S-OPT: A Points Selection Algorithm for Data-Driven Hyper-Reduction in Reduced Order Models

Siu Wun Cheung (cheung26@llnl.gov)

6th Annual MLDL Workshop – Sandia National Laboratory

July 25–28, 2022
Outline

1. Background
   - High fidelity physics simulation
   - Model order reduction

2. Hyper-reduction
   - Quasi-optimality of sampling matrix
   - Discrete Empirical Interpolation Method (DEIM)
   - S-OPT

3. Numerical examples
   - NACA0012 airfoil in compressible Navier-Stokes flow
   - Gresho vortex in compressible Euler equations
System of ordinary differential equations (ODE)

Consider a system of nonlinear ODEs with state dimension $N$:

$$M(\mu)\dot{u}(t; \mu) = f(u(t; \mu), t; \mu), \quad u(0; \mu) = u^0(\mu).$$

- $t \in [0, T]$ denotes time,
- $\mu \in \mathcal{D}$ denotes a vector of parameters,
- $M(\mu) \in \mathbb{R}^{N \times N}$ denotes the nonsingular system matrix,
- $f: \mathbb{R}^N \times \mathbb{R} \times \mathcal{D} \to \mathbb{R}^N$ is a nonlinear function,
- $u^0(\mu) \in \mathbb{R}^N$ denotes the initial condition, and
- $u(t; \mu) \in \mathbb{R}^N$ denotes the state vector.
System of difference equations (OΔE)

\[ M(\mu) \dot{u}(t; \mu) = f(u(t; \mu), t; \mu), \quad u(0; \mu) = u^0(\mu). \]

- Partition the time domain into \(0 = t_0 < t_1 < t_2 < \ldots < t_M = T\).
- Forward Euler discretization: \( \dot{u}(t_n) \approx \frac{u(t_{n+1}) - u(t_n)}{t_{n+1} - t_n} = \frac{u^{n+1} - u^n}{\Delta t_n} \).

\[ M(\mu)u^{n+1}(\mu) = M(\mu)u^n(\mu) + \Delta t_n f^n(u^n(\mu); \mu) \quad \text{(Explicit)}. \]

- Backward Euler discretization: \( \dot{u}(t_{n+1}) \approx \frac{u(t_{n+1}) - u(t_n)}{t_{n+1} - t_n} = \frac{u^{n+1} - u^n}{\Delta t_n} \).

\[ M(\mu)u^{n+1}(\mu) - \Delta t_n f^{n+1}(u^{n+1}(\mu); \mu) = M(\mu)u^n(\mu) \quad \text{(Implicit)}. \]

- High order discretization: linear multistep/Runge-Kutta methods.
- Requires to solve system of algebraic equations in \( \mathbb{R}^N \).
Reduced order model (ROM)

- Objective: reduce complexity of numerical simulation.
- Low-dimensional solution representation by $y \in \mathbb{R}^k (k \ll N)$:
  \[ u \approx \tilde{u}(y) = u_{\text{ref}} + g(y). \]
- $u_{\text{ref}} \in \mathbb{R}^N$ is a reference state.
- $g : \mathbb{R}^k \to \mathbb{R}^N$ denotes a function.
- Linear subspace (LS)-ROM: $g(y) = \Phi y$, with $\Phi \in \mathbb{R}^{N \times k}$ orthogonal.
- Galerkin projection – solve the reduced-dimensional ODE for $y \in \mathbb{R}^k$:
  \[ \Phi^\top M(\mu) \Phi \dot{y}(t; \mu) = \Phi^\top f(u_{\text{ref}} + \Phi y(t; \mu), t; \mu). \]
- Other ROM approaches:
  - Nonlinear manifold (NM)-ROM: $g$ is a decoder neural network.
Proper orthogonal decomposition (POD)

- Snapshot data from direct numerical simulation or measurement

\[ X = \begin{bmatrix} u^1(\mu_{\text{train}}) - u_{\text{ref}} & \cdots & u^M(\mu_{\text{train}}) - u_{\text{ref}} \end{bmatrix} \in \mathbb{R}^{N \times M} (M \leq N). \]

- Snapshot singular value decomposition (SVD)

\[ X = U\Sigma V^T. \]

- Energy criteria: determine smallest integer \( 1 \leq k \leq N \) such that

\[ \frac{\sum_{i=1}^{k} \sigma_i}{\sum_{i=1}^{M} \sigma_i} \geq \epsilon_\sigma. \]

- Solution basis \( \Phi = U \begin{bmatrix} e_1 & \cdots & e_k \end{bmatrix} \in \mathbb{R}^{N \times k} \) is orthogonal.
POD modes of NACA0012 laminar airfoil

**Figure:** First four POD modes for the density variable for the laminar airfoil.
Hyper-reduction

- No speed-up due to evaluation of source term $f \in \mathbb{R}^N$.

$$\Phi^\top M(\mu)\Phi \dot{y}(t; \mu) = \Phi^\top f(u_{\text{ref}} + \Phi y(t; \mu), t; \mu).$$

- Low-dimensional source representation by $\hat{f} \in \mathbb{R}^{n_f}$ ($n_f \ll N$):

$$f \approx \Phi_f \hat{f}, \quad \hat{f} = \arg\min_{a \in \mathbb{R}^{n_f}} \| Z^\top (\Phi_f a - f) \| = (Z^\top \Phi_f)\dagger Z^\top f,$$

where $Z = [e_{i_1}, \cdots, e_{i_n}] \in \mathbb{R}^{N \times n}$ ($n_f \leq n \ll N$) is a sampling matrix.

- Achievable speed-up due to evaluation of $Z^\top f \in \mathbb{R}^n$.

$$\Phi^\top M(\mu)\Phi \dot{y}(t; \mu) = \Phi^\top \Phi_f (Z^\top \Phi_f)\dagger Z^\top f(u_{\text{ref}} + \Phi y(t; \mu), t; \mu).$$

- Keep the index set $Z = \{i_1, \cdots, i_n\}$ instead of full matrix $Z$.

- $(Z^\top \Phi_f)\dagger \in \mathbb{R}^{n_f \times n}$ is offline precomputed and stored.
Quasi-optimality of sampling matrix

Given $\Phi_f \in \mathbb{R}^{N \times n_f}$ and $f \in \mathbb{R}^N$, the optimal sampling matrix is

$$Z^*(f) = \arg\min_{Z=[e_{i_1}, \ldots, e_{i_n}] \in \mathbb{R}^{N \times n}} \| (I - \Phi_f(Z^\top \Phi_f)\dagger Z^\top) f \|.$$ 

- The true optimum $Z^*(f)$ depends on $f$ and is impractical.
- Quasi-optimality: oblique projection error bound for generic $f$.

**Theorem**

Let $\Phi_f = QR$ be the QR factorization of $\Phi_f$.

$$\| (I - \Phi_f(Z^\top \Phi_f)\dagger Z^\top) f \| \leq \| (Z^\top Q)\dagger \| \| (I - QQ^\top) f \|.$$
Discrete Empirical Interpolation Method (DEIM)

\[ \mathbf{Z}_{\text{DEIM}}^* = \arg\min_{\mathbf{Z} = [\mathbf{e}_{i_1}, \cdots, \mathbf{e}_{i_n}] \in \mathbb{R}^{N \times n}} \| (\mathbf{Z}^\top \mathbf{Q})^\dagger \|. \]

Greedy sampling procedure to select the index with largest error.

1. Initialize \( Z = \{i^*\} \) where \( i^* = \arg\max_i |Q_{i,1}| \)
2. For \( j = 1, \cdots, n_f \),
   1. Construct \( A = \mathbf{Q} [\mathbf{e}_1, \cdots, \mathbf{e}_j] \).
   2. For \( k = 1, \cdots, K_j \),
      1. Construct \( \mathbf{Z} = [\mathbf{e}_i]_{i \in Z} \).
      2. Compute \( \tilde{q} = \mathbf{A}(\mathbf{Z}^\top \mathbf{A})^\dagger \mathbf{Z}^\top \mathbf{Q} \mathbf{e}_j \).
      3. Find \( i^* = \arg\max_{i \notin Z} |Q_{i,j} - \tilde{q}_i| \).
      4. Enrich \( Z \leftarrow Z \cup \{i^*\} \).
3. Set \( n = |Z| \) and output \( Z \).

\( K_1 = 0 \) and \( K_j = 1 \) for \( j > 1 \) \( \implies \) original DEIM with \( n = n_f \).
\( K_1 > 0 \) or \( K_j > 1 \) for \( j > 1 \) \( \implies \) oversampling DEIM with \( n > n_f \).
Another measure for quasi-optimality

**Theorem**

Let $\Phi_f = QR$ be the QR factorization of $\Phi_f$.

$$\| (I - \Phi_f(Z^\top \Phi_f)^+ Z^\top) f \|_2^2 = \| (I - QQ^\top) f \|_2^2 + \| \epsilon(f, Z) \|_2^2,$$

where $\epsilon(f, Z) = ((Z^\top Q)^\top (Z^\top Q))^{-1}(Z^\top Q)^\top Z^\top (I - QQ^\top) f$.

- Maximize the column orthogonality of $Z^\top Q$ to make residual small.
- Maximize the determinant of $(Z^\top Q)^\top Z^\top Q$ for non-singularity.
- Define the measure $S : \mathbb{R}^{m \times p} \rightarrow [0, 1]$ by:

$$S(A) = \left( \frac{\sqrt{\det A^\top A}}{\prod_{k=1}^p \| Ae_k \|} \right)^{\frac{1}{p}} \quad \text{for all} \ A \in \mathbb{R}^{m \times p}.$$
Another measure for quasi-optimality

Theorem

Let \( A \in \mathbb{R}^{m \times p} \) with \( m < p \). Then \( S(A) = 0 \).

Theorem

Let \( A \in \mathbb{R}^{m \times m} \) be full-rank, \( r, c \in \mathbb{R}^m \), and \( \gamma \in \mathbb{R} \). Define

\[
\tilde{A} = \begin{bmatrix} A & c \\ r^\top & \gamma \end{bmatrix} \in \mathbb{R}^{(m+1) \times (m+1)}.
\]

Then

\[
(S(\tilde{A}))^{2(m+1)} \left( \prod_{k=1}^m (\|Ae_k\|^2 + r_k^2) \right) = \frac{(\det(A^\top A)) \left( 1 + r^\top b \right)}{\left( \frac{c^\top c + \gamma^2 - \alpha}{c^\top c + \gamma^2} \right)} ,
\]

\[
b = (A^\top A)^{-1} r, \quad g = (A^\top A)^{-1} A^\top c, \]

\[
\alpha = (c^\top A + \gamma r^\top)(I - (1 + r^\top b)^{-1} br^\top)(g + \gamma b).
\]
Another measure for quasi-optimality

Theorem

Let $A \in \mathbb{R}^{m \times p}$ with $m \geq p$. Then $S(A) = 1 \iff A^\top A = I$.

Theorem

Let $A \in \mathbb{R}^{m \times p}$ with $m \geq p$ be full-rank and $r \in \mathbb{R}^p$. Define

$$\tilde{A} = \begin{bmatrix} A \\ r^\top \end{bmatrix} \in \mathbb{R}^{(m+1) \times p}.$$

Then

$$\left(S(\tilde{A})\right)^{2p} = \left(\det(A^\top A)\right) \frac{1 + r^\top (A^\top A)^{-1} r}{\prod_{k=1}^{p} (\|Ae_k\|^2 + r_k^2)}.$$
S-OPT sampling algorithm

\[
S_{\text{S-OPT}} = \arg\max_{Z \in \mathbb{R}^{N \times n}} S(Z^T Q).
\]

Greedy sampling procedure to select the index with largest measure \( S \).

1. Initialize \( \mathcal{Z} = \{ i^* \} \) where \( i^* = \arg\max_i |Q_{i,1}| \)
2. For \( j = 1, \ldots, n_f - 1 \),
   1. Construct \( Z = [e_i]_{i \in \mathcal{Z}}, A = Z^T Q [e_1, \cdots, e_j], c = Z^T Q e_{j+1} \), and \( g = (A^T A)^{-1} A^T c \).
   2. Find \( i^* = \arg\max_{i \notin \mathcal{Z}} \frac{1 + r^T b}{\prod_{k=1}^{j} (\|Ae_k\|^2 + r_k^2)} \frac{c^T c + \gamma^2 - \alpha}{c^T c + \gamma^2} \),
   \[
   r^T = e_i^T Q [e_1, \cdots, e_j], \quad b = (A^T A)^{-1} r, \quad \gamma = Q_{i,j+1},
   \]
3. Enrich \( \mathcal{Z} \leftarrow \mathcal{Z} \cup \{ i^* \} \).
S-OPT sampling algorithm

\[ Z_{S-\text{OPT}}^* = \arg\max_{Z = [e_{i_1}, \cdots, e_{i_n}] \in \mathbb{R}^{N \times n}} S(Z^\top Q). \]

Greedy sampling procedure to select the index with largest measure \( S \).

3. For \( j = 1, \cdots, n - n_f \),
   1. Construct \( Z = [e_i]_{i \in Z} \) and \( A = Z^\top Q \).
   2. Find \( i^* = \arg\max_{i \notin Z} \frac{1 + r^\top (A^\top A)^{-1}r}{\prod_{k=1}^{n_f} (\|Ae_k\|^2 + r_k^2)} \), with \( r^\top = e_i^\top Q \).
   3. Enrich \( Z \leftarrow Z \cup \{i^*\} \).
4. Output \( Z \).
Perspective of optimal design

- Ordinary least squares with design matrix $X = Z^T Q$.
- Mean-unbiased estimator $\beta = \hat{R} \hat{f}$.
- Minimizing the variance $\iff$ Maximizing the Fisher information.
- Information matrix $(Z^T Q)^T (Z^T Q)$.
- Optimal design $\iff$ Optimization of certain statistical criteria.
- E-optimality: Maximize the smallest eigenvalue of information matrix.

$$Z^*_{\text{DEIM}} = \arg\min_{Z = [e_{i_1}, \ldots, e_{i_n}] \in \mathbb{R}^{N \times n}} \| (Z^T Q)^\dagger \|. $$

- S-optimality: Maximizing column orthogonality of design matrix and determinant of information matrix.

$$Z^*_{\text{S-OPT}} = \arg\max_{Z = [e_{i_1}, \ldots, e_{i_n}] \in \mathbb{R}^{N \times n}} S(Z^T Q).$$
NACA0012 airfoil in compressible Navier-Stokes flow

**Figure:** Partial domain near the airfoil. Selected nodes are in yellow. Neighboring nodes required for nonlinear term calculation are in cyan.
NACA0012 airfoil in compressible Navier-Stokes flow

Figure: Comparison of oversampled DEIM (dashed) versus S-OPT (solid line) algorithms in maximum error (left) and simulation wall clock in seconds (right).
Gresho vortex in compressible Euler equations

Figure: Initial condition (left) and final-time solution (right).
Gresho vortex in compressible Euler equations

Figure: Nodes selected by DEIM (left) and S-OPT (right).
Gresho vortex in compressible Euler equations
References


libROM

GitHub repo: https://github.com/LLNL/libROM
Website: https://www.librom.net
Thank you for your attention. Any questions?
Siu Wun Cheung (cheung26@llnl.gov)