Multiscale Model Reduction and Sparse Operator Compression for Multiscale Problems

Thomas Y. Hou

Applied and Comput. Math, Caltech

April 3, 2017

Joint work with Qin Li and Pengchuan Zhang

Workshop I: Multiphysics, Multiscale, and Coupled Problems in Subsurface Physics
IPAM Long Program on Computational Issues in Oil Field Applications

Motivations

Dimension reduction appears nearly everywhere in science and engineering.

- Solving elliptic equations with multiscale coefficients: multiscale finite element basis for the elliptic operator.
- Principal component analysis (PCA): principle modes of the covariance operator.
- Quantum chemistry: eigen states of the Hamiltonian.

For computational efficiency and/or good interpretability, localized basis functions are preferred.

- Localized multiscale finite element basis: Babuska-Caloz-Osbron-94, Hou-Wu-1997, Hughes-Feijóo-Mazzei-98, E-Engquist-03, Owhadi-Zhang-07, Målqvist-Peterseim-14, Owhadi-15, etc.
- Sparse principle modes obtained by Sparse PCA or sparse dictionary learning: Zou-Hastie-Tibshirani-04, Witten-Tibshirani-Hastie-09, etc.
- Compressed Wannier modes: Ozoliņš-Lai-Caflisch-Osher-13, E-Li-Lu-10, Lai-Lu-Osher-15, etc.

Schrödinger Equation

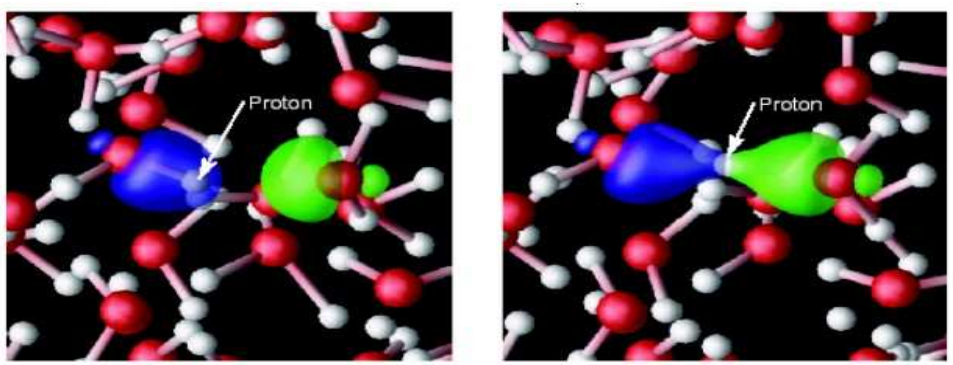
Maximally-localized generalized Wannier functions for composite energy bands

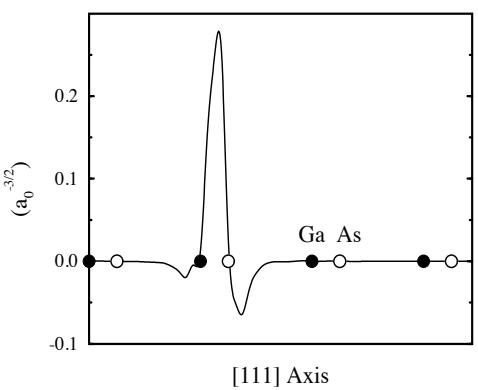
Nicola Marzari and David Vanderbilt
Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08855-0849, USA
(July 10, 1997)

We discuss a method for determining the optimally-localized set of generalized Wannier functions associated with a set of Bloch bands in a crystalline solid. By "generalized Wannier functions" we mean a set of localized orthonormal orbitals spanning the same space as the specified set of Bloch bands. Although we minimize a functional that represents the total spread $\sum_n \langle r^2 \rangle_n - \langle \mathbf{r} \rangle_n^2$ of the Wannier functions in real space, our method proceeds directly from the Bloch functions as represented on a mesh of k-points, and carries out the minimization in a space of unitary matrices $U_{mn}^{(\mathbf{k})}$ describing the rotation among the Bloch bands at each k-point. The method is thus suitable for use in connection with conventional electronic-structure codes. The procedure also returns the total electric polarization as well as the location of each Wannier center. Sample results for Si, GaAs, molecular C_2H_4 and LiCl will be presented.

[1]

$$\mathrm{i}\hbar\partial_t u(t,x) = \underbrace{\left(-\frac{\hbar^2}{2}\Delta_x + V(x)\right)}_{\text{Schrödinger operator: } \mathcal{L}} u \quad \Rightarrow \quad \begin{cases} \mathcal{L}\psi_m = \lambda_m \psi_m \\ u = \sum \alpha_m(t)\psi_m(x) \end{cases}$$





[1] Marzari-Vanderbilt, PRB (56), 97

Methods Based on L^1 Minimization: motivation

- Min-max principle: Ψ with the first d eigenvectors as columns is the optimizer for

  $$\text{minimize} \quad -\text{trace}(\Psi^T \text{Cov} \Psi) \qquad \text{subject to:} \quad \Psi \in \mathbb{R}^{n\times d},\ \Psi^T\Psi = \mathbb{I}_d.$$

- Consider $X = \Psi\Psi^T \in \mathbb{R}^{n\times n}$ and define fantope

  $$\mathcal{F}^d := \{X \in \mathbb{R}^{n\times n} \ : \ X = X^T,\ 0 \preceq X \preceq I_n,\ \text{trace}(X) = d\}.$$

  (Lai-Lu-Osher-14): projection to the first d-dimensional eigenspace is the optimizer for

  $$\text{minimize}_{X\in\mathcal{F}^d} \quad -\text{trace}(\text{Cov}X).$$

- Idea: add L^1 penalty to force Ψ or X sparse.

Methods Based on L^1 Minimization: methods

- Sparse PCA (Zou-Hastie-Tibshirani-04) and Convex relaxation of Sparse PCA (d'Aspremont et al 07, Vu et al 13): Given a symmetric input matrix $S \in \mathbb{R}^{n\times n}$, seek for a d-dimensional sparse principal space estimator $\widehat{X}$ by

$$\text{minimize}_{X\in\mathcal{F}^d} \quad -\text{trace}(SX) + \lambda\|X\|_1$$

  Take the first d eigenvectors of X as the estimated sparse basis elements.

- Compressed modes for variational problems (Ozolins-Lai-Caflisch-Osher-13) and its convex relaxation (Lai-Lu-Osher-14): Given a symmetric input matrix $H \in \mathbb{R}^{n\times n}$, seek for a d-dimensional sparse principal space estimator $\widehat{P}$ by

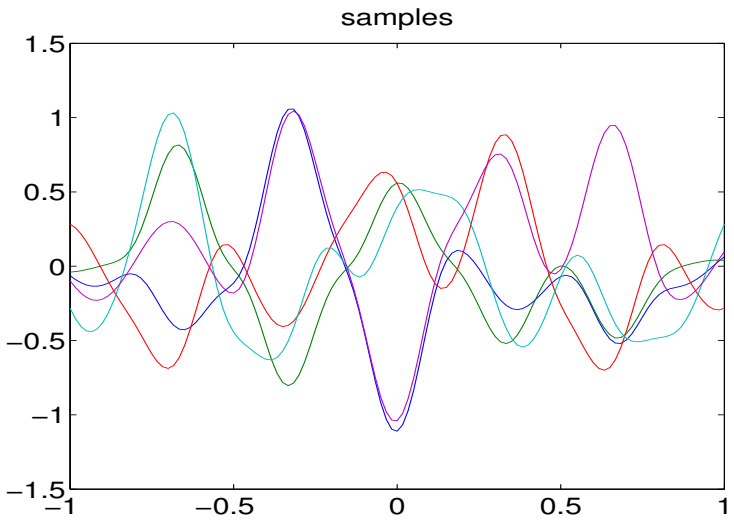
$$\text{minimize}_{P\in\mathcal{F}^d} \quad \text{trace}(HP) + \frac{1}{\mu}\|P\|_1$$

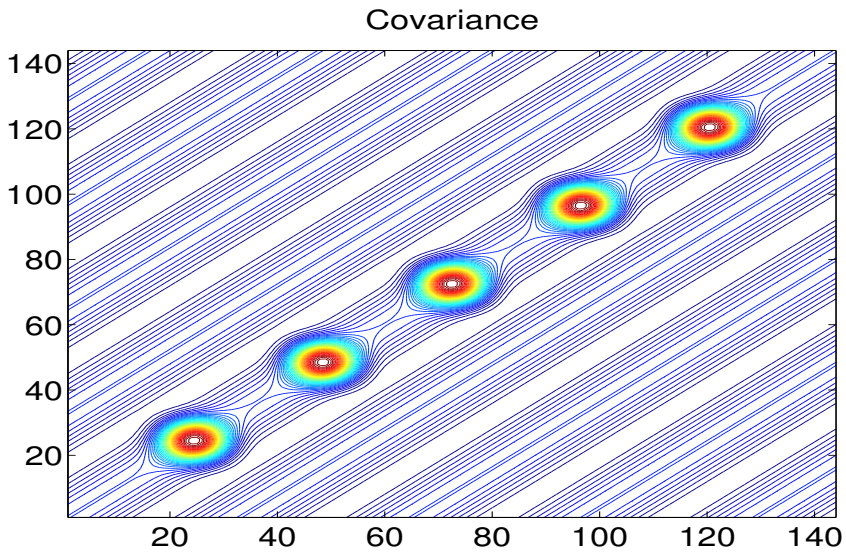
  Take d columns of P as the estimated sparse basis elements.

Random Field Parametrization

$$\kappa(x,\omega) = \mathbb{E}[\kappa(x,\omega)] + \sum_k g_k(x)\eta_k(\omega), \quad \mathbb{E}[\eta_k] = 0,\ \mathbb{E}[\eta_i\eta_j] = \delta_{ij}.$$

What we have:





Q: How to parametrize the random field?

Random Field Parametrization, continued

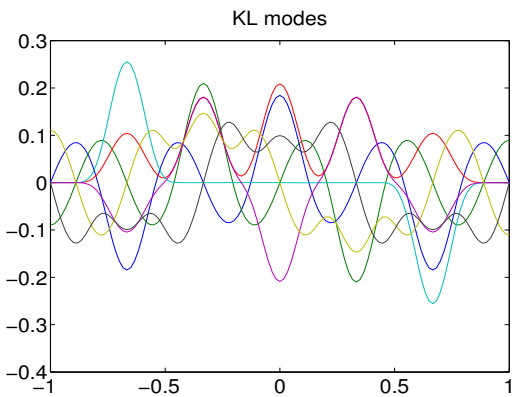
Karhunen-Loéve expansion[3] :

$$\kappa(x,\omega) = \mathbb{E}[\kappa(x,\omega)] + \sum_k \sqrt{\lambda_k} f_k(x)\xi_k$$

$$f_k(x) \ : \ \text{eigenfunctions of Cov}(x,y)$$

Given error tolerance $\epsilon > 0$,

$$\text{minimize} \quad K\,, \quad \text{subject to:} \quad \|\text{Cov} - \sum_{k=1}^{K} \lambda_k f_k(x) f_k(y)\|_2 \leq \epsilon\,.$$



[3] Loéve, 77

Random Field Parametrization, continued

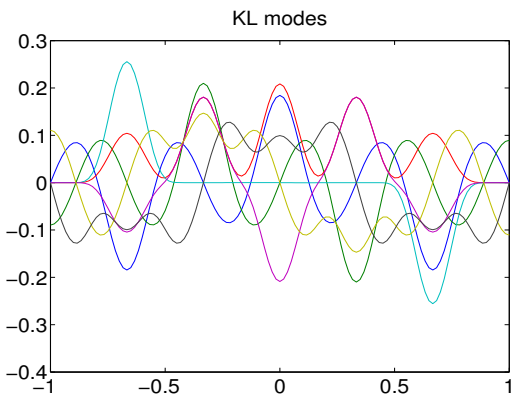
Karhunen-Loéve expansion[3] :

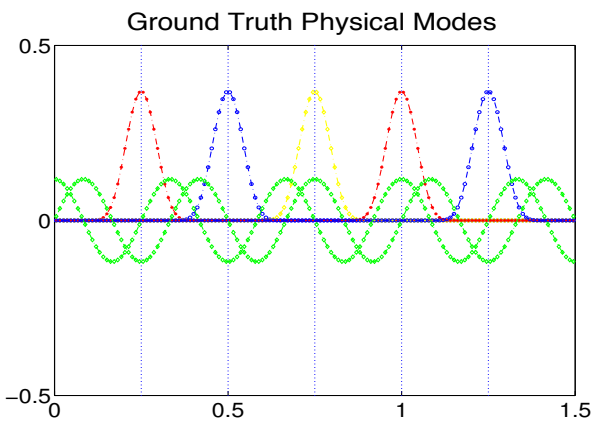
$$\kappa(x,\omega) = \mathbb{E}[\kappa(x,\omega)] + \sum_k \sqrt{\lambda_k} f_k(x)\xi_k$$

$$f_k(x) \ : \ \text{eigenfunctions of Cov}(x,y)$$

Given error tolerance $\epsilon > 0$,

$$\text{minimize} \quad K\,, \quad \text{subject to:} \quad \|\text{Cov} - \sum_{k=1}^{K} \lambda_k f_k(x) f_k(y)\|_2 \leq \epsilon\,.$$





[3] Loéve, 77

I. Intrinsic Sparse Mode Decomposition (ISMD)

Given a low rank symmetric positive semidefinite matrix Cov $\in \mathbb{R}^{N\times N}$,

- Our goal is to decompose Cov into rank-one matrices, i.e.

$$\text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,.$$

- The modes $\{g_k\}_{k=1}^{K}$ are required to be as sparse as possible.

$$\min \quad \sum_{k=1}^{K} \|g_k\|_0 \,, \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,.$$

This minimization problem is extremely difficult:

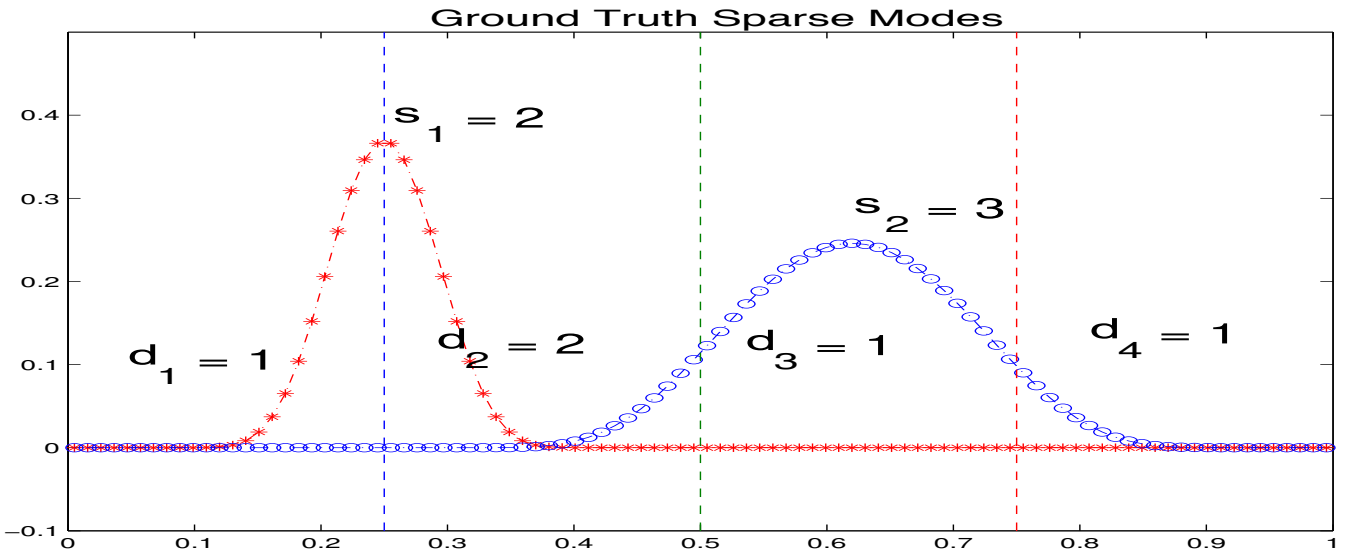
- Number of modes K is unknown, typically $K \sim 100$;
- Each unknown physical mode $g_k(x)$ is represented by a long vector ($g_k \in \mathbb{R}^N$, $N \approx 10,000$);
- l^0 minimization problem with 10^6 unknowns, NP hard!

From l^0 Norm to Patch-wise Sparseness

Definition (Sparseness of a physical mode g)

Given domain partition: $\mathcal{D} = \sqcup_{m=1}^{M} P_m$, define sparseness as

$$s(g) = \#\{P_m : g|_{P_m} \neq 0,\ m = 1, 2, \cdots, M\}.$$



d_m is the local dimension on the m'th patch P_m.

$$\sum_{k=1}^{K} s_k = \sum_{m=1}^{M} d_m$$

Intrinsic Sparse Mode Decomposition

$$\text{minimize} \quad \sum_{k=1}^{K} \|g_k\|_0 \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \ .$$

Intrinsic Sparse Mode Decomposition

$$\text{minimize} \quad \sum_{k=1}^{K} \|g_k\|_0 \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,.$$

$$\Downarrow$$

$$\text{minimize} \quad \sum_{k} s_k = \sum_{m} d_m \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,. \qquad (1)$$

Intrinsic Sparse Mode Decomposition

$$\text{minimize} \quad \sum_{k=1}^{K} \|g_k\|_0 \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,.$$

$$\Downarrow$$

$$\text{minimize} \quad \sum_{k} s_k = \sum_{m} d_m \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,. \qquad (1)$$

Theorem (Theorem 3.5 in [Hou et al., 2017a])

Under the regular sparse assumption (sparse modes are linear independent on every patch), ISMD generates one minimizer of the patch-wise sparseness minimization problem (1).

ISMD details: Construct Local Pieces by Local Rotations

Let $\text{Cov}_{m,m}$ be the local covariance matrix on patch P_m.

1. Want: local sparse physical modes $\boldsymbol{G}_m \equiv [g_{m,1}, g_{m,2}, \ldots, g_{m,d_m}]$

$$\text{Cov}_{m,m} = \boldsymbol{G}_m \boldsymbol{G}_m^T \equiv \sum_{k=1}^{d_m} g_{m,k} g_{m,k}^T.$$

2. Have: local eigen-decomposition $\boldsymbol{H}_m \equiv [h_{m,1}, h_{m,2}, \ldots, h_{m,K_m}]$

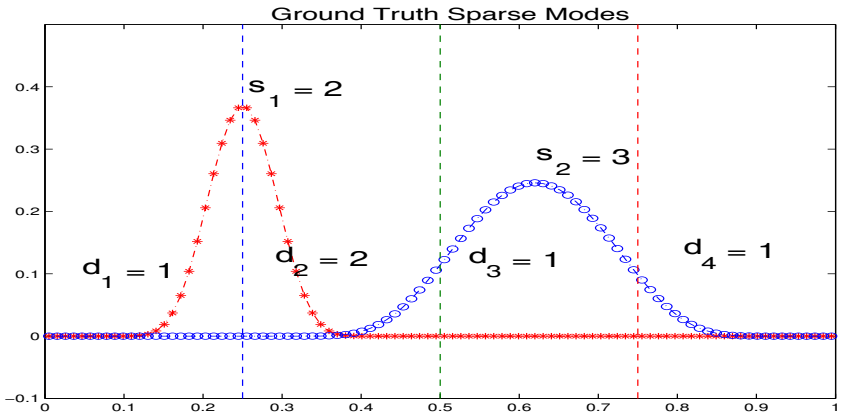
$$\text{Cov}_{m,m} = \boldsymbol{H}_m \boldsymbol{H}_m^T \equiv \sum_{k=1}^{K_m} h_{m,k} h_{m,k}^T.$$

3. $$\boldsymbol{G}_m = \boldsymbol{H}_m D_m, \quad D_m D_m^T = \mathbb{I}_{K_m}.$$

4. Aim: find the right local rotation[4] D_m on every patch P_m!

[4] Regular sparse assumption

ISMD Details: "Patch-up" Local Pieces



$$[g_1, g_2] = \begin{bmatrix} g_{1,1} & g_{2,1} \\ g_{1,2} & g_{2,2} \\ g_{1,3} & g_{2,3} \\ g_{1,4} & g_{2,4} \end{bmatrix} = \begin{bmatrix} g_{1,1} & 0 \\ g_{1,2} & g_{2,2} \\ 0 & g_{2,3} \\ 0 & g_{2,4} \end{bmatrix}$$

$$= \left[\begin{array}{c|cc|c|c} g_{1,1} & 0 & 0 & 0 & 0 \\ \hline 0 & g_{1,2} & g_{2,2} & 0 & 0 \\ \hline 0 & 0 & 0 & g_{2,3} & 0 \\ \hline 0 & 0 & 0 & 0 & g_{2,4} \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ \hline 0 & 1 \\ 0 & 1 \end{array}\right]$$

$$\Rightarrow [g_1, g_2] = \mathsf{diag}(\boldsymbol{G}_1, \boldsymbol{G}_2, \boldsymbol{G}_3, \boldsymbol{G}_4) L$$

Aim: find the right "patch-up" matrix L!

ISMD Details: Flowchart of ISMD

Step 1 Local eigen decomposition: $\text{Cov}_{m,m} = \boldsymbol{H}_m \boldsymbol{H}_m^T$.
Output: $\boldsymbol{H}_{ext} := \text{diag}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_M)$

Step 2 Assemble correlation matrix: $\text{Cov} = \boldsymbol{H}_{ext} \Lambda \boldsymbol{H}_{ext}^T$.
Output: Λ

Step 3 Identify local rotations by joint diagonalization:

$$D_m = \underset{V \in \mathcal{O}(K_m)}{\arg\min} \sum_{n=1}^{M} \sum_{i \neq j} |(V^T \Lambda_{mn} \Lambda_{mn}^T V)_{i,j}|^2 .$$

Output: $D_{ext} \leftarrow \text{diag}(D_1, \cdots, D_M)$

Step 4 "Patch-up" by pivoted Cholesky decomposition

$$D_{ext}^T \Lambda D_{ext} = P L L^T P^T$$

Output: $L \leftarrow PL$.

Step 5 Assemble the intrinsic sparse modes: $[g_1, g_2, \ldots, g_K] \leftarrow \boldsymbol{H}_{ext} D_{ext} L$.

ISMD: Numerical Example 1

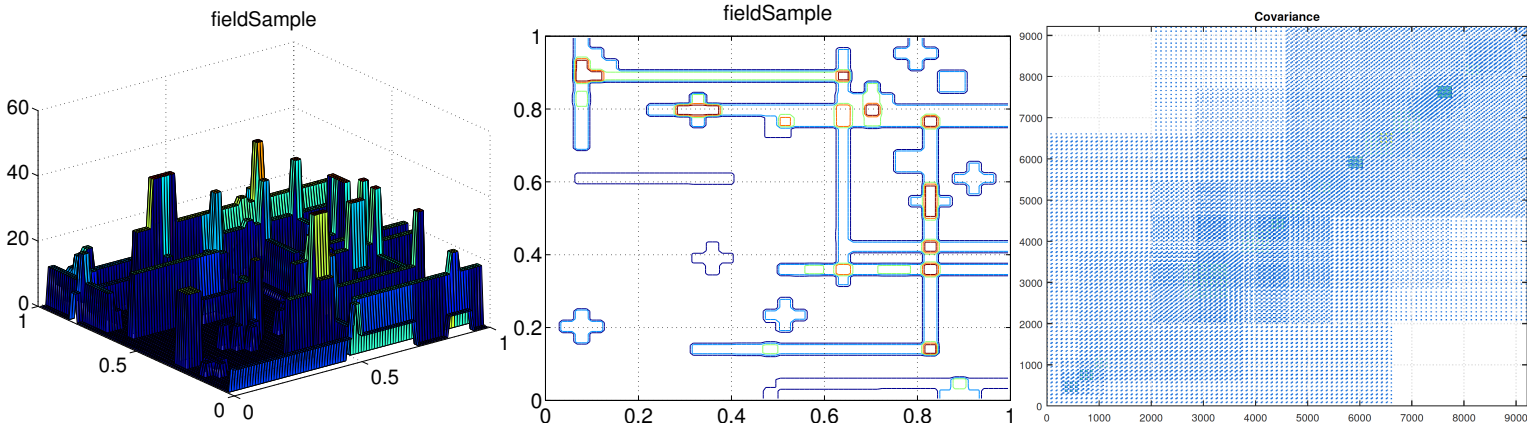


Figure: One sample and the bird's-eye view. The covariance matrix is plotted on the right.

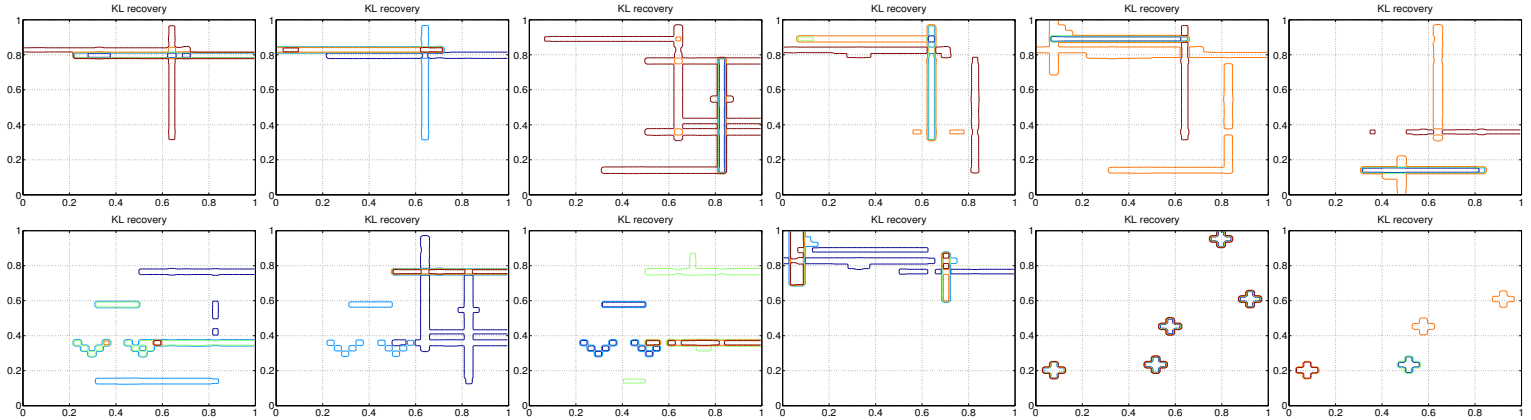


Figure: Eigen-decomposition: 12 out of 35 KL modes. Each mode contains multiple pieces of the ground-truth media.

ISMD: Numerical Example 1 continued

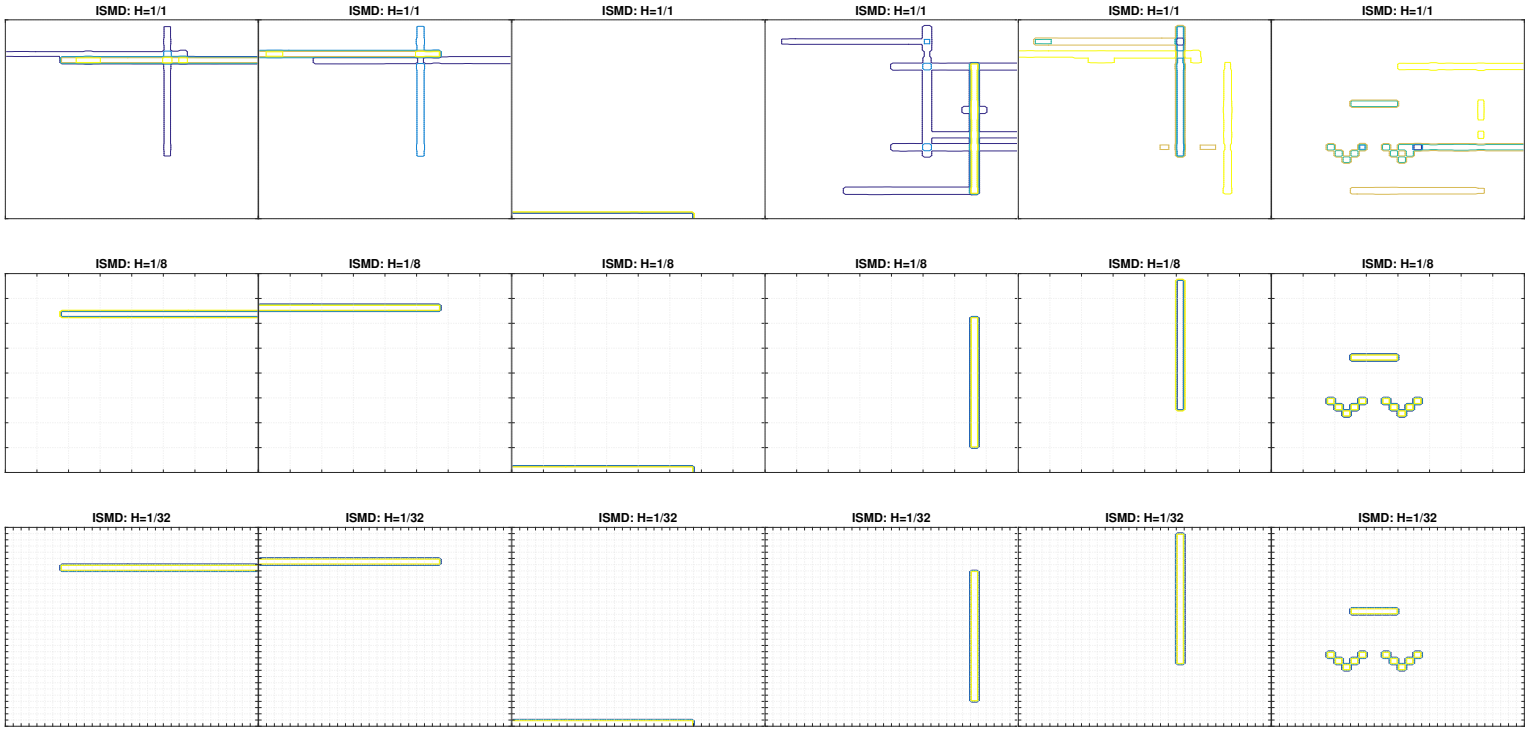


Figure: First 6 eigenvectors (H=1); First 6 intrinsic sparse modes (H=1/8, regular sparse); First 6 intrinsic sparse modes (H=1/32; not regular sparse)

Comments:

- ISMD is proven to give the optimal sparse decomposition under the regular sparse assumption.
- Even when the regular sparse assumption fails (the partition is too fine), ISMD still performs well in our numerical examples.

Other topics about ISMD

- Computational complexity: much cheaper than the eigen-decomposition.

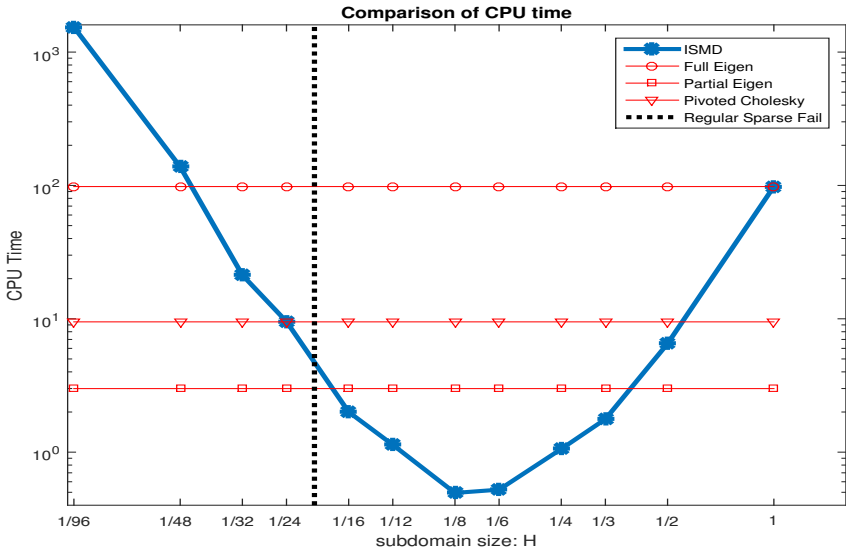


Figure: CPU time (unit: second) for different partition sizes H.

Other topics about ISMD

- Computational complexity: much cheaper than the eigen-decomposition.

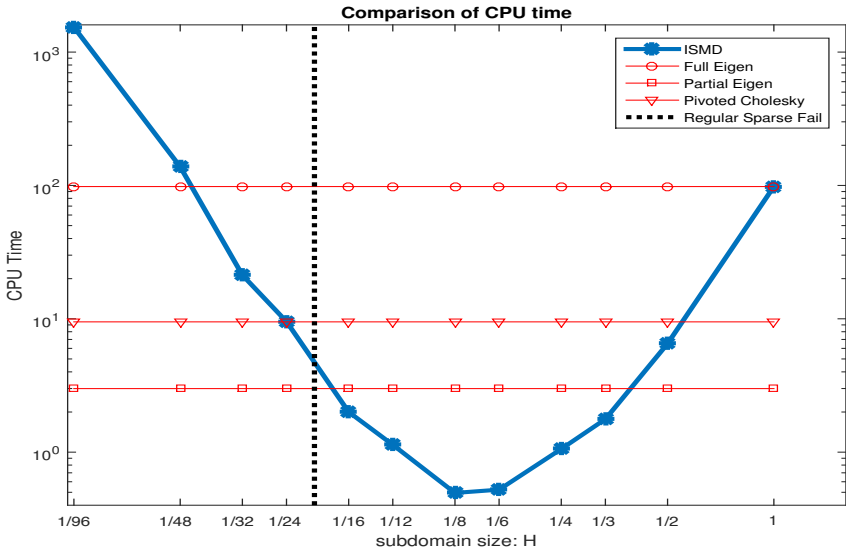


Figure: CPU time (unit: second) for different partition sizes H.

- ISMD is insensitive to the partition of the domain, see Theorem 3.6 in [Hou et al., 2017a].

Other topics about ISMD

- Computational complexity: much cheaper than the eigen-decomposition.

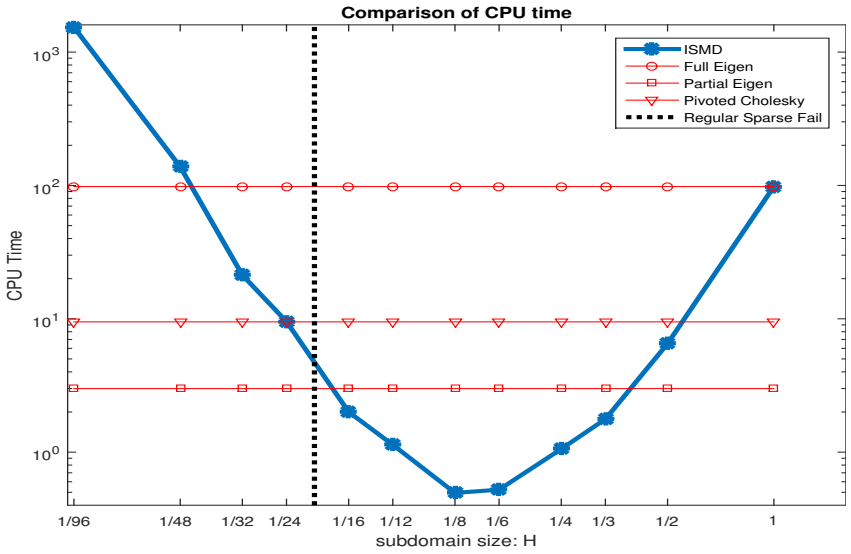


Figure: CPU time (unit: second) for different partition sizes H.

- ISMD is insensitive to the partition of the domain, see Theorem 3.6 in [Hou et al., 2017a].
- Robustness : ISMD is robust against small error, see Lemma 4.1 in [Hou et al., 2017a].

$$\text{Cov} = \sum_{k=1}^{K} g_k g_k^T + \text{Error}.$$

Results from L^1 Minimization

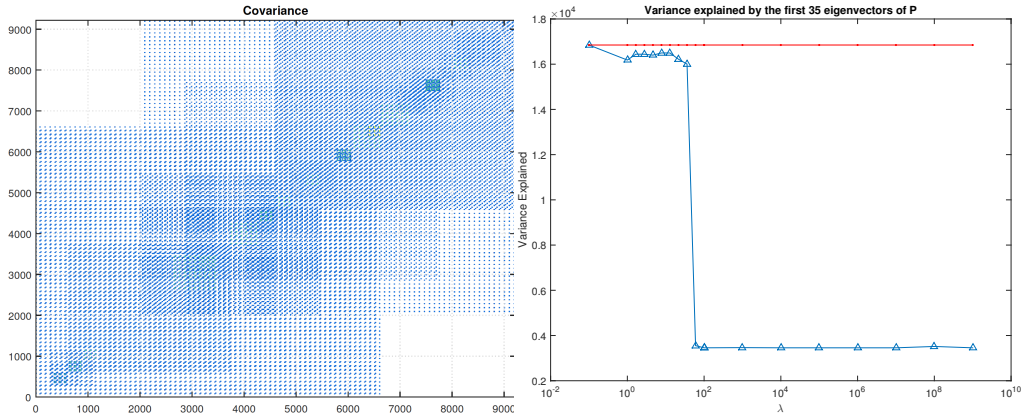


Figure: Covariance Matrix and Variance Explained by the First 35 Eigenvectors of P

Choose $\lambda = 35.94$ for the following result.

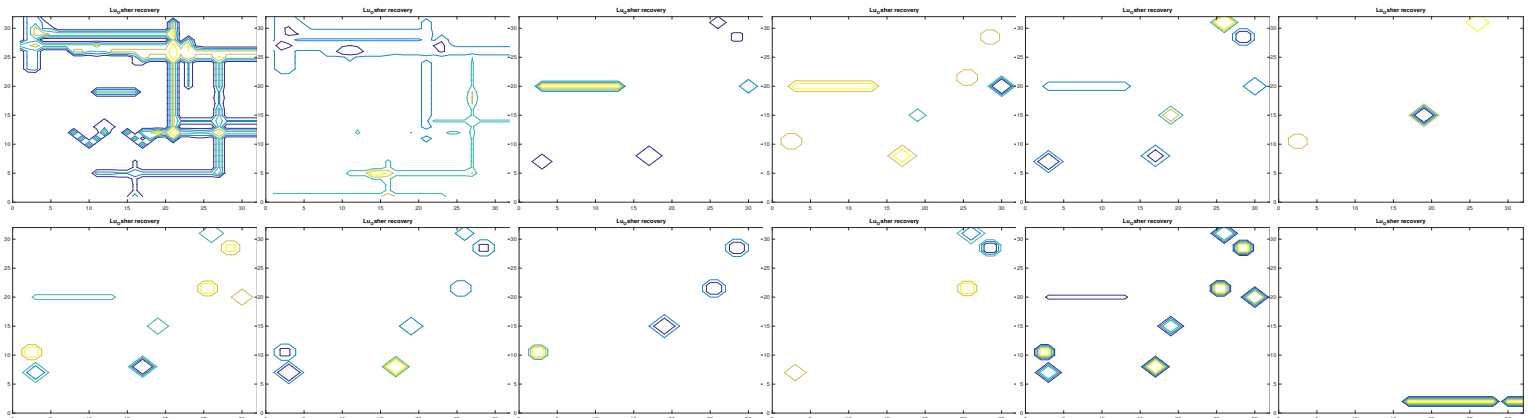


Figure: The First 12 Eigenvectors of P. The first 35 eigenvectors of P explain 95% of the variance.

Results from L^1 Minimization, continued

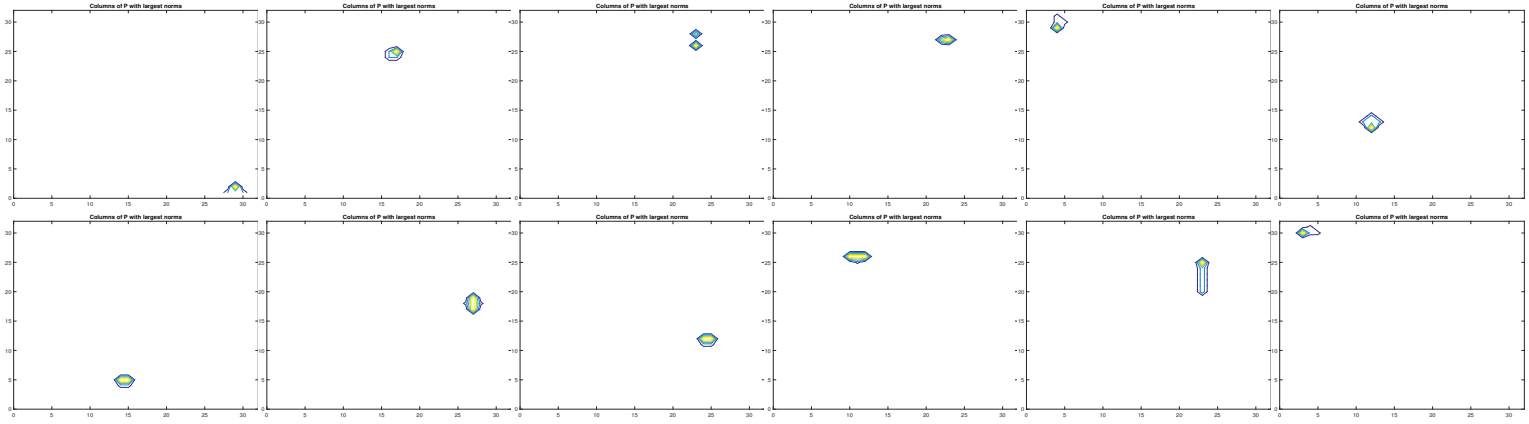


Figure: 12 Columns of P with largest norms. The first 35 columns with largest norms only explain 31.46% of the variance.

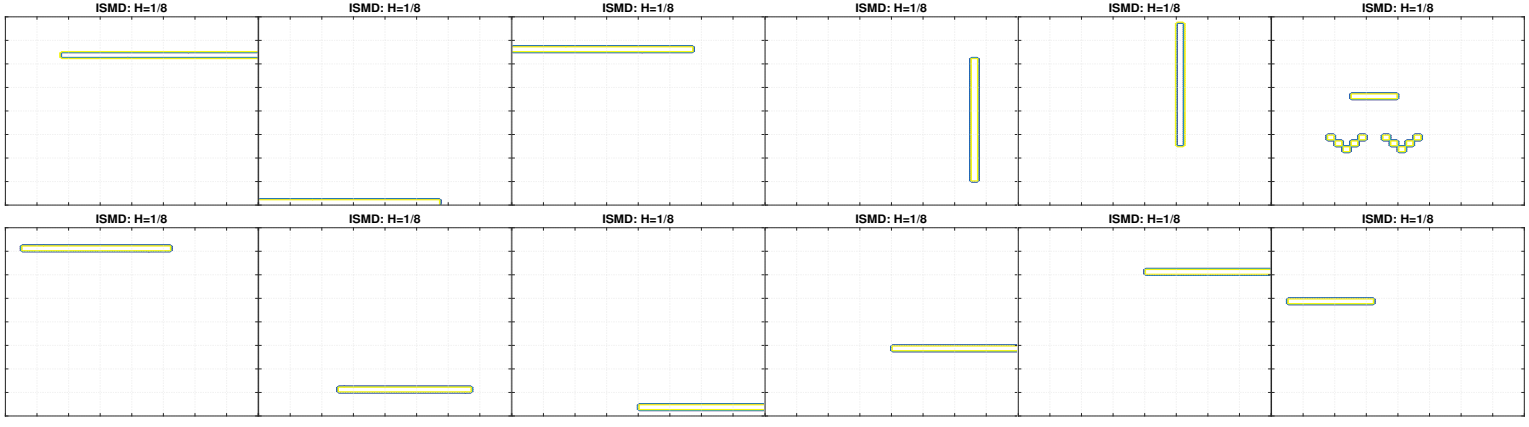


Figure: The First 12 Intrinsic Sparse Modes from ISMD ($H = 1/8$). The first 35 intrinsic sparse modes explain 100% of the variance, exact recovery!

Discussion on Computational Complexity

- The main computational cost in ISMD is the first step: compute local eigen-decomposition, which is about $M\text{Eig}(n/M)$. n is the total number of fine grid nodes in the spatial domain. Here $\text{Eig}(n)$ is the computational cost of eigen-decomposition of a matrix of size n.

- ADMM algorithm is used to solve the semidefinite program in convex relaxation of sparse PCA, and the main computational cost is a global eigen-decomposition $\text{Eig}(n)$ in each iteration. Suppose N iterations are needed to achieve convergence, the computational cost is about $N\text{Eig}(n)$. Partial eigen-decomposition is used to reduced the computational cost.

- For the specific problem we consider here, our algorithm for ISMD takes into account the special structure of the problem, and is more efficient than the ADMM algorithm (sparse PCA). In a typical test of the high contrast example, our method is 652 times faster than the ADMM algorithm.

II. Stochastic Multiscale Model Reduction

$$\text{minimize} \quad \sum_k s_k = \sum_m d_m \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \ ,$$

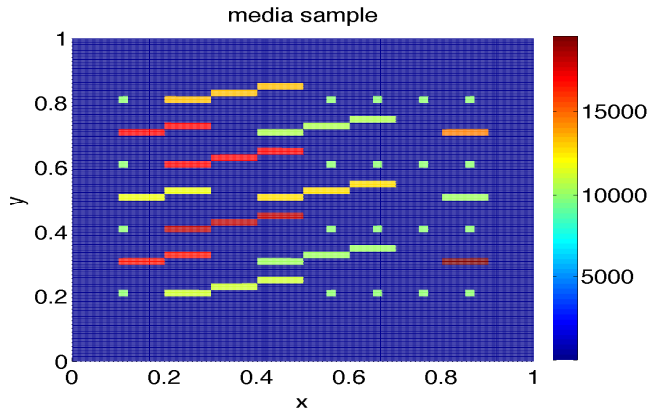


Figure: One sample of permeability field

Random PDEs

$$\begin{cases} -\nabla_x \cdot (\kappa^\epsilon(x,\omega)\nabla_x u(x,\omega)) = b(x), \quad x \in D, \omega \in \Omega, \\ u(x,\omega) = 0, \quad x \in \partial D \end{cases}$$

Random coefficient $\kappa^\epsilon(x,\omega)$ has high dimensionality in the stochastic space and has multiple scales (channels and inclusions) in the physical space.

- Multiple scales in physical space: fine mesh required for standard FEM, large computational cost for one sample.
- Stochastic high dimensional:
  - Monte Carlo : large number of samples needed.
  - gPC based methods, curse of dimensionality! Prohibitively huge number of collocation nodes.
- We propose a stochastic multiscale method that explores the locally low dimensional feature of the solution in the stochastic space and upscales the solution in the physical space simultaneously for all random samples.

Stochastic Multiscale Model Reduction

$$\begin{cases} -\nabla_x \cdot (\kappa^\epsilon(x,\omega)\nabla_x u(x,\omega)) = b(x), \quad x \in D, \omega \in \Omega, \\ u(x,\omega) = 0, \quad x \in \partial D \end{cases}$$

Step 0. Locally low dimensional parametrization of $\kappa^\epsilon(x,\omega)$ by ISMD, local KL expansion, sparse PCA, etc.

Step 1. Offline stage: prepare high-order polynomial interpolants for local upscaled quantities. Various local upscaling methods are available, see e.g. Babuska-Caloz-Osborn 1997, Hou-Wu 1997, Hughes et al 1998, E-Engquist 2003, Owhadi-Zhang 2014, Owhadi-2015.

Step 2. Online stage: for each parameter configuration, interpolate the upscaled quantities from the pre-computed interpolants in Step 1 and solve the upscaled system. Online cost saving is of order $(H/h)^d/(\log(H/h))^k$, where H/h is the ratio between the coarse and the fine gird sizes, d is the physical dimension and k is the local stochastic dimension.

For details, see [Hou et al., 2017b].

A synthetic example with high contrast random medium

$$\kappa(x,\omega) = \xi(\omega) + \sum_{k=1}^{K} g_k(x)\eta_k(\omega)$$

where global variable ξ uniformly distributed in $[0.1, 1.1]$, local variables $\{\eta_k\}_{k=1}^{K}$ are i.i.d. variables uniformly distributed in $[10^4, 2\times 10^4]$.

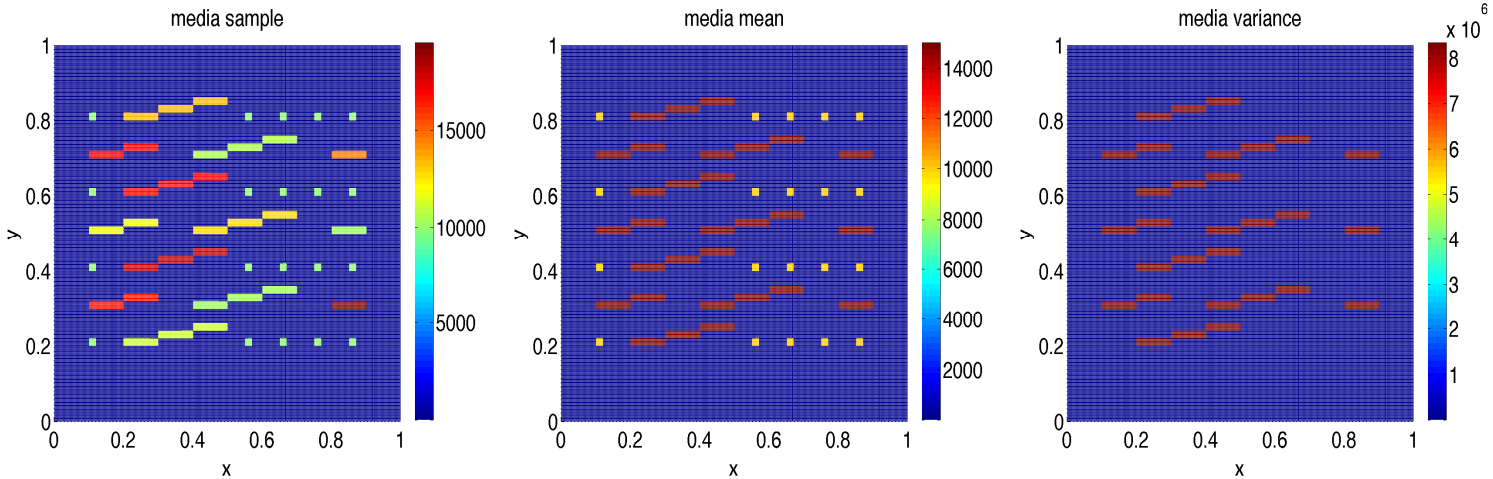


Figure: Left: one sample medium; middle: medium mean; right: medium variance. There are 13 high permeability (of order 10^4) channels in the x direction and a few high permeability inclusions. The background permeability is of order 1.

Step 0: Medium Parametrization by ISMD

$$\text{minimize} \quad \sum_k s_k = \sum_m d_m \,, \quad \text{subject to:} \quad \text{Cov} = \sum_{k=1}^{K} g_k g_k^T \,.$$

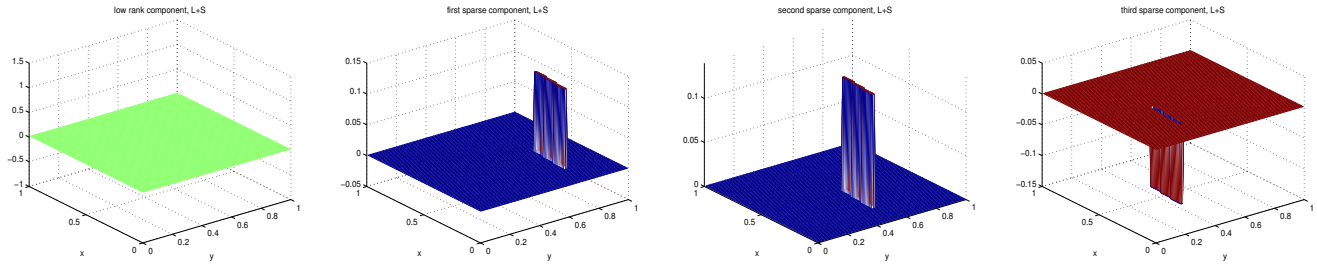


Figure: ISMD: the constant background mode and three channels. The number of local parameters is much less than the total number of parameters.

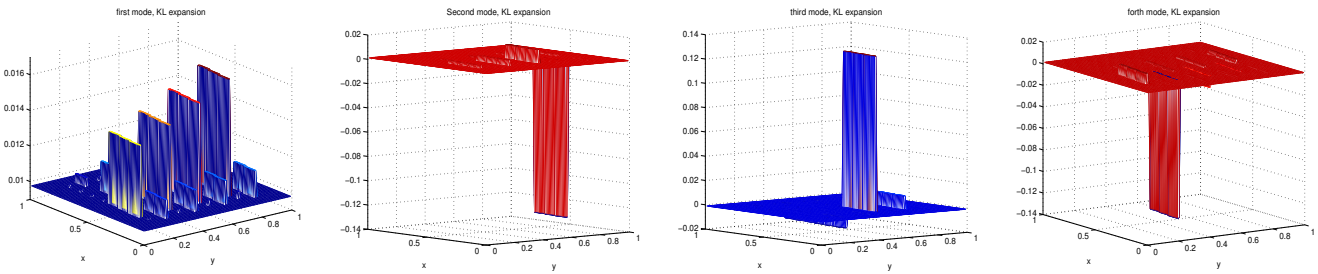


Figure: KL expansion: KL modes mixed the global and local modes together. The number of local parameters equals to the total number of parameters.

Step 1: Local Upscaling

Coarse mesh size: $H = 0.05$, fine mesh size $h = H/20$, oversampling domain size $H_{os} = 3H$.

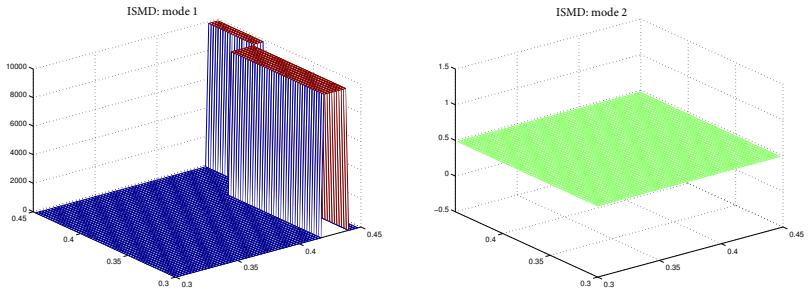


Figure: Two physical modes on patch (14,9) : $d_m = 2$

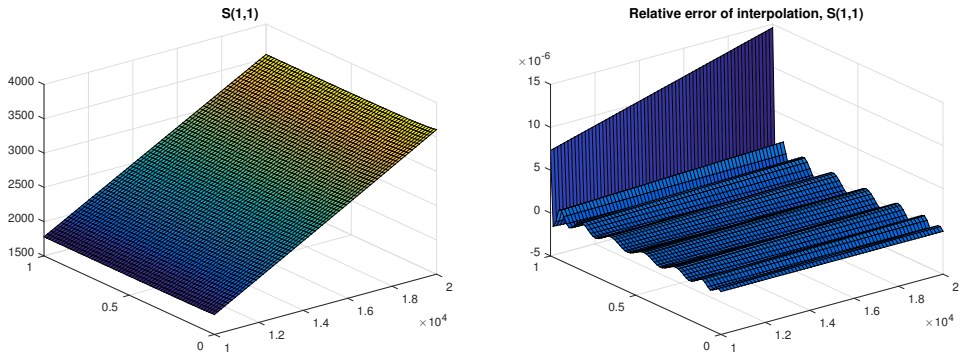


Figure: Left: the (1,1) element of local stiffness matrix on patch (14,9); Right: relative error of surrogate by Chebyshev interpolation on a 5×16 grid. The maximal relative error is of order 10^{-6}.

Step 2: Solve the upscaled system, one sample

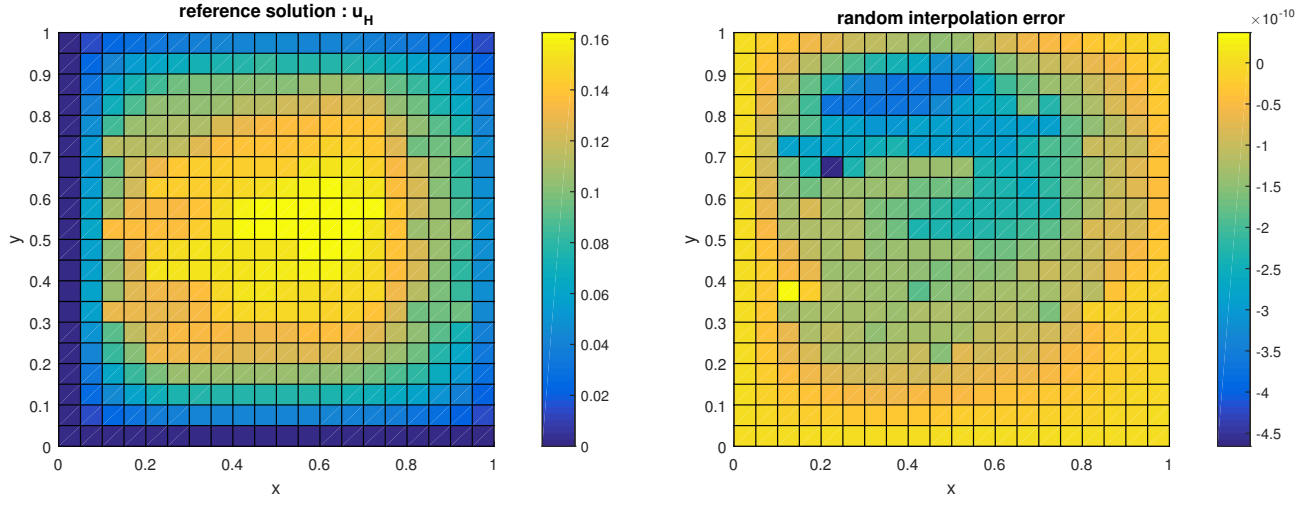


Figure: Left: one sample solution from direct MsFEM; right: solution absolute error due to local interpolation. The error introduced by interpolation is of order 10^{-10}.

Table: Online cost saving for one sample

Direct MsFEM	Local Interpolation	Cost Saving
27.38 (s)	0.0133 (s)	2059

Step 2: Solve the upscaled system, solution statistics

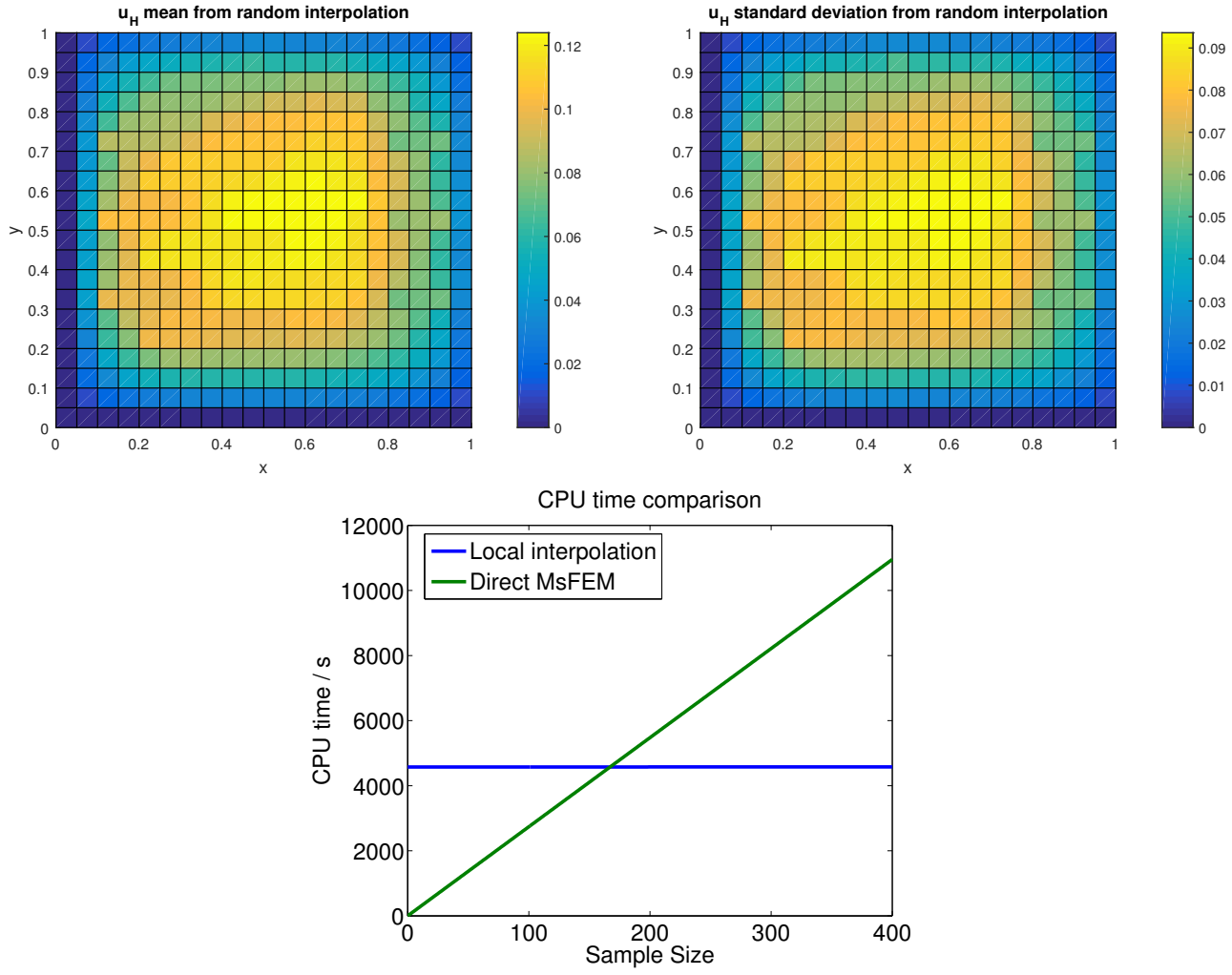


Figure: The statistics of the numerical solutions using 10^5 samples. Top left: mean; top right: standard deviation; bottom: computational cost comparison.

III. Sparse Operator Compression and Concluding Remarks

From ISMD to Sparse Operator Compression

- $\kappa(x,\omega) \approx \sum_k g_k(x)\eta_k(\omega)$, $\{\eta_k\}_{k=1}^K$ uncorrelated or indepedent.
- ISMD

$$\min_{g_1,\ldots,g_K} \quad \sum_{k=1}^K s(g_k),$$

$$\text{s.t.} \quad \text{Cov} \approx \sum_{k=1}^K g_k g_k^T.$$

- Applications in stochastic multiscale mode reduction.

- $\kappa(x,\omega) \approx \sum_k g_k(x)\eta_k(\omega)$, $\{\eta_k\}_{k=1}^K$ can be correlated.
- Sparse Operator Compression

$$\min_{g_1,\ldots,g_K, B\succeq 0} \quad \sum_{k=1}^K s(g_k),$$

$$\text{s.t.} \quad \left\| \text{Cov} - \sum_{k,k'} B_{k,k'} g_k g_k^T \right\| \leq \epsilon$$

- B is the covariance matrix of the factors $\{\eta_k\}_{k=1}^K$.
- More applications in solving deterministic elliptic equations, constructing localized Wannier functions.

For details, see [Hou and Zhang, 2017a, Hou and Zhang, 2017b].

Random field parametrization

The Matérn class covariance

$$K_\nu(x,y) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|x-y|}{\rho} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{|x-y|}{\rho} \right),$$

- Widely used in spatial statistics, geoscience.
- Compress the covariance operator with localized basis functions.
- When $\nu + \frac{d}{2}$ is an integer, we construct nearly optimally localized basis functions $\{g_i\}_{i=1}^n$:

$$|\mathsf{supp}(g_i)| \leq \frac{C_l \log(n)}{n} \quad 1 \leq i \leq n.$$

- We can approximate K_ν by rank-n operator with optimal accuracy:

$$\min_{B \succeq 0} \left\| K_\nu - G B G^T \right\|_2 \leq C_e \lambda_{n+1}(K_\nu).$$

- Sparsity/locality: better interpretability and computational efficiency.

Solving deterministic elliptic equations.

$\mathcal{L}$ is an elliptic operator of order $2k$ ($k \geq 1$) with rough multiscale coefficients in $L^\infty(D)$, and the load $f \in L^2(D)$.

$$\mathcal{L}u = f, \qquad u \in H_0^k(D). \tag{2}$$

- $k = 1$: heat equation, subsurface flow; $k = 2$: beam equation, plate equation, etc...
- Compress the Green's function ($\mathcal{L}^{-1}$) with localized basis functions.
- We construct nearly optimally localized basis functions $\{g_i\}_{i=1}^n \subset H_0^k(D)$. For a given mesh h, we have

$$|\text{supp}(g_i)| \leq C_l h \log(1/h) \quad 1 \leq i \leq n.$$

- The Galerkin finite element solution $u_{ms} := G L_n^{-1} G^T f$ satisfies

$$\|u - u_{ms}\|_H \leq C_e h^k \|f\|_2 \quad \forall f \in L^2(D),$$

  where $\|\cdot\|_H$ is the energy norm, C_e is indep. of small scale of $a_{\sigma\gamma}$.
- Sparsity/locality: computational efficiency.

Our construction and theoretical results

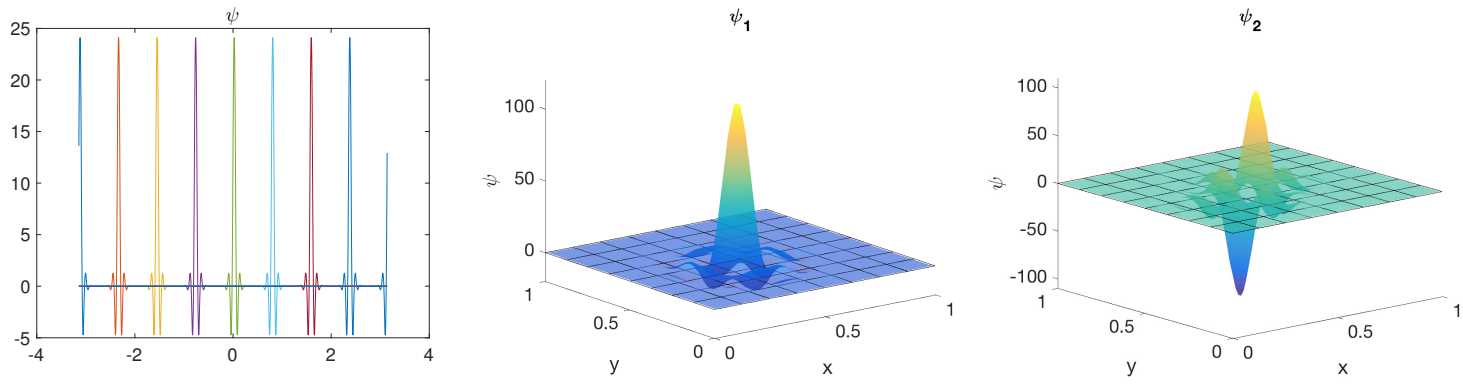


Figure: Left: 8 localized basis functions for $-\Delta$ with periodic BC. Middle and right: 2 localized basis functions for Δ^2 with homogeneous Dirichlet BC.

Our construction of $\{\psi_i^{loc}\}_{i=1}^n$

1. Choose $h > 0$. Partition the physical domain D using a regular partition $\{\tau_i\}_{i=1}^m$ with mesh size h.
2. Choose $r > 0$, say $r = 2h\log(1/h)$. For each patch τ_i, S_r is the union of the subdomains $\tau_{i'}$ intersecting $B(x_i, r)$ (for some $x_i \in \tau_i$).
3. $\mathcal{P}_{k-1}(\tau_i)$ is the space of all d-variate polynomials of degree at most $k-1$ on the patch τ_i. $Q = \binom{k+d-1}{d}$ is its dimension. $\{\varphi_{i,q}\}_{q=1}^Q$ is a set of orthogonal basis functions for $\mathcal{P}_{k-1}(\tau_i)$.

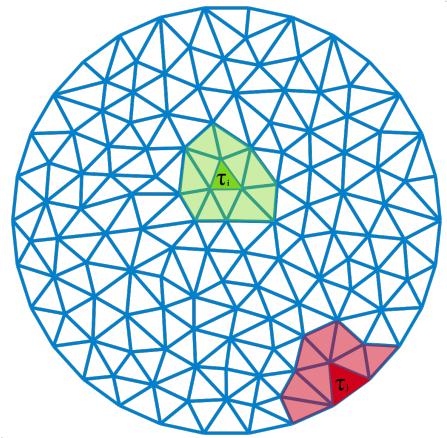


Figure: A regular partition, local patch τ_i and its associated S_r.

$$\psi_{i,q}^{loc} = \arg\min_{\psi \in H_{\mathcal{B}}^k} \quad \|\psi\|_H^2$$

$$\text{s.t.} \quad \int_{S_r} \psi\varphi_{j,q'} = \delta_{iq,jq'}, \quad \forall 1 \le j \le m,\ 1 \le q' \le Q,$$

$$\psi(x) \equiv 0, \quad x \in D \backslash S_r,$$

where $H_{\mathcal{B}}^k$ is the solution space (with some prescribed BC), $\|\cdot\|_H$ is the energy norm associated with $\mathcal{L}$ and the BC.

Our construction $\Psi^{loc} := \{\psi_{i,q}^{loc}\}_{i=1,q=1}^{m,Q}$

Theorem (Hou-Zhang-2016)

Suppose $H^k_{\mathcal{B}} = H^k_0(D)$ *and* $\mathcal{L}u = (-1)^k \sum_{|\sigma|=|\gamma|=k} D^\sigma(a_{\sigma\gamma} D^\gamma u)$. *Assume that* $\mathcal{L}$ *is self-adjoint, positive definite and strongly elliptic, and that there exists* $\theta_{min}, \theta_{max} > 0$ *such that*

$$\theta_{min}\|\xi\|^{2k} \leq \sum_{|\sigma|=|\gamma|=k} a_{\sigma\gamma}\xi^\sigma\xi^\gamma \leq \theta_{max}\|\xi\|^{2k}, \quad \forall \xi \in \mathbb{R}^d.$$

Then for $r \geq C_r h \log(1/h)$, *we have*

1. $$\|\mathcal{L}^{-1}f - \Psi^{loc} L_n^{-1} (\Psi^{loc})^T f\|_H \leq \frac{C_e h^k}{\sqrt{\theta_{min}}}\|f\|_2 \quad \forall f \in L^2(D), \tag{7}$$

   where L_n *is the stiffness matrix under basis functions* Ψ^{loc}.

2. $$E_{oc}(\Psi^{loc}; \mathcal{L}^{-1}) \leq \frac{C_e^2 h^{2k}}{\theta_{min}}. \tag{8}$$

Here, the constant C_r *only depends on the contrast* $\frac{\theta_{max}}{\theta_{min}}$, *and* C_e *is independent of the coefficients.*

Several remarks

- Theorem (Hou-Zhang-2016) also applies to $\mathcal{L}$ with low order terms, i.e. $\mathcal{L}u = (-1)^k \sum_{|\sigma|,|\gamma|\leq k} D^\sigma(a_{\sigma\gamma} D^\gamma u)$.
- Theorem (Hou-Zhang-2016) also applies to other homogeneous boundary conditions, like periodic BC, Robin BC and mixed BC.
- For $H^k_{\mathcal{B}} = H^1_0(D)$, i.e. second order elliptic operators with zero Dirichlet BC, Theorem (Hou-Zhang-2016) have been proved in Owhadi-2015. A similar result for $H^k_{\mathcal{B}} = H^1_0(D)$ was also provided in Målqvist-Peterseim-2014. In this case, Our proof improves the estimates of the constants C_r and C_e.
- For other BCs, operators with lower order terms, and high-order elliptic operators, new techniques and concepts have been developed. Among them, the most important three new techniques are
  - a projection-type polynomial approximation property in $H^k(D)$,
  - the notion of the strong ellipticity [3],
  - an inverse energy estimate for functions in $\Psi := \text{span}\{\psi_{i,q} : 1 \leq i \leq m, 1 \leq q \leq Q\}$.

[3] Equivalent to uniform ellipticity when $d = 1, 2$ or $k = 1$. Slightly stronger than uniform ellipticity in other cases; counter examples exist but difficult to construct.

Roadmap of the proof: Error estimate

Theorem (An error estimate based on projection-type approximation)

Suppose there is a n-dimensional subspace $\Phi \subset L^2(D)$ with basis $\{\varphi_i\}_{i=1}^n$ such that

$$\|u - \mathcal{P}_\Phi^{(L^2)} u\|_{L^2} \leq k_n \|u\|_H \qquad \forall u \in H^k(D). \tag{9}$$

Let Ψ be the n-dimensional subspace in $H^k(D)$ (also in $H^k_{\mathcal{B}}(D)$) spanned by $\{\mathcal{L}^{-1}\varphi_i\}_{i=1}^n$. Then

1. *For any $f \in L^2(D)$ and $u = \mathcal{L}^{-1} f$, we have*

$$\|u - \mathcal{P}_\Psi^{(H^k_{\mathcal{B}})} u\|_H \leq k_n \|f\|_{L_2}. \tag{10}$$

2. *We have*

$$E_{oc}(\Psi; \mathcal{L}^{-1}) \leq k_n^2. \tag{11}$$

- $k = 1$: Φ piecewise constant functions. By the Poincare inequality, it is easy to obtain $\|u - \mathcal{P}_\Phi^{(L^2)} u\|_{L^2} \leq \frac{C_p h}{\sqrt{\theta_{min}}} \|u\|_H$.
- $k \geq 2$: Φ piecewise polynomials with degree no more than $k-1$. By a projection-type polynomial approximation property in $H^k(D)$, see Thm 3.1 in Hou-Zhang-PartII, we have $\|u - \mathcal{P}_\Phi^{(L^2)} u\|_{L^2} \leq \frac{C_p h^k}{\sqrt{\theta_{min}}} \|u\|_H$.

Roadmap of the proof: Error estimate, discussions

Take $H^k_{\mathcal{B}} = H^1_0(D)$ as an example, where Φ is the space of piecewise constant functions.

- Based on a projection-type approximation property, we obtain the error estimates of the GFEM in the energy norm, i.e.

  $$\|u - \mathcal{P}^{(L^2)}_{\Phi} u\|_{L^2} \leq C_{proj} h \|u\|_H \Rightarrow \|u - \mathcal{P}^{(H^1_0)}_{\Psi} u\|_H \leq C_{proj} h \|f\|_{L_2}.$$

  C_{proj} does not depends on the small scales in the coefficients.
- Tranditional interpolation-type estimation requires higher regularity of the solution u: assume $u \in H^2(D)$

  $$|u - \mathcal{I}_h u|_{1,2,D} \leq Ch|u|_{2,2,D} \Rightarrow \|u - \mathcal{I}_h u\|_H \leq C_{interp} h \|f\|_{L_2}.$$

  C_{interp} depends on the small scales in the coefficients.
- Basis functions for $\mathcal{I}_h u$: optimally localized linear nodal basis
  Basis functions for $\mathcal{P}^{(H^1_0)}_{\Psi} u$: global basis functions $\{\mathcal{L}^{-1}\varphi_i\}_{i=1}^n$

Discrete setting: graph Laplacians

$$\mathcal{L}u = -\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x)\frac{\mathrm{d}u}{\mathrm{d}x}\right),$$
$$u(0) = u(1).$$

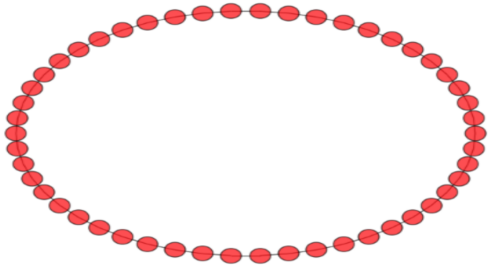


Figure: A 1D circular graph.

$$\mathcal{L}u = -\nabla\cdot(a(x)\nabla u),$$
$$u|_{\partial D} = 0.$$

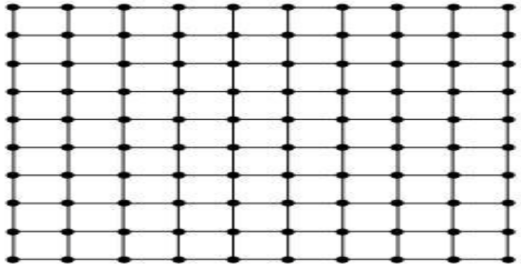


Figure: A 2D lattice graph.

$$\mathcal{L}u = f$$

$\mathcal{L}$: a graph Laplacian

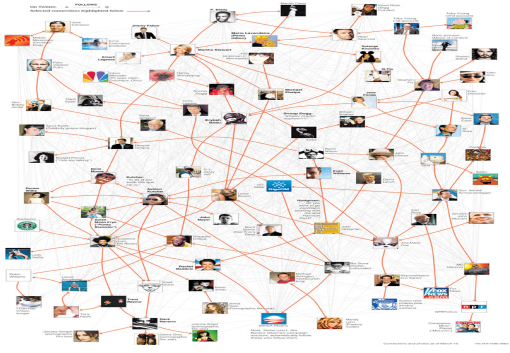


Figure: A social network graph.

- Social networks and transportation networks; genetic data and web pages; spectral clustering of images; electrical resistor circuits; elliptic partial differential equations discretized by finite elements; etc
- Fundamental problems: fast algorithms for $\mathcal{L}u = f$ and eigen decomposition of $\mathcal{L}$.

Discrete setting: graph Laplacians

$$\mathcal{L}u = f$$

- Spielman-Teng (STOC-04, SICOMP-13, SIMAX-14): Nearly-Linear Time Algorithms for Graph Partitioning and Solving Linear Systems
  - Maximal spanning tree, support-graph preconditioners, graph sparsification, etc.
  - Theoretical results, impractical algorithms.
  - Gödel Prize 2008, 2015.
- Livne-Brandt-2012: Lean Algebraic Multigrid. Practical nearly-linear time algorithm, no theoretical guarantee.
- Sparse operator compression for graph Laplacians? The key is an efficient algorithm to find a partition $\{\tau_i\}_{i=1}^m$ of the graph vertices such that

  $$\|u - \mathcal{P}_\Phi^{(L^2)} u\|_{L^2} \leq C_p \sqrt{\lambda_n(\mathcal{L}^{-1})} \|u\|_H,$$

  which is the Poincare inequality on graphs.
- Implementing the sparse operator compression in a multigrid manner leads to a nearly-linear time algorithm.

Concluding Remarks

1. We propose an intrinsic sparse mode decomposition method (ISMD) to decompose a symmetric positive semidefinite matrix into a finite sum of sparse rank one components. We apply ISMD to the covariance matrix of a random field and obtain a locally low dimensional parametrization of the random field.
2. Based on ISMD-like locally low dimensional parametrization methods, we propose the Stochastic Multiscale Finite Element Method (StoMsFEM). It can significantly reduce the computational cost for every single evaluation of the solution sample.
3. Our recent work Sparse Operator Compression shows that many symmetric positive semidefinite operator can be localized, for example, the Green's function of elliptic operators and the Matérn class kernels.
4. Representing/approximating an operator with localized basis functions has potential applications in solving deterministic elliptic equations with rough coefficients and constructing localized orbits for Hamiltonians in quantum physics.


Hou, T. Y., Li, Q., and Zhang, P. (2017a).
A sparse decomposition of low rank symmetric positive semidefinite matrices.
Multiscale Modeling & Simulation, 15(1):410–444.


Hou, Y. T., Li, Q., and Zhang, P. (2017b).
Exploring the locally low dimensional structure in solving random elliptic PDEs.
Multiscale Modeling & Simulation.


Hou, Y. T. and Zhang, P. (2017a).
Sparse operator compression of elliptic operators – Part I : Second order elliptic operators.
Research in the Mathematical Sciences.


Hou, Y. T. and Zhang, P. (2017b).
Sparse operator compression of elliptic operators – Part II : High order elliptic operators.
Research in the Mathematical Sciences.