \text{constraint} & F(\mathbf{z}) = 0 \end{cases}$$

  - Θ dense kernel matrix with entries $K(x_m, x_n), \Delta_{x_m} K(x_m, x_n)$, ...
  - F encodes the corresponding finite-dim constraint

- Solved by sequential quadratic programming

How Does It Perform?

1. Numerical experiments for solving nonlinear PDEs
2. Numerical experiments for Darcy flow inverse problems
3. Theoretical guarantee

How Does It Perform?

1. Numerical experiments for solving nonlinear PDEs
2. Numerical experiments for Darcy flow inverse problems
3. Theoretical guarantee

Nonlinear Elliptic Equation Example

The Laplacian equation with cubic nonlinearity:

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

$$\begin{aligned} \underset{u \in \mathcal{U}}{\text{minimize}} \quad & \|u\|_K \\ \text{constraint} \quad & -\Delta u(\mathbf{x}_m^{\text{int}}) + u(\mathbf{x}_m^{\text{int}})^3 = f(\mathbf{x}_m^{\text{int}}) \ \text{ for some } \mathbf{x}_m^{\text{int}} \in \Omega \\ & u(\mathbf{x}_m^{\text{bd}}) = g(\mathbf{x}_m^{\text{bd}}) \ \text{ for some } \mathbf{x}_m^{\text{bd}} \in \partial\Omega \end{aligned}$$

Numerical Experiments: Nonlinear Elliptic Equation

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

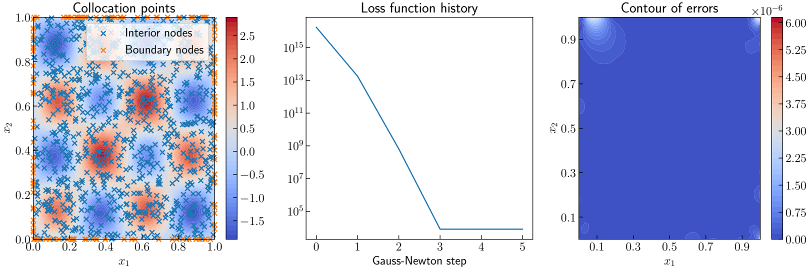


Figure: $N_{\text{domain}} = 900, N_{\text{boundary}} = 124$

- Solution is smooth, well approximated by Gaussian kernels

How Does It Perform?

1. Numerical experiments for solving nonlinear PDEs
2. Numerical experiments for Darcy flow inverse problems
3. Theoretical guarantee

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 $u(\mathbf{x}_j^{\text{data}}) = o_j + \mathcal{N}(0, \gamma^2), 1 \leq j \leq N_{\text{data}}$

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

Numerical Experiments: Darcy Flow

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

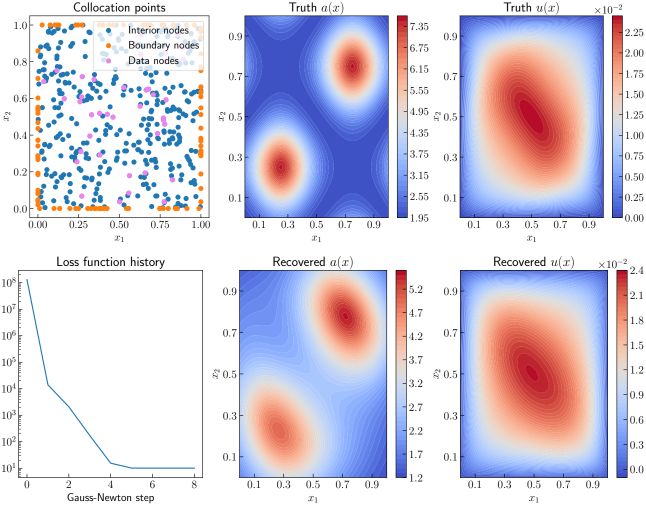


Figure: $N_{\text{domain}} = 400, N_{\text{boundary}} = 100, N_{\text{data}} = 50$

How Does It Perform?

1. Numerical experiments for solving nonlinear PDEs
2. Numerical experiments for Darcy flow inverse problems
3. Theoretical guarantee

Convergence Theory for Solving PDEs

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

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

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

Other Numerical Examples for Solving Nonlinear and Parametric PDEs

Reported in [Chen, Hosseni, Owhadi, Stuart 2021], [Batlle, Chen, Hosseni, Owhadi, Stuart 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: $\nabla_x \cdot (a(x,\theta)\nabla_x u(x,\theta)) = f$
- 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 pretty smooth and Matérn/Gaussian kernels are chosen

Efforts for Further Improvements

Adapt the model:

- Numerous approaches for learning the kernel to adapt to the problem
- **Challenges:** nonlinear procedure, limited theory

Efforts for Further Improvements

Adapt the model:

- Numerous approaches for learning the kernel to adapt to the problem
- **Challenges:** nonlinear procedure, limited theory

Sample more data:

- With enough data, any reasonable kernel functions can approximate well
- **Challenges:** dense kernel matrices with derivatives

Efforts for Further Improvements

Adapt the model:

- Numerous approaches for learning the kernel to adapt to the problem
- **Challenges:** nonlinear procedure, limited theory

Hierarchical kernel learning
First rigorous analysis of large data consistency and implicit bias for kernel flow algorithms for a Matérn-like model. Investigation of robustness to model misspecification

[Chen, Owhadi, Stuart 2020]

Sample more data:

- With enough data, any reasonable kernel functions can approximate well
- **Challenges:** dense kernel matrices with derivatives

Efforts for Further Improvements

Adapt the model:

- Numerous approaches for learning the kernel to adapt to the problem
- **Challenges:** nonlinear procedure, limited theory

Hierarchical kernel learning

First rigorous analysis of large data consistency and implicit bias for kernel flow algorithms for a Matérn-like model. Investigation of robustness to model misspecification

[Chen, Owhadi, Stuart 2020]

Sample more data:

- With enough data, any reasonable kernel functions can approximate well
- **Challenges:** dense kernel matrices with derivatives

Sparse Cholesky factorization

A new multiscale ordering of columns (with derivative entries) leading to approximately sparse Cholesky factors. Achieve the state-of-the-art near-linear complexity

[Chen, Owhadi, Schäfer 2023]

Outline

Further Directions and Future Work

High dimensional scientific computing:

- e.g., applications in Chemistry
- Very different to low dimensional PDE setting
- Use randomness to balance exploration and exploitation in high dimensional Gaussian process and kernel methods
  [Chen, Epperly, Tropp, Webber 2022]

Uncertainty quantification and posterior sampling:

- Fully exploit the potential of a Bayesian statistical framework
- Efficient numerical algorithm for sampling?
  [Chen, Huang, Huang, Reich, Stuart 2023]

Summary

Multiscale Numerical Methods:

- Construct specialized basis functions adapted to the equation
- Local structures and global decomposition for Helmholtz's equation (exponential convergence)

Statistical Numerical Methods:

- Flexible Gaussian process framework for general PDE problems
- Convergence, adaptivity, and scalable algorithms for Gaussian process and kernel methods in low and high dimensions
- Further direction: posterior sampling

Goal: enhance specialized and flexible numerical methods rigorously