Random Sketching for Model Order Reduction

MoRePas IV, 11th April 2018, Nantes, France

Oleg Balabanov
joint work with Anthony Nouy

Centrale Nantes, GeM (UMR 6183), LMJL(UMR 6629), France

Summary

Classical RB and POD methods involve the following computations:

- **Offline** (computed once):
  evaluating solution samples (snapshots)
- **Online** (for each parameter value):
  solving reduced systems of equations

Summary

Classical RB and POD methods involve the following computations:

- **Offline** (computed once):
  evaluating solution samples (snapshots),
  computing low-rank approximation of snapshot matrix for POD,
  evaluating numerous inner products.
- **Online** (for each parameter value):
  solving reduced systems of equations

Summary

Classical RB and POD methods involve the following computations:

- **Offline** (computed once):
  evaluating solution samples (snapshots),
  computing low-rank approximation of snapshot matrix for POD,
  evaluating numerous inner products.
- **Online** (for each parameter value):
  solving reduced systems of equations,
  evaluating the reduced quantities from their affine decompositions.

Summary

Classical RB and POD methods involve the following computations:

- **Offline** (computed once):
  evaluating solution samples (snapshots),
  computing low-rank approximation of snapshot matrix for POD,
  evaluating numerous inner products.
- **Online** (for each parameter value):
  solving reduced systems of equations,
  evaluating the reduced quantities from their affine decompositions.

We propose a probabilistic way for reducing the cost of all the computations besides solving linear systems.

Summary

Classical RB and POD methods involve the following computations:

- **Offline** (computed once):
  evaluating solution samples (snapshots),
  computing low-rank approximation of snapshot matrix for POD,
  evaluating numerous inner products.
- **Online** (for each parameter value):
  solving reduced systems of equations,
  evaluating the reduced quantities from their affine decompositions.

We propose a probabilistic way for reducing the cost of all the computations besides solving linear systems.

Our methodology can be beneficial in any computational environment.

Outline

Classical projection-based MOR

ℓ_2 embeddings

Random Sketching for MOR

Numerical experiments

Outline

Parameter-dependent problem

Parameter-dependent system of equations

Let $\xi \in \Xi$ denote the parameters. Find $\mathbf{u}(\xi) \in U$ such that

$$\mathbf{A}(\xi)\mathbf{u}(\xi) = \mathbf{b}(\xi),$$

where $\mathbf{A}(\xi) : U \to U'$ and $\mathbf{b}(\xi) \in U'$.

Parameter-dependent problem

Parameter-dependent system of equations

Let $\xi \in \Xi$ denote the parameters. Find $\mathbf{u}(\xi) \in U$ such that

$$\mathbf{A}(\xi)\mathbf{u}(\xi) = \mathbf{b}(\xi),$$

where $\mathbf{A}(\xi) : U \to U'$ and $\mathbf{b}(\xi) \in U'$.

Output quantity

Let $\mathbf{l}(\xi) \in U'$ be the extractor of a quantity of interest.

$$s(\xi) := \langle \mathbf{l}(\xi), \mathbf{u}(\xi) \rangle.$$

Classical projection-based MOR

Galerkin projection

$\mathbf{u}(\xi)$ is approximated by its projection $\mathbf{u}_r(\xi)$ onto r-dimensional subspace $U_r \subseteq U$ defined by

$$\langle \mathbf{r}(\mathbf{u}_r(\xi); \xi), \mathbf{v} \rangle = 0, \ \forall \mathbf{v} \in U_r,$$

where $\mathbf{r}(\mathbf{x}; \xi) := \mathbf{b}(\xi) - \mathbf{A}(\xi)\mathbf{x}$.

Classical projection-based MOR

Galerkin projection

$\mathbf{u}(\xi)$ is approximated by its projection $\mathbf{u}_r(\xi)$ onto r-dimensional subspace $U_r \subseteq U$ defined by

$$\langle \mathbf{r}(\mathbf{u}_r(\xi);\xi), \mathbf{v} \rangle = 0, \ \forall \mathbf{v} \in U_r,$$

where $\mathbf{r}(\mathbf{x};\xi) := \mathbf{b}(\xi) - \mathbf{A}(\xi)\mathbf{x}$.

Error estimation/certification

For $\mathbf{u}_r^*(\xi) \in U_r$,

$$\|\mathbf{u}(\xi) - \mathbf{u}_r^*(\xi)\|_U \leq \Delta_r(\mathbf{u}_r^*(\xi);\xi) := \frac{\|\mathbf{r}(\mathbf{u}_r^*(\xi);\xi)\|_{U'}}{\eta(\xi)}.$$

Classical projection-based MOR

Assume, $\mathbf{A}(\xi)$ and $\mathbf{b}(\xi)$ are parameter-separable with m_A and m_b terms:

$$\mathbf{A}(\xi) = \sum_{i=1}^{m_A} \mathbf{A}_i \phi_i(\xi), \qquad \mathbf{b}(\xi) = \sum_{i=1}^{m_b} \mathbf{b}_i \theta_i(\xi)$$

Offline computational cost

- The snapshots can be computed with a standard solver on a server or multiple workstations. r snapshots: $\mathcal{O}(rn \log n)$ time.
- Evaluating numerous inner products between high-dimensional vectors: $\mathcal{O}(nr^2 m_A^2 + nm_b^2)$ flops.
- Many passes over large data sets.
- For distributed computing, extremely high amount of communication between machines.

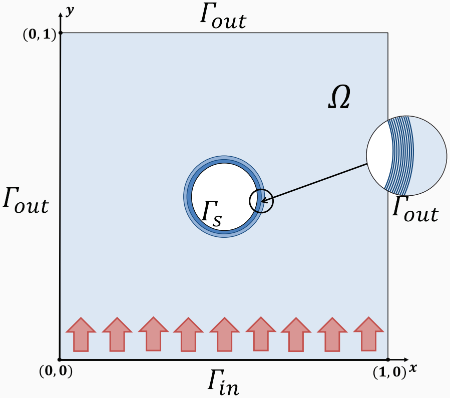
Classical projection-based MOR

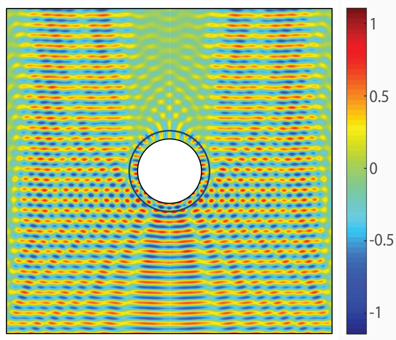
Online computational cost

- Evaluating the reduced quantities from their affine expansions.
  A reduced system: $\mathcal{O}(rm_A + m_b)$ time, the residual norm: $\mathcal{O}(r^2 m_A^2 + m_b^2)$ time.
- Solving a reduced system of equations: $\mathcal{O}(r^2 \log r)$ flops with iterative solver.

The cost of evaluating inner products between high-dimensional vectors is dominant for both offline and online stages.

Illustration





We consider the following Helmholtz equation:

$$\Delta u + \kappa^2 u = 0,$$

with first order absorbing b.c.'s and wave initialization on Γ_{in}.

The background has $\kappa = \kappa_0 := 50$. The cloak consists of 10 layers.

The i-th layer has $\kappa = \kappa_i$. Define $\xi := (\kappa_1, ..., \kappa_{10}) \in [\kappa_0, \sqrt{2}\kappa_0]^{10} := \Xi$.

Illustration

Discretization: $n \approx 200000$ complex degrees of freedom. $r = 150$ iterations of the classical greedy algorithm were performed. We chose $\eta(\xi) = 1$.

offline total	150 snapshots	online solution	reduced quantities
5402s	336s	0.6ms	1.9ms

Table 1: CPU times

- Computing the snapshots occupied only 6% of the overall runtime of the classical greedy algorithm.
- Evaluating the residual terms from the precomputed affine expansions has the dominant cost in the online stage.

Outline

ε-subspace embeddings

Consider a discrete setting $U := \mathbb{K}^n$, with a "natural" inner product for U:

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle$$

where $\mathbf{R}_U : U \to U'$ is self-adjoint positive definite matrix.

ε-subspace embeddings

Consider a discrete setting $U := \mathbb{K}^n$, with a "natural" inner product for U:

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle$$

where $\mathbf{R}_U : U \to U'$ is self-adjoint positive definite matrix.

$\langle \cdot, \cdot \rangle_U$ can be extremely expensive to operate with.

ε-subspace embeddings

Consider a discrete setting $U := \mathbb{K}^n$, with a "natural" inner product for U:

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle$$

where $\mathbf{R}_U : U \to U'$ is self-adjoint positive definite matrix.

$\langle \cdot, \cdot \rangle_U$ can be extremely expensive to operate with.

Construction of a reduced model requires evaluation of inner products only between vectors lying in low-dimensional subspaces of U.

ε-subspace embeddings

Consider a discrete setting $U := \mathbb{K}^n$, with a "natural" inner product for U:

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle$$

where $\mathbf{R}_U : U \to U'$ is self-adjoint positive definite matrix.

$\langle \cdot, \cdot \rangle_U$ can be extremely expensive to operate with.

Construction of a reduced model requires evaluation of inner products only between vectors lying in low-dimensional subspaces of U.

For these vectors, $\langle \cdot, \cdot \rangle_U$ can be efficiently approximated by

$$\langle \cdot, \cdot \rangle_U^{\mathbf{\Theta}} := \langle \mathbf{\Theta} \cdot, \mathbf{\Theta} \cdot \rangle$$

where $\mathbf{\Theta} \in \mathbb{K}^{k \times n}$, with $k \ll n$.

ε-subspace embeddings

Consider a discrete setting $U := \mathbb{K}^n$, with a "natural" inner product for U:

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle$$

where $\mathbf{R}_U : U \to U'$ is self-adjoint positive definite matrix.

$\langle \cdot, \cdot \rangle_U$ can be extremely expensive to operate with.

Construction of a reduced model requires evaluation of inner products only between vectors lying in low-dimensional subspaces of U.

For these vectors, $\langle \cdot, \cdot \rangle_U$ can be efficiently approximated by

$$\langle \cdot, \cdot \rangle_U^{\mathbf{\Theta}} := \langle \mathbf{\Theta} \cdot, \mathbf{\Theta} \cdot \rangle$$

where $\mathbf{\Theta} \in \mathbb{K}^{k \times n}$, with $k \ll n$.

Let V be a subspace of U. If

$$\forall \mathbf{x}, \mathbf{y} \in V, \ \left| \langle \mathbf{x}, \mathbf{y} \rangle_U - \langle \mathbf{x}, \mathbf{y} \rangle_U^{\mathbf{\Theta}} \right| \leq \varepsilon \|\mathbf{x}\|_U \|\mathbf{y}\|_U,$$

then $\mathbf{\Theta}$ is called $U \to \ell_2$ ε-subspace embedding for V.

Data-oblivious embeddings

$\boldsymbol{\Theta}$ is a realization of a probability distribution over matrices.

$\boldsymbol{\Theta}$ is called a (ε, δ, d) oblivious $U \to \ell_2$ subspace embedding if for any d-dimensional subspace V of U

$$\mathbb{P}\left(\boldsymbol{\Theta} \text{ is a } U \to \ell_2 \ \varepsilon\text{-subspace embedding for } V\right) \geq 1 - \delta.$$

Data-oblivious embeddings

$\mathbf{\Theta}$ is a realization of a probability distribution over matrices.

$\mathbf{\Theta}$ is called a (ε, δ, d) oblivious $U \to \ell_2$ subspace embedding if for any d-dimensional subspace V of U

$$\mathbb{P}\left(\mathbf{\Theta} \text{ is a } U \to \ell_2 \ \varepsilon\text{-subspace embedding for } V\right) \geq 1 - \delta.$$

Classical oblivious $\ell_2 \to \ell_2$ ε-subspace embeddings:

- The rescaled Gaussian and Rademacher distributions if $k \geq 7.87\varepsilon^{-2}(6.9d + \log(1/\delta))$ for $\mathbb{K} = \mathbb{R}$ or $k \geq 7.87\varepsilon^{-2}(13.8d + \log(1/\delta))$ for $\mathbb{K} = \mathbb{C}$
- The partial Subsampled Randomized Hadamard Transform (P-SRHT) $k \geq 6\varepsilon^{-2}\left[\sqrt{d} + \sqrt{8\log(6n/\delta)}\right]^2 \log(3d/\delta)$ (for $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$)

The lower bounds are independent or only weakly (logarithmically) dependent on the dimension n and the probability of failure δ.

Data-oblivious embeddings

Oblivious $U \to \ell_2$ subspace embedding $\mathbf{\Theta}$ can be built from classical $\ell_2 \to \ell_2$ subspace embeddings:

$$\mathbf{\Theta} = \mathbf{\Omega}\mathbf{Q},$$

where $\mathbf{Q} \in \mathbb{K}^{s \times n}$ is rectangular matrix such that $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{R}_U$.

Data-oblivious embeddings

Oblivious $U \to \ell_2$ subspace embedding $\boldsymbol{\Theta}$ can be built from classical $\ell_2 \to \ell_2$ subspace embeddings:

$$\boldsymbol{\Theta} = \boldsymbol{\Omega}\mathbf{Q},$$

where $\mathbf{Q} \in \mathbb{K}^{s \times n}$ is rectangular matrix such that $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{R}_U$.

Computational aspects

- The rescaled Rademacher distribution can be efficiently implemented using standard SQL primitives.
- The P-SRHT has a hierarchical structure needing just $\mathcal{O}(n \log k)$ flops for multiplication by a vector.
- The products with Gaussian or Rademacher matrices are embarrassingly parallel.
- The product $\boldsymbol{\Omega}\mathbf{Q}$ should not be evaluated explicitly.
- The random sequences can be generated using a seeded random number generator with negligible communication (for parallel and distributed computing) and storage costs.

Outline

Sketched Galerkin projection

Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|_{U_r'} = 0, \qquad \|\mathbf{w}\|_{U_r'} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w},\mathbf{x}\rangle_U|}{\|\mathbf{x}\|_U}.$$

Sketched Galerkin projection

Sketched Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|^{\boldsymbol{\Theta}}_{U'_r} = 0, \qquad \|\mathbf{w}\|^{\boldsymbol{\Theta}}_{U'_r} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w}, \mathbf{x}\rangle^{\boldsymbol{\Theta}}_U|}{\|\mathbf{x}\|^{\boldsymbol{\Theta}}_U}.$$

Sketched Galerkin projection

Sketched Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|^{\Theta}_{U_r'} = 0, \qquad \|\mathbf{w}\|^{\Theta}_{U_r'} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w}, \mathbf{x}\rangle^{\Theta}_U|}{\|\mathbf{x}\|^{\Theta}_U}.$$

Quasi-optimality of Galerkin projection

$$\|\mathbf{u}(\xi) - \mathbf{u}_r(\xi)\|_U \leq (1 + \frac{\beta_r(\xi)}{\alpha_r(\xi)})\|\mathbf{u}(\xi) - \mathbf{P}_{U_r}\mathbf{u}(\xi)\|_U.$$

$$\alpha_r(\xi) := \min_{\mathbf{x}\in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|_{U_r'}}{\|\mathbf{x}\|_U}, \qquad \beta_r(\xi) := \max_{\mathbf{x}\in \mathrm{span}\{\mathbf{u}(\xi)\}+U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|_{U_r'}}{\|\mathbf{x}\|_U}$$

Sketched Galerkin projection

Sketched Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|^{\Theta}_{U_r'} = 0, \qquad \|\mathbf{w}\|^{\Theta}_{U_r'} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w}, \mathbf{x}\rangle^{\Theta}_U|}{\|\mathbf{x}\|^{\Theta}_U}.$$

Quasi-optimality of sketched Galerkin projection

$$\|\mathbf{u}(\xi) - \mathbf{u}_r(\xi)\|_U \leq (1 + \frac{\beta_r^{\Theta}(\xi)}{\alpha_r^{\Theta}(\xi)})\|\mathbf{u}(\xi) - \mathbf{P}_{U_r}\mathbf{u}(\xi)\|_U.$$

$$\alpha_r^{\Theta}(\xi) := \min_{\mathbf{x}\in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\Theta}_{U_r'}}{\|\mathbf{x}\|_U}, \qquad \beta_r^{\Theta}(\xi) := \max_{\mathbf{x}\in \mathrm{span}\{\mathbf{u}(\xi)\}+U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\Theta}_{U_r'}}{\|\mathbf{x}\|_U}$$

Sketched Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|^{\Theta}_{U_r'} = 0, \qquad \|\mathbf{w}\|^{\Theta}_{U_r'} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w}, \mathbf{x}\rangle^{\Theta}_U|}{\|\mathbf{x}\|^{\Theta}_U}.$$

Quasi-optimality of sketched Galerkin projection

$$\|\mathbf{u}(\xi) - \mathbf{u}_r(\xi)\|_U \leq (1 + \frac{\beta_r^{\Theta}(\xi)}{\alpha_r^{\Theta}(\xi)})\|\mathbf{u}(\xi) - \mathbf{P}_{U_r}\mathbf{u}(\xi)\|_U.$$

$$\alpha_r^{\Theta}(\xi) := \min_{\mathbf{x}\in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\Theta}_{U_r'}}{\|\mathbf{x}\|_U}, \qquad \beta_r^{\Theta}(\xi) := \max_{\mathbf{x}\in \mathrm{span}\{\mathbf{u}(\xi)\}+U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\Theta}_{U_r'}}{\|\mathbf{x}\|_U}$$

For any $\mathbf{x} \in U_r$ the residual $\mathbf{r}(\mathbf{x};\xi)$ belongs to $Y_r(\xi)' := \mathbf{R}_U Y_r(\xi)$, with

$$Y_r(\xi) := U_r + \mathrm{span}\{\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{x} : \mathbf{x} \in U_r\} + \mathrm{span}\{\mathbf{R}_U^{-1}\mathbf{b}(\xi)\}.$$

Sketched Galerkin projection

Sketched Galerkin projection

$$\|\mathbf{r}(\mathbf{u}_r(\xi);\xi)\|^{\mathbf{\Theta}}_{U_r'} = 0, \qquad \|\mathbf{w}\|^{\mathbf{\Theta}}_{U_r'} := \max_{\mathbf{x}\in U_r\backslash\{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1}\mathbf{w}, \mathbf{x}\rangle^{\mathbf{\Theta}}_U|}{\|\mathbf{x}\|^{\mathbf{\Theta}}_U}.$$

Quasi-optimality of sketched Galerkin projection

$$\|\mathbf{u}(\xi) - \mathbf{u}_r(\xi)\|_U \leq (1 + \frac{\beta_r^{\mathbf{\Theta}}(\xi)}{\alpha_r^{\mathbf{\Theta}}(\xi)})\|\mathbf{u}(\xi) - \mathbf{P}_{U_r}\mathbf{u}(\xi)\|_U.$$

$$\alpha_r^{\mathbf{\Theta}}(\xi) := \min_{\mathbf{x}\in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\mathbf{\Theta}}_{U_r'}}{\|\mathbf{x}\|_U}, \qquad \beta_r^{\mathbf{\Theta}}(\xi) := \max_{\mathbf{x}\in \mathrm{span}\{\mathbf{u}(\xi)\}+U_r} \frac{\|\mathbf{A}(\xi)\mathbf{x}\|^{\mathbf{\Theta}}_{U_r'}}{\|\mathbf{x}\|_U}$$

For any $\mathbf{x} \in U_r$ the residual $\mathbf{r}(\mathbf{x};\xi)$ belongs to $Y_r(\xi)' := \mathbf{R}_U Y_r(\xi)$, with

$$Y_r(\xi) := U_r + \mathrm{span}\{\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{x} : \mathbf{x} \in U_r\} + \mathrm{span}\{\mathbf{R}_U^{-1}\mathbf{b}(\xi)\}.$$

If $\mathbf{\Theta}$ is a $U \to \ell_2$ ε-embedding for $Y_r(\xi)$, then

$$\alpha_r^{\mathbf{\Theta}}(\xi) \geq \frac{1}{\sqrt{1+\varepsilon}}(1 - \varepsilon a_r(\xi))\alpha_r(\xi), \quad \beta_r^{\mathbf{\Theta}}(\xi) \leq \frac{1}{\sqrt{1-\varepsilon}}(\beta_r(\xi) + \varepsilon\beta(\xi)),$$

where $a_r(\xi) := \max_{\mathbf{w}\in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{w}\|_{U'}}{\|\mathbf{A}(\xi)\mathbf{w}\|_{U_r'}}$.

Error estimation with sketched norm

Error estimation/certification

$$\Delta_r(\mathbf{u}_r^*(\xi);\xi) := \frac{\|\mathbf{r}(\mathbf{u}_r^*(\xi);\xi)\|_{U'}}{\eta(\xi)}.$$

Error estimation with sketched norm

Error estimation/certification with sketched norm

$$\Delta_r^{\mathbf{\Theta}}(\mathbf{u}_r^*(\xi);\xi) := \frac{\|\mathbf{r}(\mathbf{u}_r^*(\xi);\xi)\|_{U'}^{\mathbf{\Theta}}}{\eta(\xi)}.$$

Error estimation with sketched norm

Error estimation/certification with sketched norm

$$\Delta_r^{\mathbf{\Theta}}(\mathbf{u}_r^*(\xi);\xi) := \frac{\|\mathbf{r}(\mathbf{u}_r^*(\xi);\xi)\|_{U'}^{\mathbf{\Theta}}}{\eta(\xi)}.$$

If $\mathbf{\Theta}$ is a $U \to \ell_2$ ε-embedding for $Y_r(\xi)$, then

$$\sqrt{1-\varepsilon}\Delta_r(\mathbf{u}_r^*(\xi);\xi) \leq \Delta_r^{\mathbf{\Theta}}(\mathbf{u}_r^*(\xi);\xi) \leq \sqrt{1+\varepsilon}\Delta_r(\mathbf{u}_r^*(\xi);\xi).$$

ε-embedding for collection of subspaces

Constructing $U \to \ell_2$ ε-embedding for $Y_r(\xi)$ for all $\xi \in \Xi$

- If Ξ is of finite cardinality. Choose $(\varepsilon, \delta|\Xi|^{-1}, d)$ oblivious $U \to \ell_2$ subspace embedding, where $d := \max_{\xi \in \Xi}$ dim$(Y_r(\xi))$.

ε-embedding for collection of subspaces

Constructing $U \to \ell_2$ ε-embedding for $Y_r(\xi)$ for all $\xi \in \Xi$

- If Ξ is of finite cardinality. Choose $(\varepsilon, \delta|\Xi|^{-1}, d)$ oblivious $U \to \ell_2$ subspace embedding, where $d := \max_{\xi \in \Xi} \dim(Y_r(\xi))$.
- If Ξ is infinite. Assume $\bigcup_{\xi \in \Xi} Y_r(\xi)$ is contained in a space Y_r^* of dimension d^*. Choose $(\varepsilon, \delta, d^*)$ oblivious $U \to \ell_2$ subspace embedding.

A sketch of a reduced model

We refer to $\mathbf{U}_r^{\boldsymbol{\Theta}} := \boldsymbol{\Theta}\mathbf{U}_r$ and the affine expansions of

$$\mathbf{V}_r^{\boldsymbol{\Theta}}(\xi) := \boldsymbol{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_r, \quad \mathbf{b}^{\boldsymbol{\Theta}}(\xi) := \boldsymbol{\Theta}\mathbf{R}_U^{-1}\mathbf{b}(\xi), \quad \mathbf{l}_r(\xi)^{\mathrm{H}} := \mathbf{l}(\xi)^{\mathrm{H}}\mathbf{U}_r,$$

as $\boldsymbol{\Theta}$-sketch of a reduced model associated with $\mathbf{U}_r$.

A sketch of a reduced model

We refer to $\mathbf{U}_r^{\boldsymbol{\Theta}} := \boldsymbol{\Theta}\mathbf{U}_r$ and the affine expansions of

$$\mathbf{V}_r^{\boldsymbol{\Theta}}(\xi) := \boldsymbol{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_r, \quad \mathbf{b}^{\boldsymbol{\Theta}}(\xi) := \boldsymbol{\Theta}\mathbf{R}_U^{-1}\mathbf{b}(\xi), \quad \mathbf{l}_r(\xi)^{\mathrm{H}} := \mathbf{l}(\xi)^{\mathrm{H}}\mathbf{U}_r,$$

as $\boldsymbol{\Theta}$-sketch of a reduced model associated with $\mathbf{U}_r$.

- Given the sketch, the quantities required for the online stage can be computed with negligible cost.
- The sketch can be efficiently evaluated in any computational environment.
- Complexity with P-SRHT: $\mathcal{O}(nrm_A \log k + nm_b \log k)$. Recall, the classical complexity: $\mathcal{O}(nr^2m_A^2 + nm_b^2)$.
- A sketch of each snapshot can be obtained on a separate machine with absolutely no communication.
- No need to maintain large matrices and vectors.
- With good matrices, random projections are embarrassingly parallel.

Sketched Greedy algorithm

Standard error indicator for reduced basis generation with the Greedy algorithm

$$\widetilde{\Delta}_r(\xi) := \Delta_r(\mathbf{u}_r(\xi); \xi).$$

Sketched Greedy algorithm

Sketched error indicator for reduced basis generation with the Greedy algorithm

$$\widetilde{\Delta}_r(\xi) := \Delta_r^{\boldsymbol{\Theta}}(\mathbf{u}_r(\xi); \xi).$$

Sketched Greedy algorithm

Sketched error indicator for reduced basis generation with the Greedy algorithm

$$\widetilde{\Delta}_r(\xi) := \Delta_r^{\boldsymbol{\Theta}}(\mathbf{u}_r(\xi); \xi).$$

- If $\boldsymbol{\Theta}$ is $U \to \ell_2$ ε-subspace embedding for $Y_r(\xi)$ then $\Delta_r^{\boldsymbol{\Theta}}(\mathbf{u}_r(\xi); \xi)$ is close to optimal.

Sketched Greedy algorithm

Sketched error indicator for reduced basis generation with the Greedy algorithm

$$\widetilde{\Delta}_r(\xi) := \Delta_r^{\mathbf{\Theta}}(\mathbf{u}_r(\xi);\xi).$$

- If $\mathbf{\Theta}$ is $U \to \ell_2$ ε-subspace embedding for $Y_r(\xi)$ then $\Delta_r^{\mathbf{\Theta}}(\mathbf{u}_r(\xi);\xi)$ is close to optimal.
- Greedy algorithm is adaptive. $\mathbf{\Theta}$ has to be $U \to \ell_2$ ε-subspace embedding for $Y_r(\xi)$ for all possible outcomes.

Sketched Greedy algorithm

Sketched error indicator for reduced basis generation with the Greedy algorithm

$$\widetilde{\Delta}_r(\xi) := \Delta_r^{\boldsymbol{\Theta}}(\mathbf{u}_r(\xi);\xi).$$

- If $\boldsymbol{\Theta}$ is $U \to \ell_2$ ε-subspace embedding for $Y_r(\xi)$ then $\Delta_r^{\boldsymbol{\Theta}}(\mathbf{u}_r(\xi);\xi)$ is close to optimal.
- Greedy algorithm is adaptive. $\boldsymbol{\Theta}$ has to be $U \to \ell_2$ ε-subspace embedding for $Y_r(\xi)$ for all possible outcomes.
- Let $m = |\Xi_{train}|$. Choose $(\varepsilon, m^{-1}\binom{m}{r}^{-1}\delta, 2r+1)$ oblivious $U \to \ell_2$ subspace embedding for $\boldsymbol{\Theta}$.

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n\times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r}\mathbf{u}(\xi_i)\|_U^2.$$

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n \times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r} \mathbf{u}(\xi_i)\|_U^2.$$

Method of snapshots for POD

$$\mathbf{G}\mathbf{t} = \lambda \mathbf{t},$$

where $[\mathbf{G}]_{i,j} = \langle \mathbf{u}(\xi_i), \mathbf{u}(\xi_j) \rangle_U$.

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n\times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r}\mathbf{u}(\xi_i)\|_U^2.$$

Sketched Method of snapshots for POD

$$\mathbf{G}^{\boldsymbol{\Theta}}\mathbf{t} = \lambda \mathbf{t},$$

where $[\mathbf{G}^{\boldsymbol{\Theta}}]_{i,j} = \langle \mathbf{u}(\xi_i), \mathbf{u}(\xi_j)\rangle_U^{\boldsymbol{\Theta}}$.

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n \times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r} \mathbf{u}(\xi_i)\|_U^2.$$

Sketched Method of snapshots for POD

$$\mathbf{G}^{\boldsymbol{\Theta}} \mathbf{t} = \lambda \mathbf{t},$$

where $[\mathbf{G}^{\boldsymbol{\Theta}}]_{i,j} = \langle \mathbf{u}(\xi_i), \mathbf{u}(\xi_j) \rangle_U^{\boldsymbol{\Theta}}$.

$$U_r^* := \mathrm{range}(\mathbf{U}_m \mathbf{T}_r),$$

where $\mathbf{T}_r := [\mathbf{t}_1, ..., \mathbf{t}_r]$.

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n\times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r}\mathbf{u}(\xi_i)\|_U^2.$$

Sketched Method of snapshots for POD

$$\mathbf{G}^{\boldsymbol{\Theta}}\mathbf{t} = \lambda \mathbf{t},$$

where $[\mathbf{G}^{\boldsymbol{\Theta}}]_{i,j} = \langle \mathbf{u}(\xi_i), \mathbf{u}(\xi_j)\rangle_U^{\boldsymbol{\Theta}}$.

$$U_r^* := \mathrm{range}(\mathbf{U}_m \mathbf{T}_r),$$

where $\mathbf{T}_r := [\mathbf{t}_1, ..., \mathbf{t}_r]$.

If $\boldsymbol{\Theta}$ is $U \to \ell_2$ ε-subspace embedding for U_m, then

$$\frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}_i - \mathbf{P}_{U_r^*}\mathbf{u}_i\|_U^2 \leq \frac{1+\varepsilon}{1-\varepsilon} \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{u}_i - \mathbf{P}_{U_r}\mathbf{u}_i\|_U^2.$$

Sketched Proper Orthogonal Decomposition

Let $\mathbf{U}_m := [\mathbf{u}(\xi_1), \mathbf{u}(\xi_2), ..., \mathbf{u}(\xi_m)] \in \mathbb{K}^{n\times m}$ and $U_m := \mathrm{range}(\mathbf{U}_m)$.

$$U_r = \arg \min_{U_r \subseteq U_m} \frac{1}{m}\sum_{i=1}^{m} \|\mathbf{u}(\xi_i) - \mathbf{P}_{U_r}\mathbf{u}(\xi_i)\|_U^2.$$

Sketched Method of snapshots for POD

$$\mathbf{G}^{\mathbf{\Theta}}\mathbf{t} = \lambda\mathbf{t},$$

where $[\mathbf{G}^{\mathbf{\Theta}}]_{i,j} = \langle \mathbf{u}(\xi_i), \mathbf{u}(\xi_j)\rangle_U^{\mathbf{\Theta}}$.

$$U_r^* := \mathrm{range}(\mathbf{U}_m\mathbf{T}_r),$$

where $\mathbf{T}_r := [\mathbf{t}_1, ..., \mathbf{t}_r]$.

If $\mathbf{\Theta}$ is $U \to \ell_2$ ε-subspace embedding for U_m, then

$$\frac{1}{m}\sum_{i=1}^{m} \|\mathbf{u}_i - \mathbf{P}_{U_r^*}\mathbf{u}_i\|_U^2 \leq \frac{1+\varepsilon}{1-\varepsilon}\frac{1}{m}\sum_{i=1}^{m} \|\mathbf{u}_i - \mathbf{P}_{U_r}\mathbf{u}_i\|_U^2.$$

Moreover, quasi-optimality of U_r^* can be guaranteed even when $\mathbf{\Theta}$ is $U \to \ell_2$ ε-subspace embedding not for the whole U_m but several specific subspaces.

Sketched Proper Orthogonal Decomposition

Assume that we are given the sketch of a reduced model associated with $\mathbf{U}_m$:

$$\mathbf{U}_m^{\mathbf{\Theta}} := \mathbf{\Theta}\mathbf{U}_m, \quad \mathbf{V}_m^{\mathbf{\Theta}}(\xi) := \mathbf{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_m, \ldots$$

Sketched Proper Orthogonal Decomposition

Assume that we are given the sketch of a reduced model associated with $\mathbf{U}_m$:

$$\mathbf{U}_m^{\mathbf{\Theta}} := \mathbf{\Theta}\mathbf{U}_m, \quad \mathbf{V}_m^{\mathbf{\Theta}}(\xi) := \mathbf{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_m, \dots$$

Given $\mathbf{T}_r$, the sketch associated with $\mathbf{U}_r^* := \mathbf{U}_m\mathbf{T}_r$ can be evaluated with

$$\mathbf{\Theta}\mathbf{U}_r^* = \mathbf{U}_m^{\mathbf{\Theta}}\mathbf{T}_r, \quad \mathbf{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_r^* = \mathbf{V}_m^{\mathbf{\Theta}}(\xi)\mathbf{T}_r, \dots$$

Sketched Proper Orthogonal Decomposition

Assume that we are given the sketch of a reduced model associated with $\mathbf{U}_m$:

$$\mathbf{U}_m^{\mathbf{\Theta}} := \mathbf{\Theta}\mathbf{U}_m, \quad \mathbf{V}_m^{\mathbf{\Theta}}(\xi) := \mathbf{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_m, \ldots$$

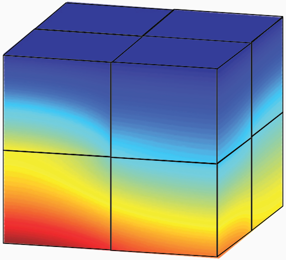
Given $\mathbf{T}_r$, the sketch associated with $\mathbf{U}_r^* := \mathbf{U}_m\mathbf{T}_r$ can be evaluated with

$$\mathbf{\Theta}\mathbf{U}_r^* = \mathbf{U}_m^{\mathbf{\Theta}}\mathbf{T}_r, \quad \mathbf{\Theta}\mathbf{R}_U^{-1}\mathbf{A}(\xi)\mathbf{U}_r^* = \mathbf{V}_m^{\mathbf{\Theta}}(\xi)\mathbf{T}_r, \ldots$$

- The sketch associated with $\mathbf{U}_r^*$ can be computed without operating with large vectors and matrices.
- With random sketching evaluating and storing POD vectors is not necessary.
- The sketch associated with $\mathbf{U}_m$ can be efficiently computed on distributed machines with no communication.
- The cost of transferring the sketches to the core is independent of n.

Outline

Thermal block benchmark



We consider the following equation:

$$-\boldsymbol{\nabla} \cdot (\kappa \boldsymbol{\nabla} T) = 0$$

with $T = 0$ on the top face, zero flux on the side faces and unit flux on the bottom face.

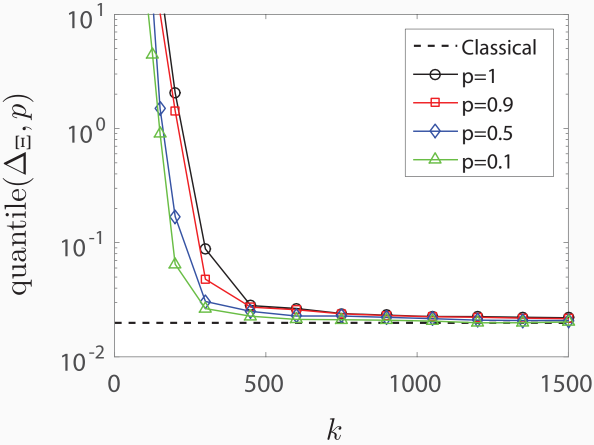
$$\kappa(x) = \kappa_i, \quad x \in \Omega_i.$$

Let $\xi := (\kappa_1, ..., \kappa_8) \in \Xi := [\frac{1}{10}, 10]^8$, $\kappa_i \sim LU[\frac{1}{10}, 10]$.

Discretization: $n \approx 120000$ degrees of freedom.

We chose $\eta(\xi) = 1$ for error estimation.

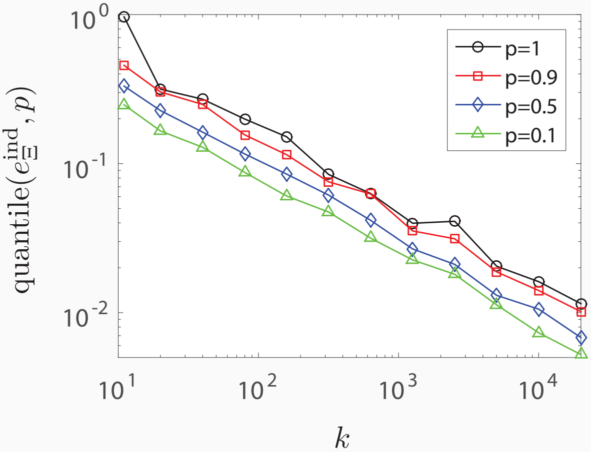
Accuracy of Galerkin projection



$\mathbf{u}(\xi)$ approximated by a projection $\mathbf{u}_r(\xi) \in U_r$ with $r = 100$.

- $\Delta_\Xi = \max_{\xi \in \Xi_{test}} \Delta_r(\mathbf{u}_r^*(\xi), \xi)$. $|\Xi_{test}| = 1000$.
- We provide results for P-SRHT. Similar performance of Gaussian and Rademacher matrices.

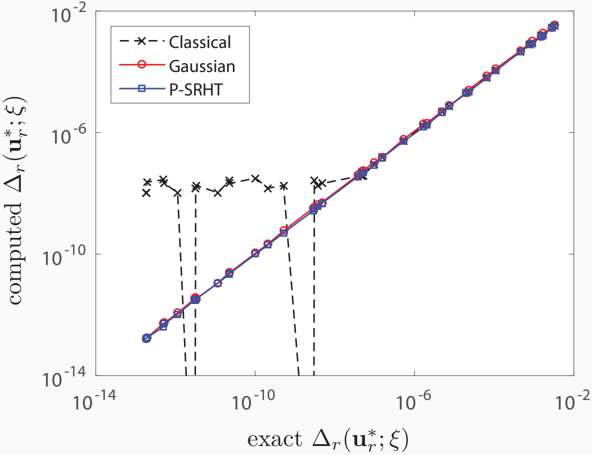
Accuracy of error indicator



- $e^{\mathrm{ind}}_{\Xi} = \max_{\xi \in \Xi_{test}} |\Delta_r(\mathbf{u}^*_r(\xi), \xi) - \Delta^{\Theta}_r(\mathbf{u}^*_r(\xi), \xi)| / \Delta_r(\mathbf{u}^*_r(\xi), \xi)$. $|\Xi_{test}| = 1000$.
- Accurate error estimation for $k \geq 100$.

Numerical stability of error indicator

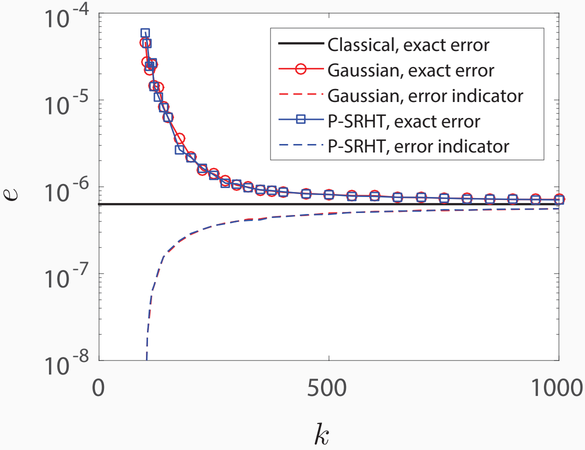
$\Delta_r(\mathbf{u}_r^*, \xi)$ and $\Delta_r^{\Theta}(\mathbf{u}_r^*, \xi)$ were evaluated for several $\mathbf{u}_r^*$ at different distances from $\mathbf{u}(\xi)$.



- The sketched error indicator is less sensitive to round off errors.

Randomized POD

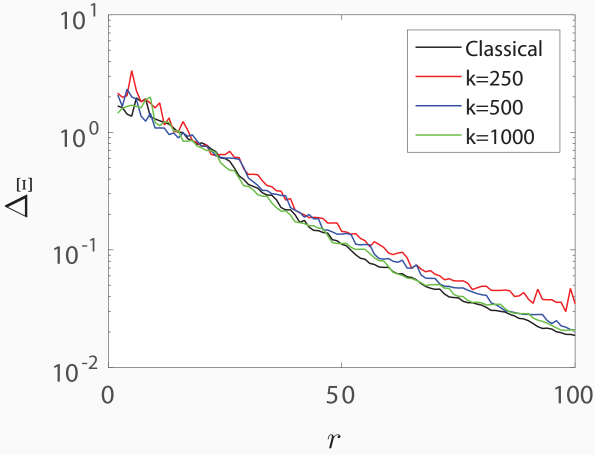
$|\Xi_{train}| = 1000$, $r = 100$.



- For $k \geq 500$, the approximate POD basis is close to optimal.

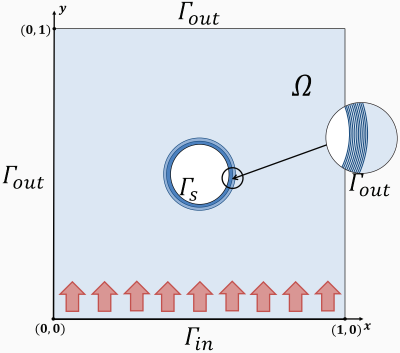
Randomized Greedy algorithm

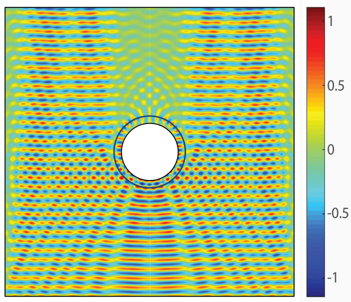
$|\Xi_{train}| = 10000$. $\Delta_\Xi := \max_{\xi \in \Xi_{train}} \Delta(\mathbf{u}_r(\xi); \xi)$.



- The convergences of the classical and the randomized (with $k \geq 500$) algorithms are almost identical.

Multi-layered acoustic cloak





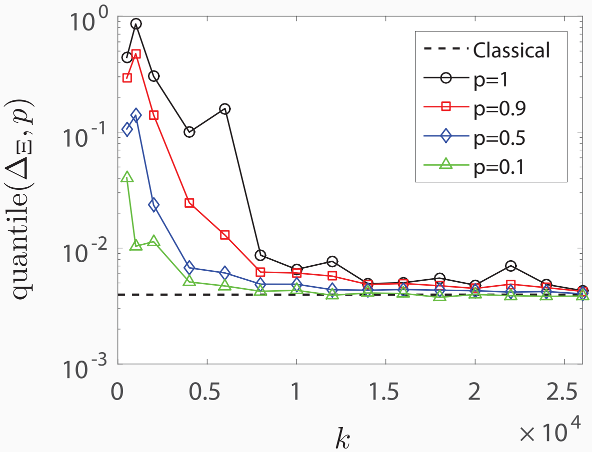
$$\Delta u + \kappa^2 u = 0,$$

with first order absorbing b.c.'s and wave initialization on Γ_{in}.

The background has $\kappa = \kappa_0 := 50$. The cloak consists of 10 layers. The i-th layer has $\kappa = \kappa_i$. Define $\xi := (\kappa_1, ..., \kappa_{10}) \in [\kappa_0, \sqrt{2}\kappa_0]^{10} := \Xi$.

Discretization: $n \approx 200000$.

Sketched Galerkin projection



$\mathbf{u}(\xi)$ approximated by a projection $\mathbf{u}_r(\xi) \in U_r$ with $r = 150$.

- The accuracy of random sketching for Galerkin projection is sensitive to operator's properties.
  More precisely, it depends on $a_r(\xi) := \max_{\mathbf{w} \in U_r} \frac{\|\mathbf{A}(\xi)\mathbf{w}\|_{U'}}{\|\mathbf{A}(\xi)\mathbf{w}\|_{U'_r}}$.

Practical computational costs

The CPU times in seconds taken by the classical greedy algorithm and the randomized greedy algorithm. $|\Xi_{train}| = 20000$.

Category	Computations	Classical	Randomized
snapshots		336	336
high-dimensional matrix-vector & inner products	sketch	–	111
	Galerkin	407	25
	error	2520	–
	remaining	185	39
	total	3111	175
provisional online solver	sketch	–	180
	Galerkin	712	712
	error	1578	373
	total	2291	1265

- We chose $k = 20000$.
- The memory consumption has been reduced from 6.29GB to only 0.96GB.
- For larger problems even more drastic reduction of computational cost is expected.

Conclusions and perspectives

- The computational cost of constructing a reduced order model is essentially reduced to evaluating the samples (snapshots).
- The reduced order model is constructed from a random sketch (a set of efficiently computable random projections).
- Our method does not require maintaining and operating with high-dimensional vectors.
- Better efficiency in terms of complexity (number of flops), memory consumption, scalability, communication cost between distributed machines, etc.

Future work

- Sketched primal-dual correction.
- Better theoretical bounds for k.
- A posteriori error indicators/certificates of accuracy of the sketch.
- Randomized minimal residual projection with random sketching insensitive to operators's properties (unlike sketched Galerkin projection).
- Efficient parameter-dependent preconditioners for projection-based MOR.

Bibliography

O. Balabanov and A. Nouy.
Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation.
arXiv preprint arXiv:1803.02602, 2018.

O. Balabanov and A. Nouy.
Randomized linear algebra for model reduction. Part II: minimal residual methods, adaptivity and efficiency.
2018.

David P Woodruff et al.
Sketching as a tool for numerical linear algebra.
Foundations and Trends® in Theoretical Computer Science, 10(1–2):1–157, 2014.

Andreas Buhr and Kathrin Smetana.
Randomized local model order reduction.
arXiv preprint arXiv:1706.09179, 2017.