Role of Hyper-Reduction in Enhancing the Efficacy of Digital Twins

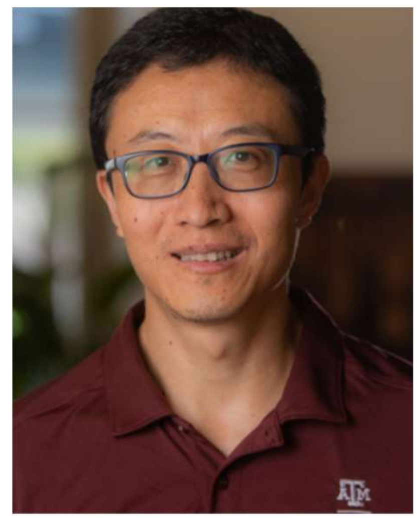
Suparno Bhattacharyya, Jian Tao, Eduardo Gildin, Jean C. Ragusa

TEXAS A&M UNIVERSITY

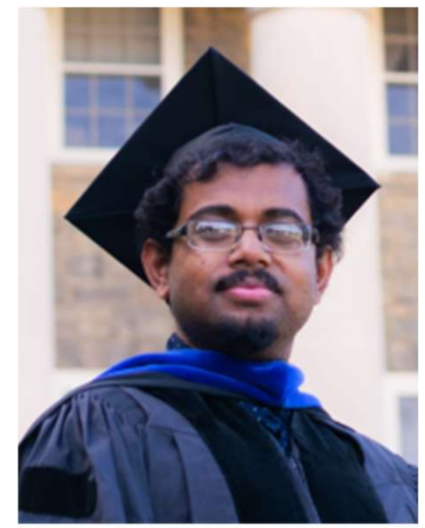
Team at Texas A&M

Jean Ragusa
Nuclear Engineering

Eduardo Gildin
Petroleum Engineering



Jian Tao
Visualization /TAMIDS



Suparno Bhattacharyya
TAMIDS

Quincy Huhn
Nuclear Engineering

Pravija Danda
Computer Science

Reduced order models

- Reduced Order Models (ROMs) are instrumental in improving computational efficiency of large-scale system simulation.

- Complex models describing neutron diffusion, neutron transport phenomena in nuclear reactors, or flow and transport models in porous media, etc. are inherently high-dimensional and/or nonlinear, and consume vast computational resources during simulation.

- ROMs, in principle, can simplify such models, improving computational efficiency.

- Hyper-reduction techniques further improves ROMs: manage non-linearities effectively.

ROMs for Digital Twin

- Digital twin (DT) technology provides a virtual duplicate of a physical asset, allowing for real-time data exchange, analysis, and control.

- In the realm of nuclear reactors, DT has shown promise with the following:
  - Real-Time Monitoring & Control for enhanced safety.
  - Forecast and mitigate potential issues, such as heat fluctuations.
  - Reduce operational costs.
  - Enhanced reactor design optimization.

- DT requires fast and precise simulations; hyper-reduced ROMs are capable of bolstering both computational efficiency and accuracy.

System of interest: nonlinear heat conduction(ss)

- Steady-state nonlinear heat conduction model provides a simplified yet representative elliptic partial differential equation (PDE) for study.

- Models describing intricate thermo-fluid dynamics, and energetics of the nuclear reactor core include nonlinear elliptic operator.

- Other large-scale models of interest such as steady-state mass transfer in porous media are often characterized by elliptic PDEs.

Nonlinear heat conduction: governing equations

$$\rho c_v \frac{\partial T}{\partial t} = \nabla \cdot (k(\mathbf{x}, T, \boxed{\mu})\nabla T) + q(\mathbf{x}, T) \quad x \in \Omega$$

Parameter

$$\boxed{0 = \nabla \cdot (k(\mathbf{x}, T, \mu)\nabla T) + q(\mathbf{x}, T)}$$

Steady-state

- **Fixed temperature at boundary** (Dirichlet) : $T|_{\Gamma_d} = T_b$
- **Fixed heat flux at boundary** (Neuman) : $-k(\mathbf{x}, T, \mu)\ \frac{\partial T}{\partial n}|_{\Gamma_n} = q_n$
- **Robin (Mixed) Condition** : $-k(\mathbf{x}, T, \mu)\ \frac{\partial T}{\partial n}|_{\Gamma_r} + h(T - T_{\text{ext}})|_{\Gamma_r} = 0$

Weak-form for FEM analysis

$$0 = \nabla \cdot (k(\mathbf{x}, T, \mu)\nabla T) + q(\mathbf{x}, T)$$

$$0 = \int_\Omega \left[\nabla \cdot (k(\mathbf{x}, T, \mu)\nabla T) + q(\mathbf{x}, T)\right] v(\mathbf{x})\, d\Omega$$

$$\int_\Omega k(\mathbf{x}, T, \mu)\nabla T \cdot \nabla v\, d\Omega - \int_\Gamma k(\mathbf{x}, T, \mu)(\nabla T \cdot \mathbf{n}) v\, d\Gamma = -\int_\Omega q(\mathbf{x}, T) v\, d\Omega$$

Stiffness

Boundary term

Source term

Weak-form for FEM analysis

$$0 = \nabla \cdot (k(\mathbf{x}, T, \mu)\nabla T) + q(\mathbf{x}, T)$$

$$0 = \int_\Omega [\nabla \cdot (k(\mathbf{x}, T, \mu)\nabla T) + q(\mathbf{x}, T)]\, v(\mathbf{x})\, d\Omega$$

Dirichlet

$$\int_\Omega k(\mathbf{x}, T, \mu)\nabla T \cdot \nabla v\, d\Omega - \int_\Gamma k(\mathbf{x}, T, \mu)(\nabla T \cdot \mathbf{n}) v\, d\Gamma = -\int_\Omega q(\mathbf{x}, T) v\, d\Omega$$

Stiffness

Boundary term

Source term

Reflective

Assembled FEM model

$$\mathsf{K}_{N_h \times N_h} \mathbf{T}_{N_h \times 1} = \mathbf{F}_{N_h \times 1}$$

$$K_{ij} = \sum_{e=1}^{n_e} \int_{\Omega_e} k(\mathbf{x}, T, \mu) \nabla N_i^e \cdot \nabla N_j^e \, d\Omega_e$$

$$F_i = \sum_{e=1}^{n_e} - \int_{\Omega_e} N_i^e q(\mathbf{x}, T) \, d\Omega_e$$

Nonlinear model solution requires iterative evaluation of T, K, F

$$\mathsf{K}_{N_h \times N_h} \mathbf{T}_{N_h \times 1} = \mathbf{F}_{N_h \times 1}$$

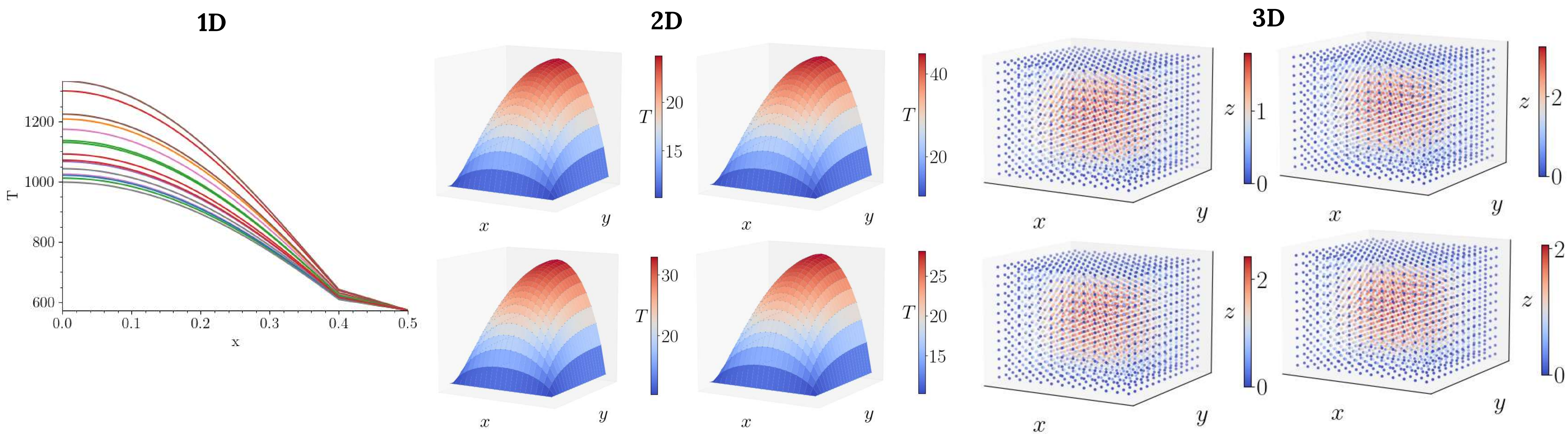
Study with large set of $\boxed{\text{parameters}}$ becomes computationally expensive for large N_h

$$K_{ij} = \sum_{e=1}^{n_e} \int_{\Omega_e} k(\mathbf{x}, \boxed{T}, \boxed{\mu}) \nabla N_i^e \cdot \nabla N_j^e \, d\Omega_e$$

$$F_i = \sum_{e=1}^{n_e} - \int_{\Omega_e} N_i^e q(\mathbf{x}, \boxed{T}) \, d\Omega_e$$

Model order reduction (MOR) for parametric nonlinear systems

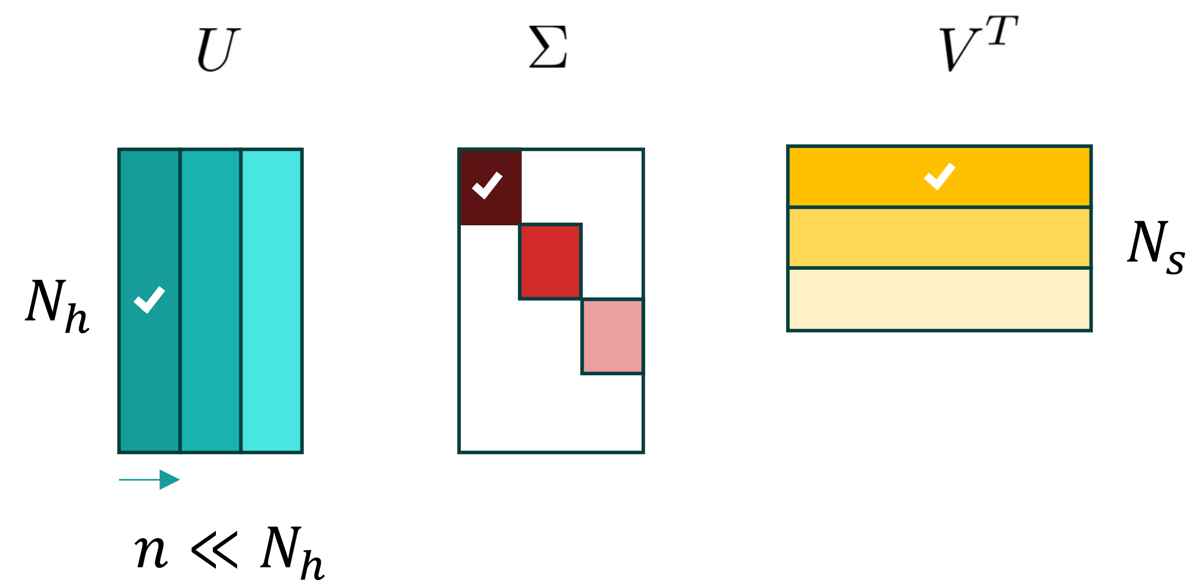
- SVD on solution snapshots corresponding to different μ.



- Determine reduced subspace, spanned by the most dominant left singular vectors.

Derive nonlinear reduced order models

$$\begin{bmatrix} | & | & & | \\ \mathbf{T}^{\mu_1} & \mathbf{T}^{\mu_2} & \cdots & \mathbf{T}^{\mu_{N_s}} \\ | & | & & | \end{bmatrix}_{N_h \times N_s} =$$



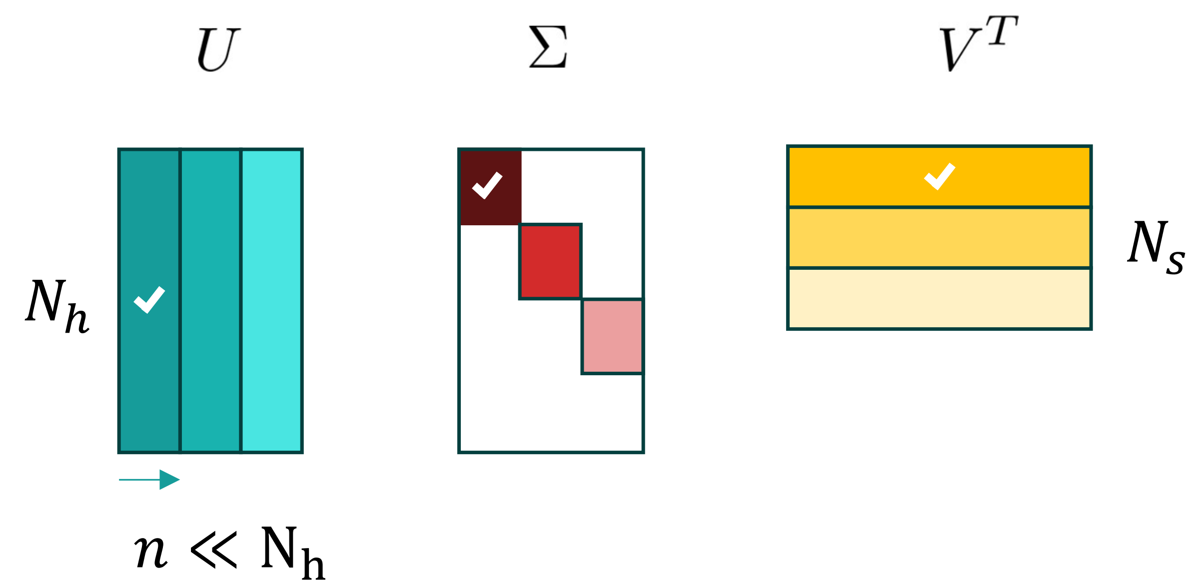
$$\widetilde{\mathsf{U}} = \mathsf{U}[:, : n] \in \mathbb{R}^{N_h \times n}$$

$$\widetilde{\mathsf{U}}^T \mathsf{K} \widetilde{\mathsf{U}}\, \mathbf{T}_n = \widetilde{\mathsf{U}}^T \mathbf{F}$$

ROM: $\mathsf{K}_n \mathbf{T}_n = \mathbf{F}_n \qquad \mathbf{T} = \widetilde{\mathsf{U}} \mathbf{T}_n$

Nonlinear ROM *still* requires iterative evaluation of T, K, F

$$\begin{bmatrix} | & | & & | \\ \mathbf{T}^{\mu_1} & \mathbf{T}^{\mu_2} & \cdots & \mathbf{T}^{\mu_{N_s}} \\ | & | & & | \end{bmatrix}_{N_h \times N_s} =$$



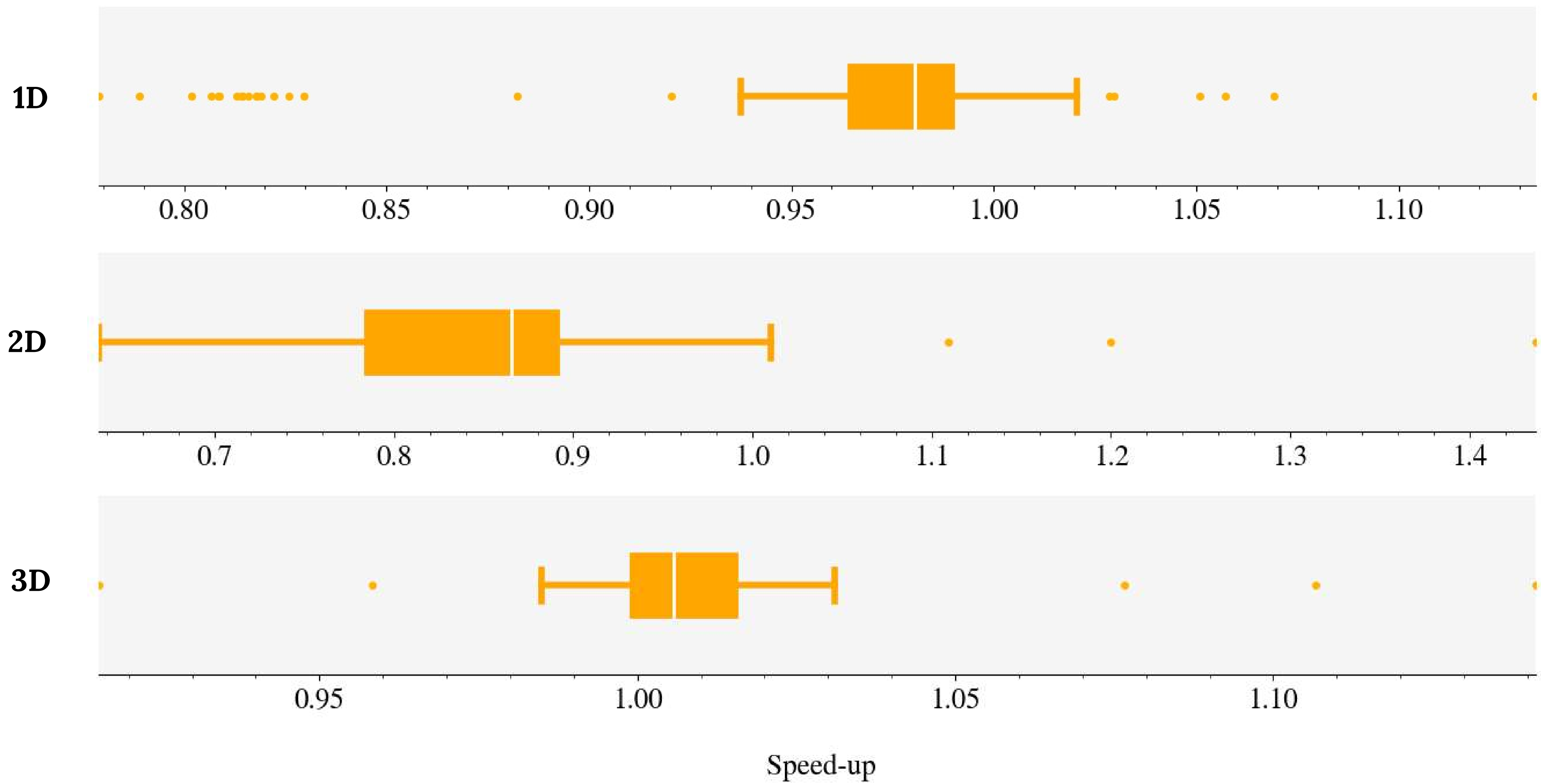
$$\widetilde{\mathsf{U}} = \mathsf{U}[:, : n] \in \mathbb{R}^{N_h \times n}$$

$$\widetilde{\mathsf{U}}^T \boxed{\mathsf{K}} \widetilde{\mathsf{U}}\, \mathbf{T}_n = \widetilde{\mathsf{U}}^T \boxed{\mathbf{F}}$$

ROM: $\mathsf{K}_n \mathbf{T}_n = \mathbf{F}_n \qquad \boxed{\mathbf{T}} = \widetilde{\mathsf{U}} \mathbf{T}_n$

Iteratively evaluate K, **F** using **T** to calculate K_n, $\mathbf{F}_n$

ROMs are hardly any faster!



Hyper-Reduction in parametric ROMs

- Minimization of computational complexity associated with evaluating nonlinear terms in ROMs.

  - The Discrete Empirical Interpolation Method (DEIM).

  - Energy conserving sampling and weighing (ECSW) method.

The Discrete Empirical Interpolation Method (DEIM)

- DEIM employs a specific kind of reduced basis, denoted as U^f, to approximate the nonlinear term $\mathbf{F}$.

$$\mathbf{F}_{N_h \times 1} \approx \mathbf{F}^{\text{approx}}_{N_h \times 1} = \mathsf{U}^f_{N_h \times r} \hat{\mathbf{F}}_{r \times 1}$$

- Basis U^f *typically* originates from snapshots of $\mathbf{F}$ (e.g., by applying SVD).

- A sampling matrix P is further calculated that selects specific components of $\mathbf{F}$, which is used to determine $\mathbf{F}^{\text{approx}}$ using

Evaluate the r rows of $\mathbf{F}$ indicated by P

$$\hat{\mathbf{F}}_{r \times 1} = (\mathsf{P}^T_{r \times N_h} \mathsf{U}^f_{N_h \times r})^{\dagger} \boxed{\mathsf{P}^T_{r \times N_h} \mathbf{F}_{N_h \times 1}}$$

S. Chaturantabut and D. C. Sorensen, “Nonlinear Model Reduction via Discrete Empirical Interpolation,” SIAM Journal on Scientific Computing, vol. 32, no. 5. Society for Industrial & Applied Mathematics (SIAM), pp. 2737–2764, Jan. 2010.

The Discrete Empirical Interpolation Method (DEIM)

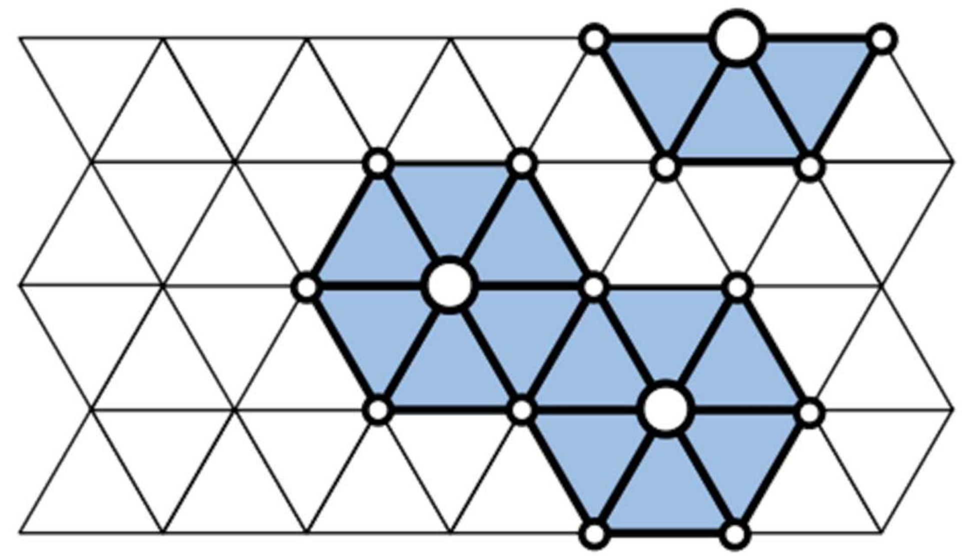
- The reduced force vector can then be calculated as

$$\mathbf{F}_n = \widetilde{\mathsf{U}}\mathbf{F}^{\mathrm{approx}} = \boxed{\widetilde{\mathsf{U}}\mathsf{U}^f(\mathsf{P}^T\mathsf{U}^f)^\dagger}\mathsf{P}^T\mathbf{F}$$

Approximate
then project

Pre-computed

DEIM with FEM

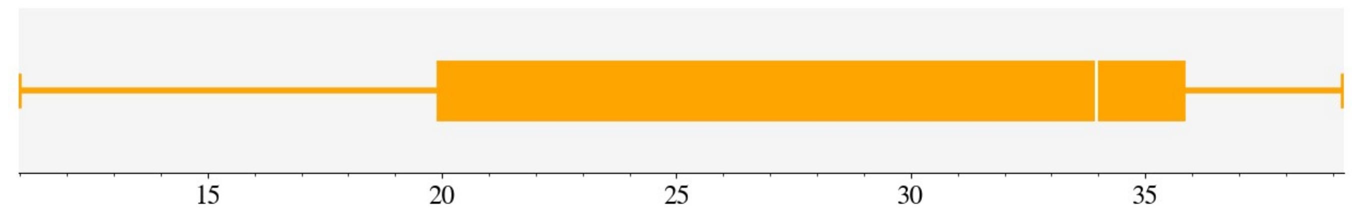


- Sampled rows of the nonlinear force vector, correspond to specific nodes.
- Calculate contribution only from elements associated with the nodes.
- Computational cost is a mere **fraction** of the total cost associated with iteratively evaluating and assembling **all** elements in the domain.

S. Jain and P. Tiso, "Hyper-Reduction Over Nonlinear Manifolds for Large Nonlinear Mechanical Systems," Journal of Computational and Nonlinear Dynamics, vol. 14, no. 8. ASME International, May 15, 2019.

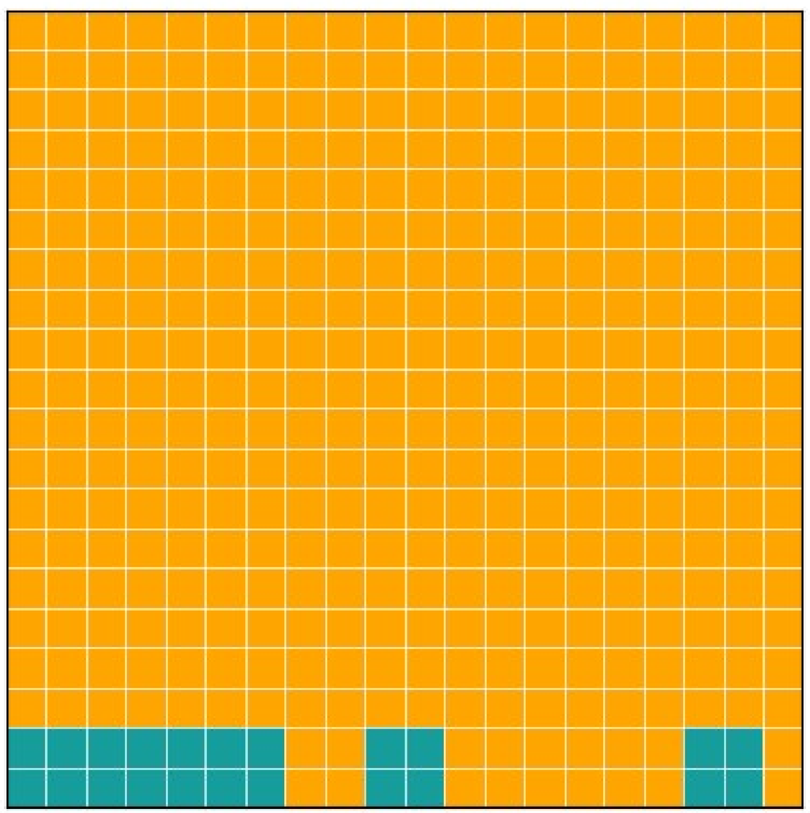
Results for Hyper-Reduced ROMs (DEIM)

1D



Active elements: 3.6%

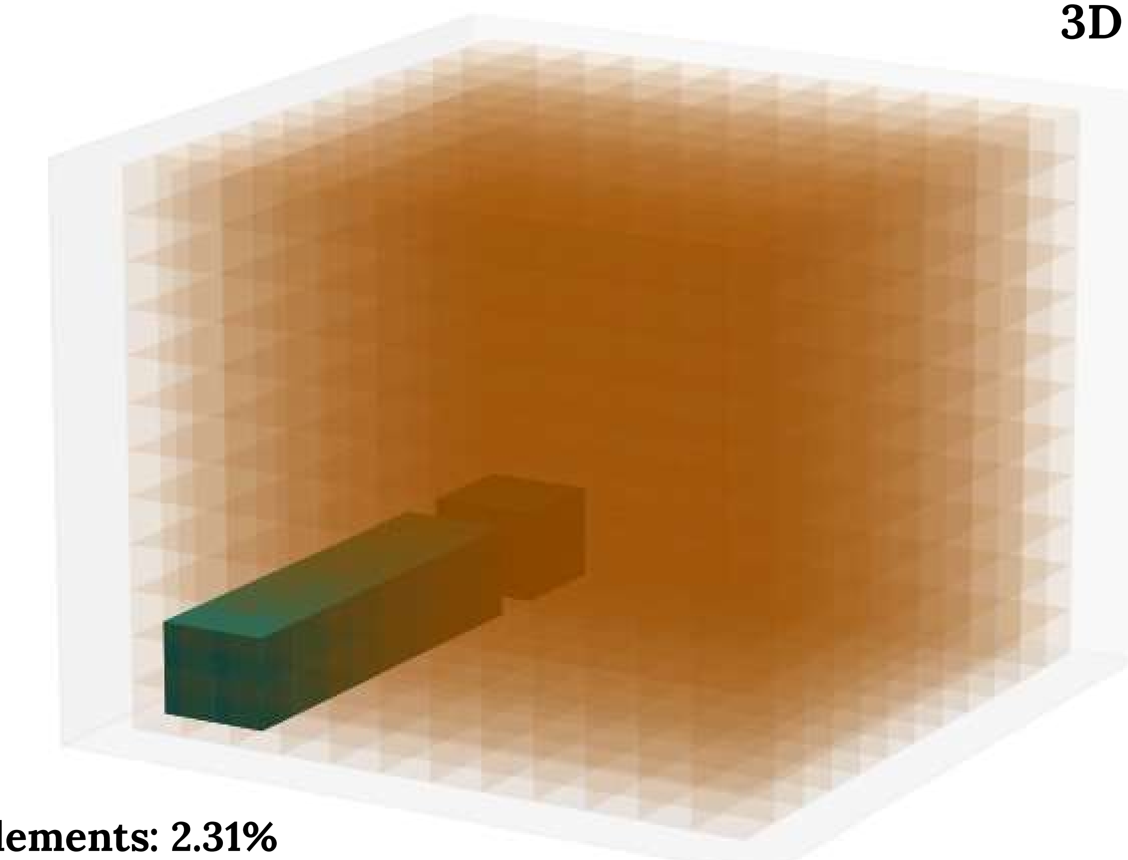
2D



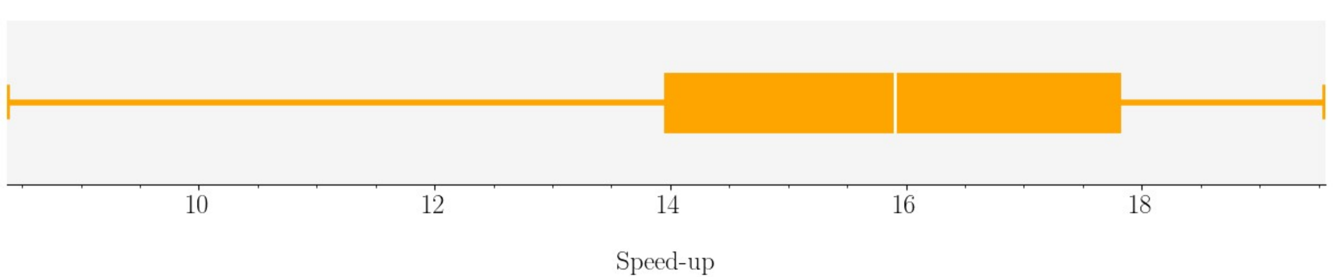
Active
elements: 5.5%

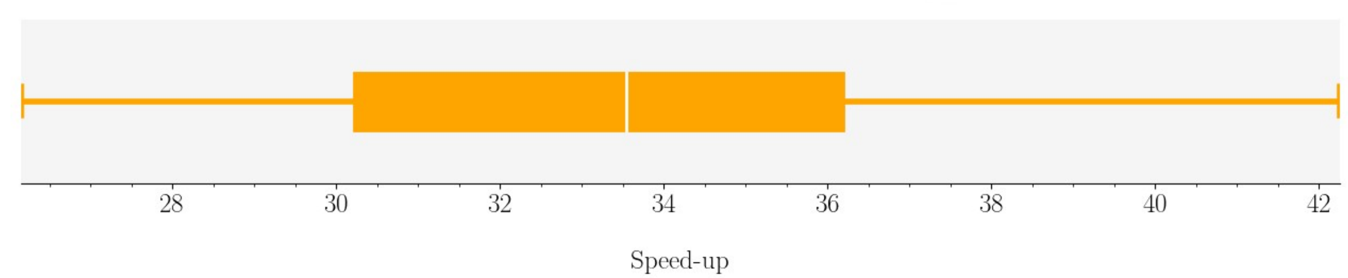
ROM error: $< 10^{-5}$

3D



Active elements: 2.31%





Energy conserving sampling and weighing (ECSW) method.

- Project-then-approximate hyper-reduction method.

- The basic idea is to match the "virtual-work" done by the internal "forces" onto the "displacements" induced by the reduction basis $\widetilde{\mathsf{U}}$.

- The nonlinear projected term: $\mathbb{F} = \sum_{e=1}^{n_e} \widetilde{\mathsf{U}}^{e^T} \left(\mathsf{K}^e \mathbf{T}^e - \mathbf{F}^e\right)$

C. Farhat, P. Avery, T. Chapman, and J. Cortial, "Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency," International Journal for Numerical Methods in Engineering, vol. 98, no. 9. Wiley, pp. 625–662, Apr. 07, 2014.

Energy conserving sampling and weighing (ECSW) method.

- Weights ξ_e are determined by matching the "work" done by the nonlinear forces onto the reduction basis for a set of N_s sampled training forces.

$$\mathbb{F} = \sum_{e=1}^{n_e} \widetilde{\mathsf{U}}^{e^T} \left(\mathsf{K}^e \mathbf{T}^e - \mathbf{F}^e \right)$$

$$\widetilde{\mathbb{F}}^{\text{approx}} = \sum_{e \in E} \xi_e \widetilde{\mathsf{U}}^{e^T} \left(\mathsf{K}^e \mathbf{T}^e - \mathbf{F}^e \right)$$

where ξ_e are positive weights and $|E| < n_e$.

C. Farhat, P. Avery, T. Chapman, and J. Cortial, "Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency," International Journal for Numerical Methods in Engineering, vol. 98, no. 9. Wiley, pp. 625–662, Apr. 07, 2014.

Energy conserving sampling and weighing (ECSW) method.

- Underlying optimization problem:

$$\xi : \arg \min_{\tilde{\xi} \in \mathbb{R}^{n_e}, \tilde{\xi} \geq 0} \left\| \mathsf{G}\tilde{\xi} - \mathbf{b} \right\|_2$$

$$\mathsf{G}_{nN_s \times n_e} = \begin{bmatrix} \mathbf{g}^{1\,1} & \cdots & \mathbf{g}^{1\,n_e} \\ \vdots & \ddots & \vdots \\ \mathbf{g}^{N_s\,1} & \cdots & \mathbf{g}^{N_s\,n_e} \end{bmatrix} \quad \mathbf{g}^{ie}_{n\times 1} = \widetilde{\mathsf{U}}^{e^T} \left(\mathsf{K}^e_i \widetilde{\mathsf{U}}^e \widetilde{\mathsf{U}}^T \mathbf{T}^{\mu_i} - \mathbf{F}^e_i \left(\widetilde{\mathsf{U}} \widetilde{\mathsf{U}}^T \mathbf{T}^{\mu_i} \right) \right)$$

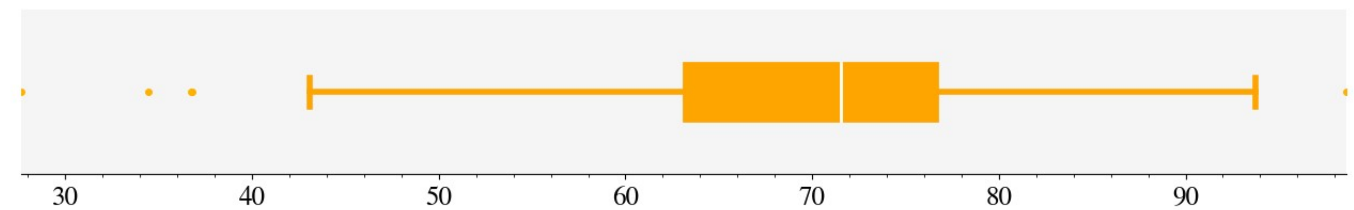
$$\mathbf{b} = \mathsf{G}\mathsf{1}$$

- Use non-negative least squares (NNLS) to solve the optimization problem and obtain sparse ξ

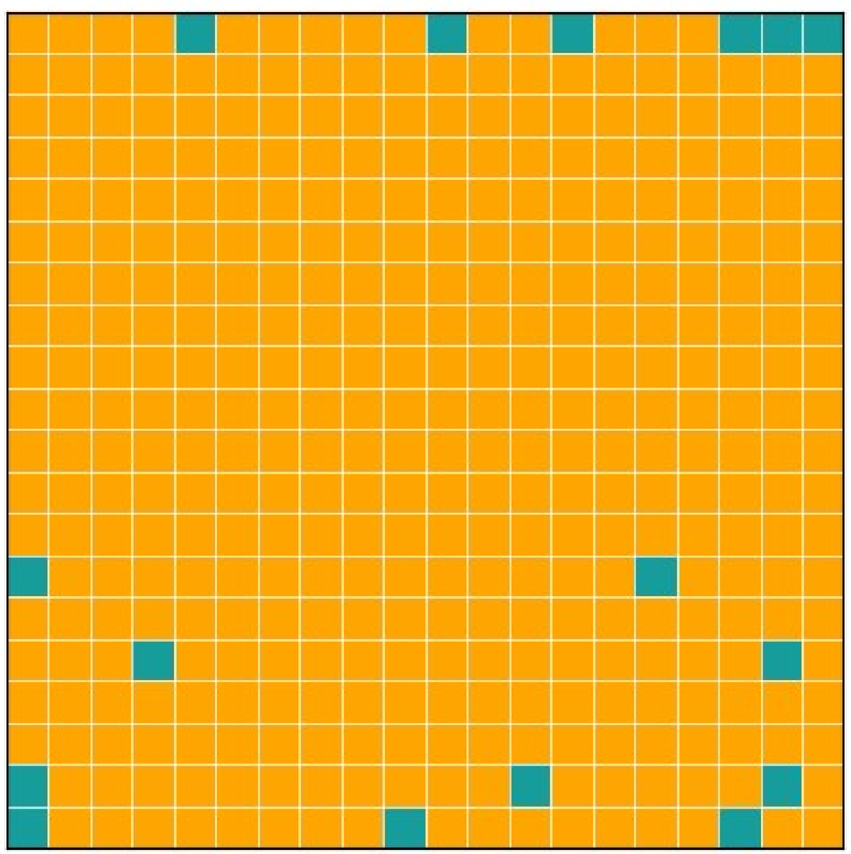
Results for Hyper-Reduced ROMs (ECSW)

1D

Active elements: 1.8 %



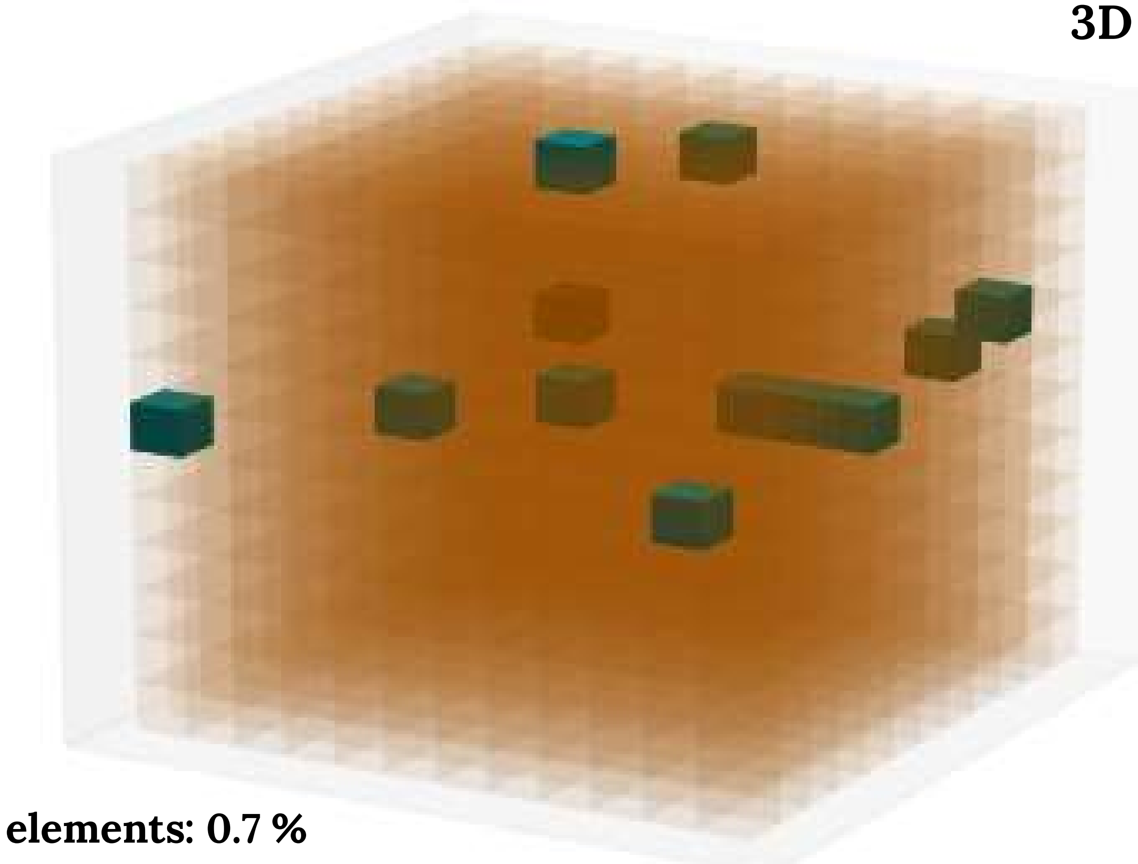
2D



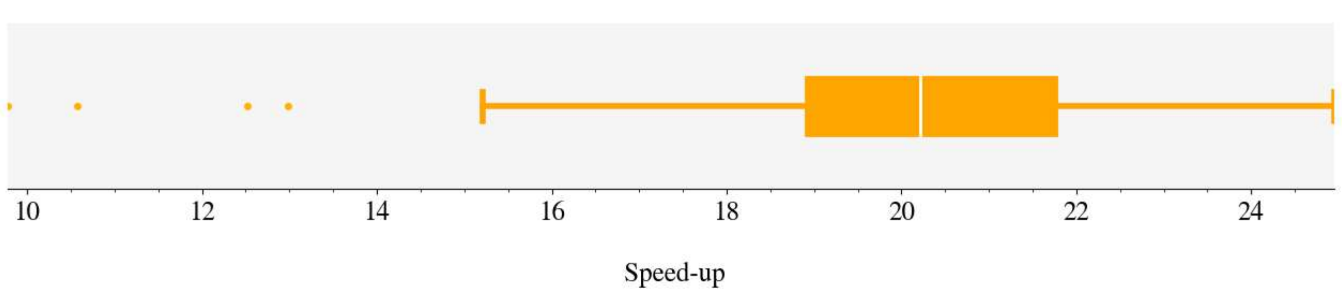
Active elements: 4.0%

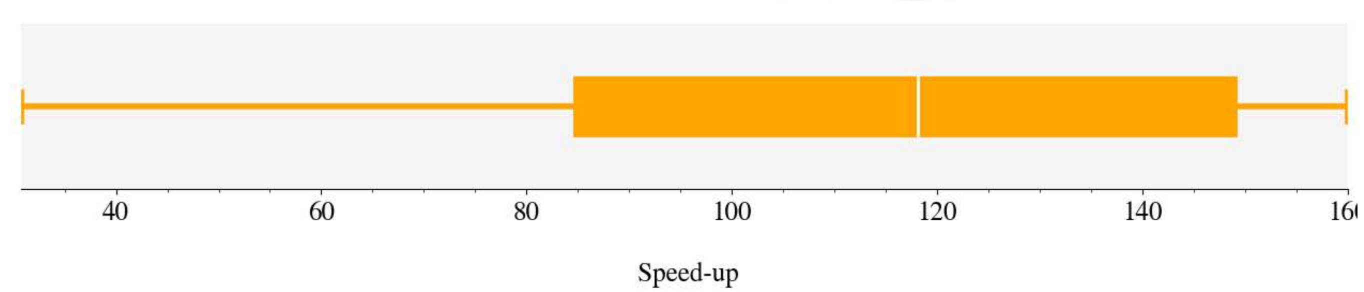
ROM error: $< 10^{-5}$

3D



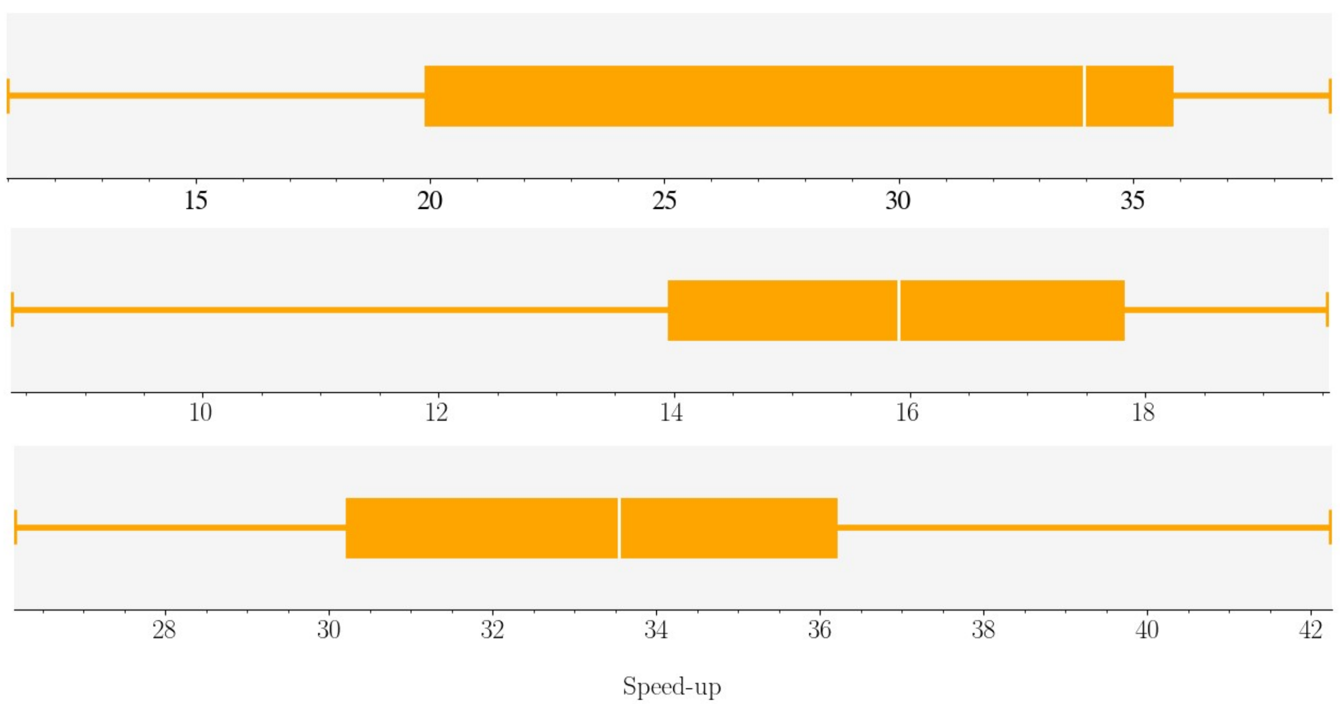
Active elements: 0.7 %



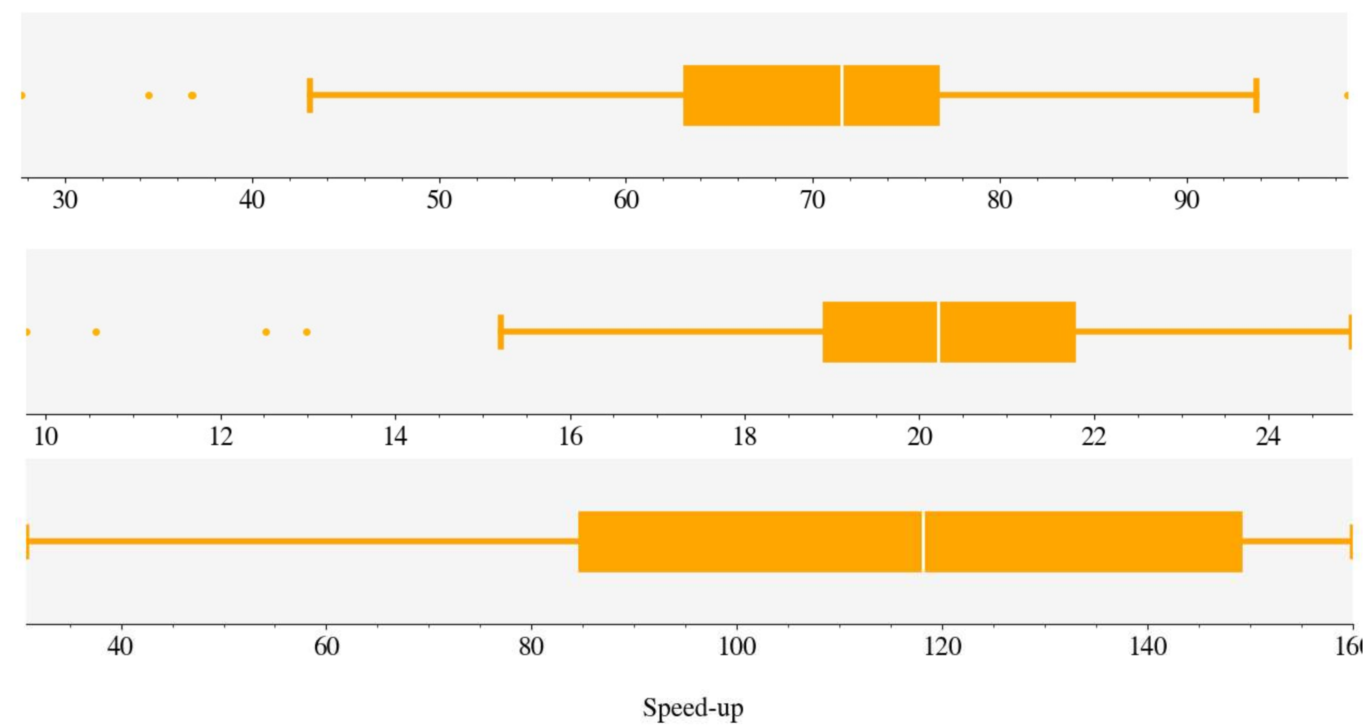


Comparison of speed-up

DEIM



ECSW



Conclusion

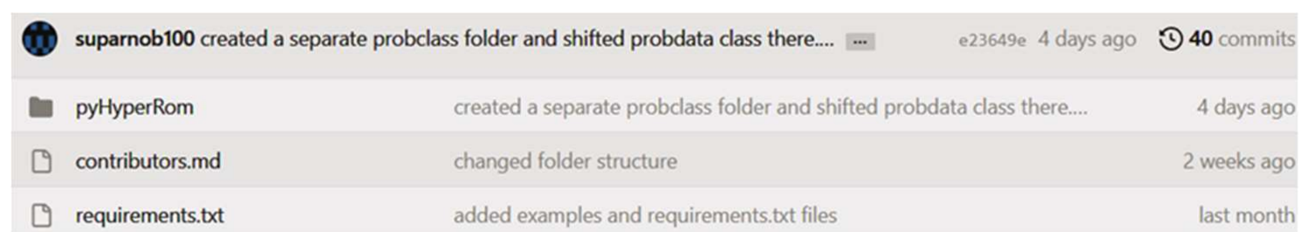
- Preliminary results on hyper-Reduced ROMs for steady-state nonlinear heat conduction problem.

- Comparison of the two popular hyperreduction techniques: DEIM, ECSW.

- Preliminary findings suggest ECSW-based hyperreduction is more effective than DEIM with similar ROM-error for these problems.

Future work

- Extend to time-dependent radiation transport models, and flow and transport models in porous media.
- Building a Github repository, which focuses on hyper-reduced ROMs that can be used for educational and computational purposes.



- Explore the scope of developing SciML-based hyper-reduction approaches.
- Repository of benchmark problems to evaluate hyper-reduction algorithms.
- **We are open to collaboration!**

suparnob@tamu.edu

Extra Slides

DEIM Optimization problem

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{y}\in\mathbb{R}^f} ||\mathsf{P}^T\mathbf{f} - \mathsf{P}^T\Phi_f\mathbf{y}||_2^2$$

Radiation transport is a much more complex equation that devolves to a diffusion operator in the limit of high amount of scattering, which is not an unreasonable approximation here.

The thermal fluid flow is typically solve using low-Mach fluid solver, so there’s a pressure-Poisson solver (against an elliptic operator) hidden in there.

Energy conservation in the solid is a nonlinear elliptic operator.

Parallel NNLS

- The solution to the assembly problem is to domain decompose the problem and solve NNLS on each of the N_d sub-domains
- We take $\widetilde{N}_e = N_e/N_d$ and assemble
$$\left[C_1, C_2, \dots, C_{N_d}\right], C_i \in \mathbb{R}^{N_s \text{ x } \widetilde{N}_e}$$
- We then form and solve the NNLS problem on each sub-domain
$$\left[C_1\xi_1^* \approx d_1, C_2\xi_2^* \approx d_2, \dots, C_{N_d}\xi_{N_d}^* \approx d_{N_d}\right]$$
- After solving for $\left[\xi_1^*, \xi_2^*, \dots, \xi_{N_d}^*\right]$ we can form ξ^* that gives the weighting for ECSW over the entire domain
- Farhat et. al. (2015) proved that you can maintain a global tolerance ε by having a tolerance on each sub-domain of
$$\varepsilon_i \equiv \left(\frac{\|d\|_2}{N_s\|d_i\|}\right)\varepsilon$$
- We should be able to calculate NNLS in parallel with no loss in accuracy

Parallel NNLS Results

	Serial	Parallel N=2	Parallel N=5
Error	1.66e-06 %	3.73e-06 %	3.72e-06 %
% Cells Retained	10.6%	20.0%	40.2%
NNLS Speedup	1x	1.76x	4.03x
ROM Speedup	9.27x	4.47x	2.06x

- Speedup is achieved in the NNLS by the parallelization
- The number of cells retained goes up however in the case where there is not a tolerance
- This tradeoff means that the system matrix must be very large before it is economical to consider parallel NNLS