

SAND2015-9973PE

Nonlinear model reduction: discrete optimality and time parallelism

Kevin Carlberg

Sandia National Laboratories

ICME LA/Opt Seminar
November 12, 2015

Collaborators

2014–2015 M. Barone

Sandia National Laboratories

2014–2015 H. Antil

George Mason University

2015 L. Brencher, B. Haasdonk

University of Stuttgart

2010–2011 C. Farhat, C. Bou-Mosleh, J. Cortial, D. Amsallem

Stanford University

Computational barrier

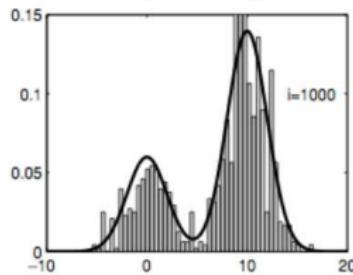
High-fidelity simulation

- + An indispensable tool
- Very high computational cost

Barrier

Time-critical applications

Many query



[Andrieu et al., 2003]

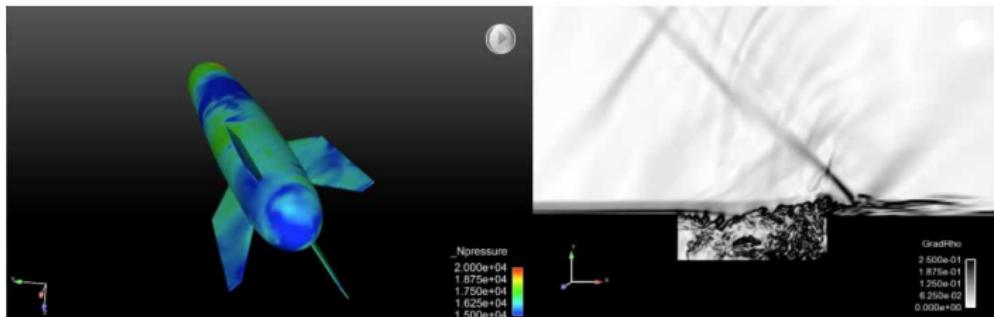
Real time



[pow,]

Objective: break barrier

Computational barrier at Sandia



- CFD model
 - 100 million cells
 - 200,000 time steps
- High simulation costs
 - 6 weeks, 5000 cores
 - 6 runs **maxes out Cielo**

Barrier

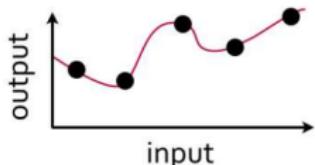
- Fast-turnaround design
- Uncertainty quantification (UQ)

Surrogate modeling

inputs $\mu \rightarrow$ **full-order model** \rightarrow outputs y

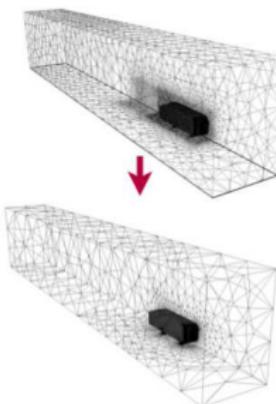
inputs $\mu \rightarrow$ **surrogate model** \rightarrow outputs y

1) Data fits



- Not physics based
- + High speedups

2) Coarsened physics



- + Physics based
- Low speedups

3) Reduced-order models (ROMs)

- + Physics based
- + High speedups
- + Preserve structure
- + Rigorous error analysis
- **Unproven for nonlinear dynamical systems**

ROM = data science + modeling and simulation

Goal: exploit simulation data to drastically reduce simulation costs

ROM: state of the art [Benner et al., 2015]

- Linear time-invariant systems: **mature** [Antoulas, 2005]
 - Balanced truncation [Moore, 1981]
 - Empirical balanced truncation [Willcox and Peraire, 2002, Rowley, 2005]
 - Moment matching
 - [Bai, 2002, Freund, 2003, Gallivan et al., 2004, Baur et al., 2011]
 - Loewner framework [Lefteriu and Antoulas, 2010, Ionita and Antoulas, 2014]
 - + *Reliable*: guaranteed stability, *a priori* error bounds
 - + *Certified*: sharp, computable *a posteriori* error bounds
- Elliptic/parabolic PDEs (FEM): **mature** [Rozza et al., 2008]
 - Reduced-basis method
 - [Prud'Homme et al., 2001, Veroy et al., 2003, Barrault et al., 2004]
 - Subsystem-based reduced-basis method
 - [Maday and Rønquist, 2002, Phuong Huynh et al., 2013, Eftang and Patera, 2013]
 - + *Reliable*: *a priori* error bounds
 - + *Certified*: sharp, computable *a posteriori* error bounds
- Nonlinear dynamical systems: **unproven**
 - Proper orthogonal decomposition (POD)–Galerkin
 - *Not reliable*: Stability and accuracy not guaranteed
 - *Not certified*: error bounds not sharp

Our research goal

Nonlinear model-reduction methods that are accurate, low cost, certified, and reliable.

+ Accuracy

- Improve projection technique [?, C. et al., 2015a]
- Preserve problem structure [C. et al., 2012, C. et al., 2015c]

+ Low cost

- Sample-mesh approach [C. et al., 2011, C. et al., 2013]
- Leverage time-domain data [C. et al., 2015b]

+ Certification

- Error bounds [C. et al., 2015a]
- Statistical error modeling [Drohmann and C., 2015]

+ Reliability

- *A posteriori* h -refinement [C., 2015]

Our research goal

Nonlinear model-reduction methods that are accurate, low cost, certified, and reliable.

+ Accuracy

- Improve projection technique [?, C. et al., 2015a]
- Preserve problem structure [C. et al., 2012, C. et al., 2015c]

+ Low cost

- Sample-mesh approach [C. et al., 2011, C. et al., 2013]
- Leverage time-domain data [C. et al., 2015b]

+ Certification

- Error bounds [C. et al., 2015a]
- Statistical error modeling [Drohmann and C., 2015]

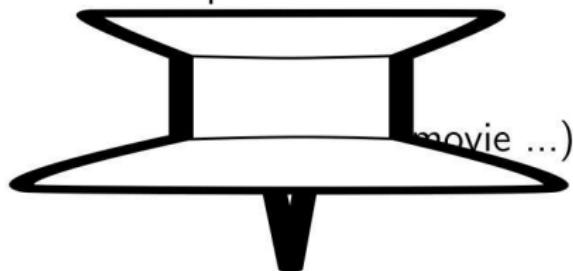
+ Reliability

- *A posteriori* h -refinement [C., 2015]

POD–Galerkin: offline data collection

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}; t, \mu); \quad \mathbf{x}(0, \mu) = \mathbf{x}^0(\mu), \quad t \in [0, T], \quad \mu \in \mathcal{D}$$

- 1 Collect 'snapshots' of the state



- 2 Data compression

- Compute SVD:

$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \mathbf{U} \ \boldsymbol{\Sigma} \ \mathbf{V}^T$$

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix of snapshots. On the left, a matrix is shown with three columns, each represented by a vertical bar of a different color: green, blue, and red. An arrow points from this matrix to the right, where it is decomposed into three components: \mathbf{U} (a black matrix), $\boldsymbol{\Sigma}$ (a diagonal matrix represented by a black rectangle with a diagonal line), and \mathbf{V}^T (a matrix with three columns, each represented by a vertical bar of a different color: green, blue, and red).

- Truncate: $\Phi = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$

POD–Galerkin: online projection

Full-order model:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}; t, \mu), \quad \mathbf{x}(0, \mu) = \mathbf{x}^0(\mu)$$

1 $\mathbf{x}(t) \approx \tilde{\mathbf{x}}(t) = \Phi \hat{\mathbf{x}}(t)$

$$\begin{array}{c|c|c} \mathbf{x}(t) & \approx & \tilde{\mathbf{x}}(t) \\ \hline \mathbf{x} & \approx & \Phi \hat{\mathbf{x}} \end{array}$$

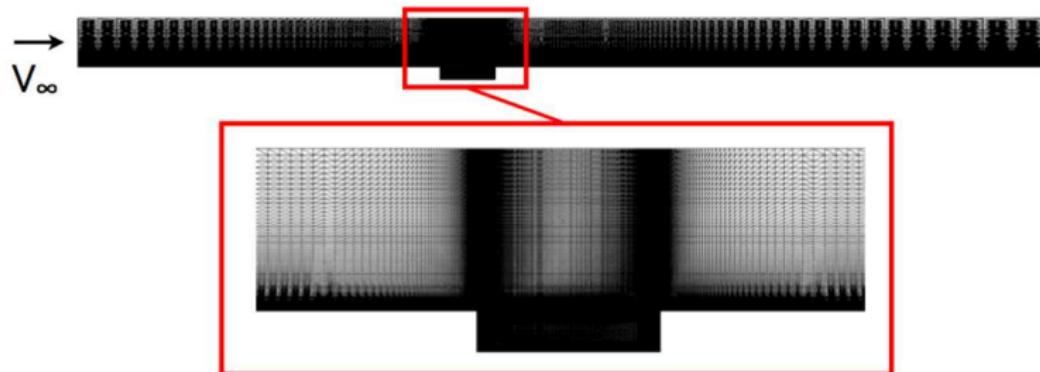
2 $\Phi^T (\mathbf{f}(\tilde{\mathbf{x}}, t) - \frac{d\tilde{\mathbf{x}}}{dt}) = 0$

$$\Phi^T \left(\mathbf{f}(\tilde{\mathbf{x}}, t) - \frac{d\tilde{\mathbf{x}}}{dt} \right) = 0$$

Galerkin ROM:

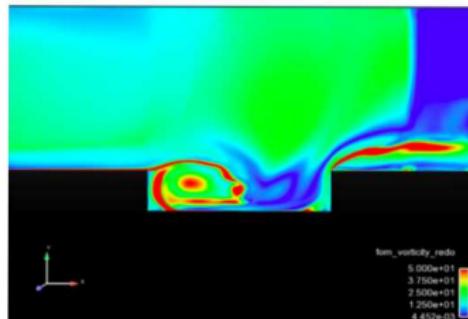
$$\frac{d\hat{\mathbf{x}}}{dt} = \Phi^T \mathbf{f}(\Phi \hat{\mathbf{x}}; t, \mu), \quad \hat{\mathbf{x}}(0, \mu) = \Phi^T \mathbf{x}^0(\mu)$$

Cavity-flow problem. Collaborator: M. Barone (SNL)

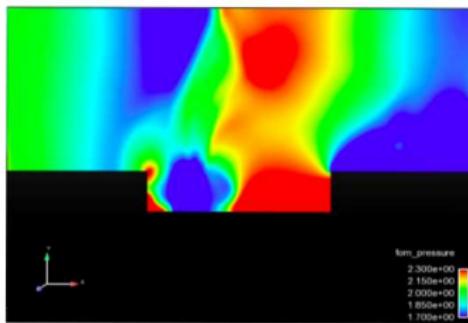


- Unsteady Navier–Stokes
- DES turbulence model
- 1.2 million degrees of freedom
- $Re = 6.3 \times 10^6$
- $M_\infty = 0.6$
- CFD code: AERO-F
[Farhat et al., 2003]

FOM responses

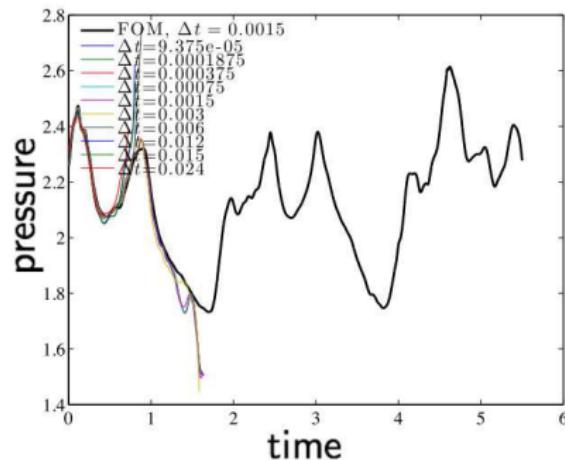


vorticity field



pressure field

POD–Galerkin failure



- Galerkin ROMs unstable

Our research goal

Nonlinear model-reduction methods that are accurate, low cost, certified, and reliable.

+ Accuracy

- Improve projection technique [?, C. et al., 2015a]
- Preserve problem structure [C. et al., 2012, C. et al., 2015c]

+ Low cost

- Sample-mesh approach [C. et al., 2011, C. et al., 2013]
- Leverage time-domain data [C. et al., 2015b]

+ Certification

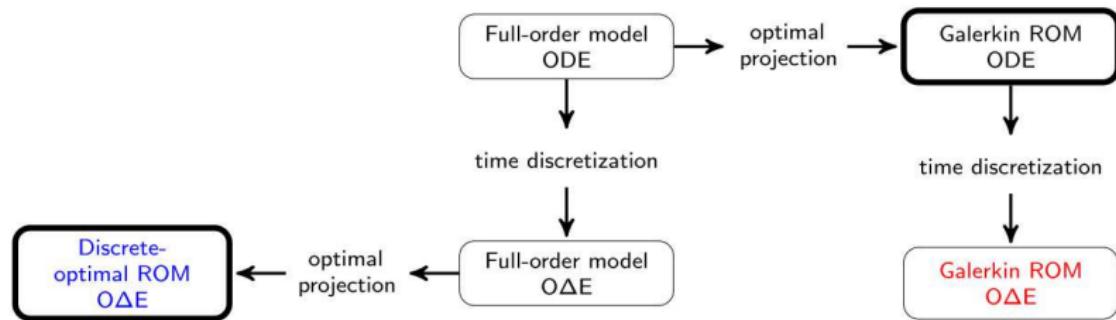
- Error bounds [C. et al., 2015a]
- Statistical error modeling [Drohmann and C., 2015]

+ Reliability

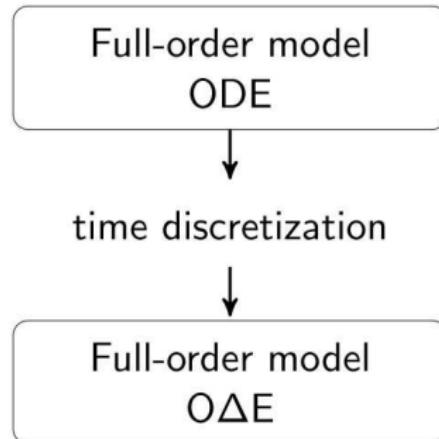
- *A posteriori* h -refinement [C., 2015]

How to construct a ROM for nonlinear dynamical systems?

- Optimize then discretize? (Galerkin)
- Discretize then optimize? (discrete optimal)



- Outstanding questions:
 - 1 Time-continuous and time-discrete representations?
 - 2 Are the two techniques ever equivalent?
 - 3 Discrete-time error bounds?



Full-order model (FOM)

- ODE: time continuous

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}^0, \quad t \in [0, T]$$

- OΔE, linear multistep schemes: $\boxed{\mathbf{r}^n(\mathbf{x}^n) = 0}, n = 1, \dots, N$

$$\mathbf{r}^n(\mathbf{x}) := \alpha_0 \mathbf{x} - \delta t \beta_0 \mathbf{f}(\mathbf{x}, t^n) + \sum_{j=1}^k \alpha_j \mathbf{x}^{n-j} - \delta t \sum_{j=1}^k \beta_j \mathbf{f}(\mathbf{x}^{n-j}, t^{n-j})$$

$$\mathbf{x}^n = \mathbf{x}^n \text{ (explicit state update)}$$

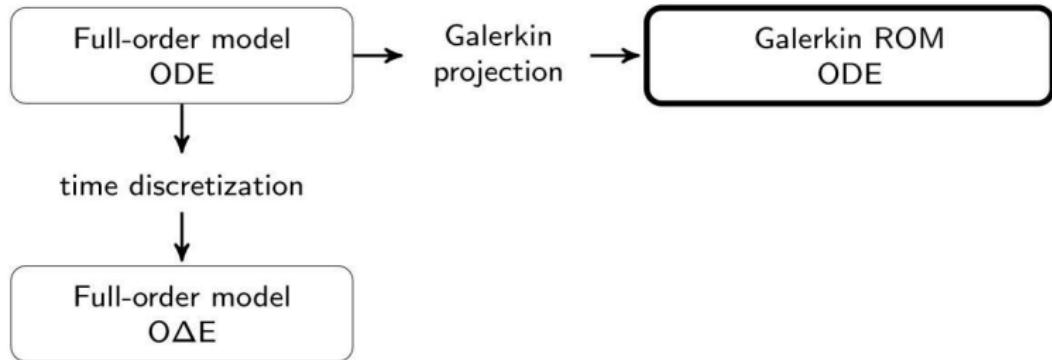
- OΔE, Runge–Kutta: $\boxed{\mathbf{r}_i^n(\mathbf{x}_1^n, \dots, \mathbf{x}_s^n) = 0}, i = 1, \dots, s$

$$\mathbf{r}_i^n(\mathbf{x}_1, \dots, \mathbf{x}_s) := \mathbf{x}_i - \mathbf{f}(\mathbf{x}^{n-1} + \delta t \sum_{j=1}^s a_{ij} \mathbf{x}_j, t^{n-1} + c_i \delta t)$$

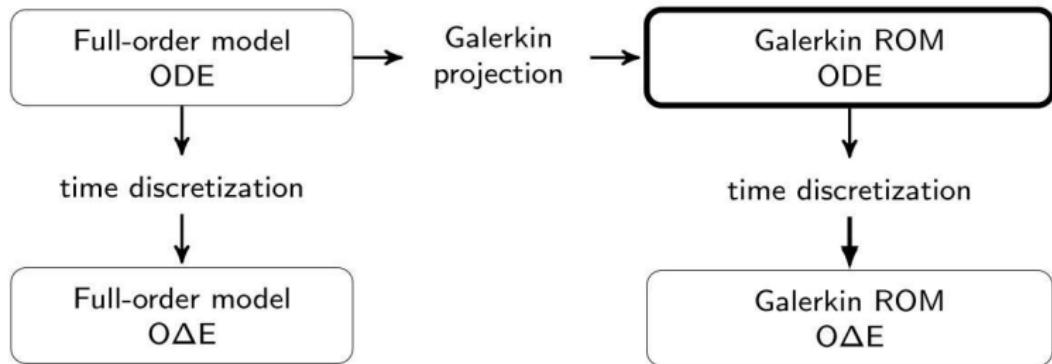
$$\mathbf{x}^n = \mathbf{x}^{n-1} + \delta t \sum_{i=1}^s b_i \mathbf{x}_i^n \text{ (explicit state update)}$$

This talk focuses on linear multistep schemes.

Galerkin ROM: first optimize



Galerkin: first optimize, then discretize



Galerkin ROM

■ ODE

$$\frac{d\hat{\mathbf{x}}}{dt} = \Phi^T \mathbf{f}(\Phi \hat{\mathbf{x}}, t), \quad \hat{\mathbf{x}}(0) = \Phi^T \mathbf{x}^0, \quad t \in [0, T]$$

+ Continuous velocity $\frac{d\hat{\mathbf{x}}}{dt}$ is optimal

Theorem (Galerkin ROM: continuous optimality)

The Galerkin ROM velocity minimizes the time-continuous FOM residual:

$$\frac{d\tilde{\mathbf{x}}}{dt}(\mathbf{x}, t) = \arg \min_{\mathbf{v} \in \text{range}(\Phi)} \|\mathbf{v} - \mathbf{f}(\mathbf{x}, t)\|_2^2$$

■ OΔE

$$\hat{\mathbf{r}}^n(\hat{\mathbf{x}}^n) = 0, \quad n = 1, \dots, N$$

$$\hat{\mathbf{r}}^n(\hat{\mathbf{x}}) := \alpha_0 \hat{\mathbf{x}} - \delta t \beta_0 \Phi^T \mathbf{f}(\Phi \hat{\mathbf{x}}, t^n) + \sum_{j=1}^k \alpha_j \hat{\mathbf{x}}^{n-j} - \delta t \sum_{j=1}^k \beta_j \Phi^T \mathbf{f}(\Phi \hat{\mathbf{x}}^{n-j}, t^{n-j})$$

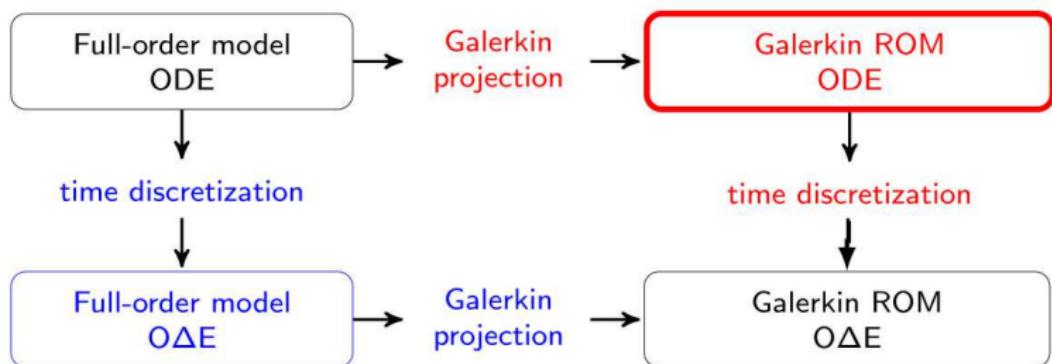
- Discrete state $\hat{\mathbf{x}}^n$ is not generally optimal

Galerkin ROM: Commutativity

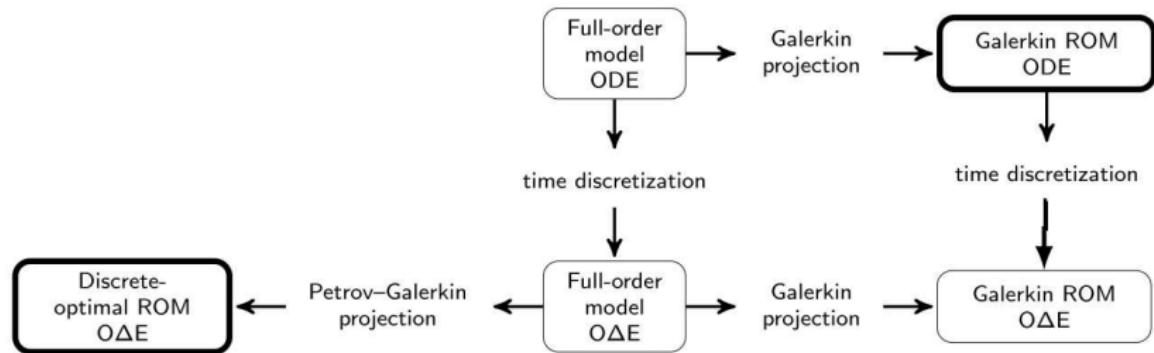
Theorem

Projection and time discretization are commutative for Galerkin ROMs:

$$\hat{r}^n(\hat{x}) = \Phi^T r^n(\Phi \hat{x})$$



Discrete-optimal ROM: first discretize, then optimize



Discrete-optimal ROM

- FOM OΔE

$$\mathbf{r}^n(\mathbf{x}^n) = 0, \quad n = 1, \dots, N$$

- Discrete-optimal ROM OΔE:

$$\hat{\mathbf{x}}^n = \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \|\mathbf{A}\mathbf{r}^n(\Phi\hat{\mathbf{z}})\|_2^2.$$

↔

$$\Psi^n(\hat{\mathbf{x}}^n)^T \mathbf{r}^n(\Phi\hat{\mathbf{x}}^n) = 0, \quad \Psi^n(\hat{\mathbf{x}}) := \mathbf{A}^T \mathbf{A} \frac{\partial \mathbf{r}^n}{\partial \mathbf{x}}(\Phi\hat{\mathbf{x}})$$

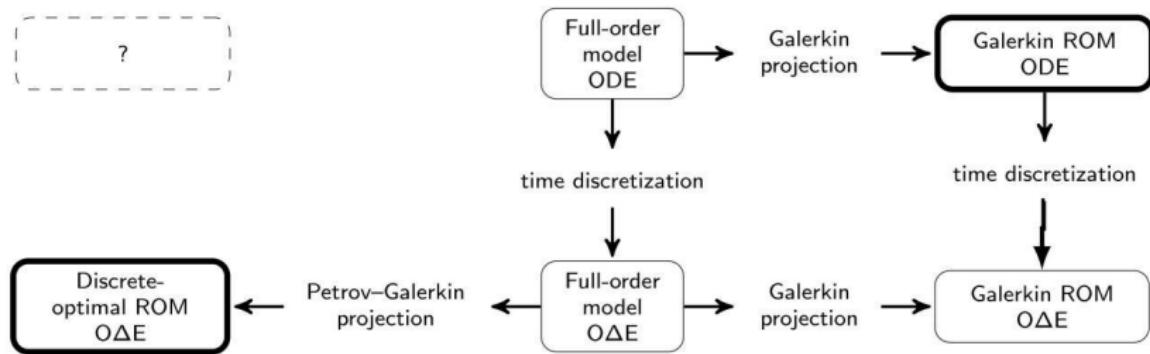
- $\mathbf{A} = \mathbf{I}$: Least-squares Petrov–Galerkin

[LeGresley, 2006, Bui-Thanh et al., 2008, ?]

- Alternative norm: ℓ^1 [Abgrall and Amsallem, 2015]

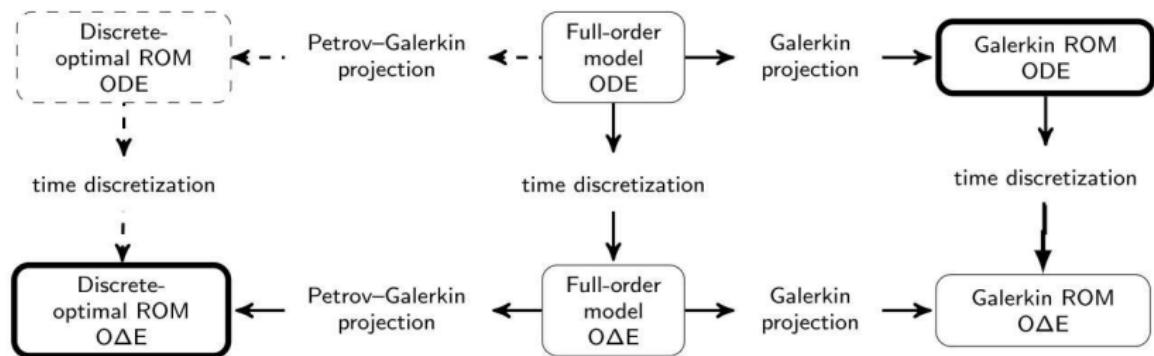
- + Discrete solution is optimal

Does the discrete-optimal ROM have a time-continuous representation?



Does the discrete-optimal ROM have a time-continuous representation?

Sometimes.



Discrete-optimal ROM: continuous representation

Theorem

The discrete-optimal ROM is equivalent to applying a Petrov–Galerkin projection to the FOM ODE with test basis

$$\Psi(\hat{\mathbf{x}}, t) = \mathbf{A}^T \mathbf{A} \left(\alpha_0 \mathbf{I} - \Delta t \beta_0 \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}^0 + \Phi \hat{\mathbf{x}}, t) \right) \Phi$$

if

- 1 $\beta_j = 0, j \geq 1$ (e.g., a single-step method),
- 2 the velocity \mathbf{f} is linear in the state, or
- 3 $\beta_0 = 0$ (i.e., explicit schemes).

Time-continuous test basis depends on
time-discretization parameters!

Are the two approaches ever equivalent?

- Galerkin: $\Phi^T \mathbf{r}^n(\Phi \hat{\mathbf{x}}^n) = 0$
- Discrete-optimal: $\Psi^n(\hat{\mathbf{x}}^n)^T \mathbf{r}^n(\Phi \hat{\mathbf{x}}^n) = 0$

Does $\Psi^n(\hat{\mathbf{x}}^n) = \Phi$ ever?

Yes.

$$\Psi^n(\hat{\mathbf{x}}) := \mathbf{A}^T \mathbf{A} \frac{\partial \mathbf{r}^n}{\partial \mathbf{x}}(\Phi \hat{\mathbf{x}}) = \mathbf{A}^T \mathbf{A} \left(\alpha_0 \mathbf{I} - \delta t \beta_0 \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\Phi \hat{\mathbf{x}}, t^n) \right) \Phi$$

Theorem

The two approaches are equivalent ($\Psi^n(\hat{\mathbf{x}}) = \Phi$)

- 1 *in the limit of $\delta t \rightarrow 0$ with $\mathbf{A} = 1/\sqrt{\alpha_0} \mathbf{I}$,*
- 2 *if the scheme is explicit ($\beta_0 = 0$) with $\mathbf{A} = 1/\sqrt{\alpha_0} \mathbf{I}$, or*
- 3 *if $\frac{\partial \mathbf{r}^n}{\partial \mathbf{x}}$ is positive definite with $[\frac{\partial \mathbf{r}^n}{\partial \mathbf{x}}]^{-1} = \mathbf{A}^T \mathbf{A}$.*

Discrete-time error bound

Theorem

If the following conditions hold:

- 1 $\mathbf{f}(\cdot, t)$ is Lipschitz continuous with Lipschitz constant κ , and
- 2 δt is such that $0 < h := |\alpha_0| - |\beta_0| \kappa \delta t$,

then

$$\|\delta \mathbf{x}_G^n\| \leq \frac{\delta t}{h} \sum_{\ell=0}^k |\beta_\ell| \|(\mathbf{I} - \textcolor{red}{V}) \mathbf{f}(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_G^{n-\ell})\| + \frac{1}{h} \sum_{\ell=1}^k (|\beta_\ell| \kappa \delta t + |\alpha_\ell|) \|\delta \mathbf{x}_G^{n-\ell}\|$$

$$\|\delta \mathbf{x}_D^n\| \leq \frac{\delta t}{h} \sum_{\ell=0}^k |\beta_\ell| \|(\mathbf{I} - \textcolor{blue}{P}^n) \mathbf{f}(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_D^{n-\ell})\| + \frac{1}{h} \sum_{\ell=1}^k (|\beta_\ell| \kappa \delta t + |\alpha_\ell|) \|\delta \mathbf{x}_D^{n-\ell}\|,$$

with

- $\delta \mathbf{x}_G^n := \mathbf{x}_*^n - \Phi \hat{\mathbf{x}}_G^n$
- $\delta \mathbf{x}_D^n := \mathbf{x}_*^n - \Phi \hat{\mathbf{x}}_D^n$
- $\textcolor{red}{V} := \Phi \Phi^T$
- $\textcolor{blue}{P}^n := \Phi ((\Psi^n)^T \Phi)^{-1} (\Psi^n)^T$

Discrete-optimal ROM yields a smaller error bound

Theorem (Backward Euler)

If conditions (1) and (2) hold, then

$$\|\delta \mathbf{x}_G^n\| \leq \delta t \sum_{j=0}^{n-1} \frac{1}{(h)^{j+1}} \underbrace{\|(\mathbf{I} - \mathbb{V}) \mathbf{f} \left(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_G^{n-j} \right)\|}_{\varepsilon_G^{n-j}}$$

$$\|\delta \mathbf{x}_D^n\| \leq \delta t \sum_{j=0}^{n-1} \frac{1}{(h)^{j+1}} \underbrace{\|(\mathbf{I} - \mathbb{P}^{n-j}) \mathbf{f} \left(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_D^{n-j} \right)\|}_{\varepsilon_D^{n-j}}$$

$$\varepsilon_G^k = \|\Phi \hat{\mathbf{x}}_G^k - \delta t \mathbf{f} \left(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_G^k \right) - \Phi \hat{\mathbf{x}}_G^{k-1}\|$$

$$\varepsilon_D^k = \|\Phi \hat{\mathbf{x}}_D^k - \delta t \mathbf{f} \left(\mathbf{x}_0 + \Phi \hat{\mathbf{x}}_D^k \right) - \Phi \hat{\mathbf{x}}_D^{k-1}\| = \min_{\mathbf{y}} \|\Phi \mathbf{y} - \delta t \mathbf{f} \left(\mathbf{x}_0 + \Phi \mathbf{y} \right) - \Phi \hat{\mathbf{x}}_D^{k-1}\|$$

Corollary (Discrete-optimal smaller error bound)

If $\hat{\mathbf{x}}_D^{k-1} = \hat{\mathbf{x}}_G^{k-1}$, then $\varepsilon_D^k \leq \varepsilon_G^k$.

Corollary (Backward Euler)

Define

- $\Delta \hat{\mathbf{x}}_D^j := \hat{\mathbf{x}}_D^j - \hat{\mathbf{x}}_D^{j-1}$ and
- $\Delta \bar{\mathbf{x}}^j$: full-space solution increment from $\hat{\mathbf{x}}_D^{j-1}$.

Then, the discrete-optimal error can also be bounded as

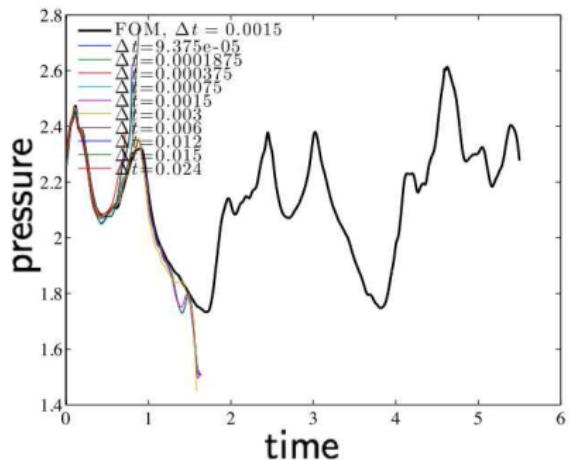
$$\|\delta \mathbf{x}_D^n\| \leq \delta t(1 + \kappa \delta t) \sum_{j=0}^{n-1} \frac{\mu^{n-j}}{(h)^{j+1}} \|\mathbf{f}(\hat{\mathbf{x}}_D^{j-1} + \Delta \bar{\mathbf{x}}^{n-j})\|$$

with $\mu^j := \|\Phi \Delta \hat{\mathbf{x}}_D^j - \Delta \bar{\mathbf{x}}^j\| / \Delta \bar{\mathbf{x}}^j$.

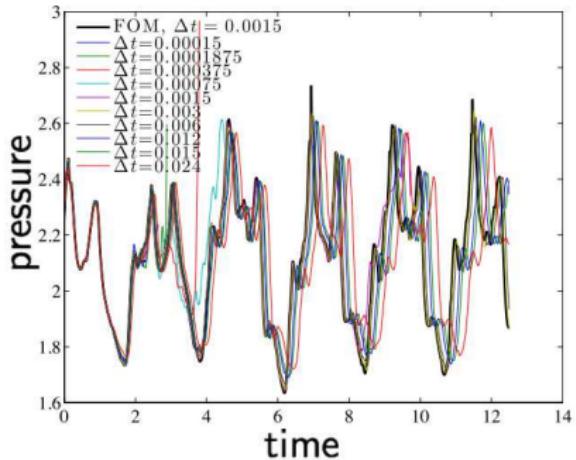
Effect of decreasing δt :

- + The terms $\delta t(1 + \kappa \delta t)$ and $1/(h)^{j+1}$ decrease
- The number of total time instances n increases
- ? The term μ^{n-j} may **increase** or **decrease**, depending on the spectral content of the basis Φ

Galerkin and discrete-optimal responses for basis dimension $p = 204$



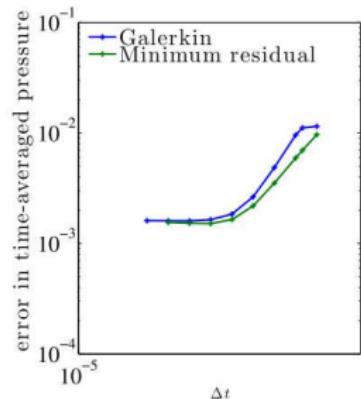
(a) Galerkin



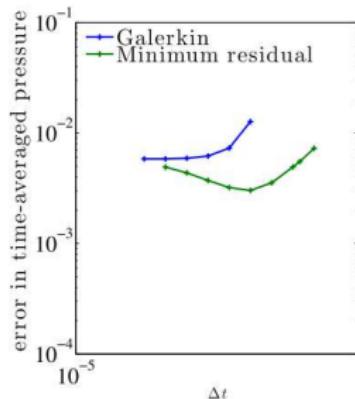
(b) Discrete optimal

- Galerkin ROMs unstable for long time intervals
(consistent with previous results [C. et al., 2013, ?, C., 2011])
- + Discrete-optimal ROMs accurate and stable (most time steps)

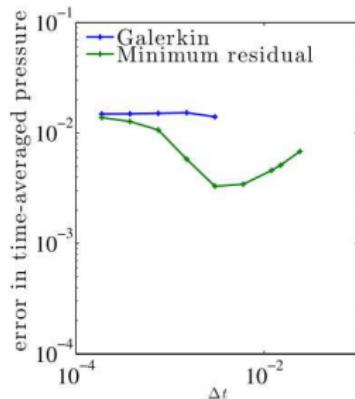
Discrete-optimal ROM: superior performance



(c) $0 \leq t \leq 0.55$



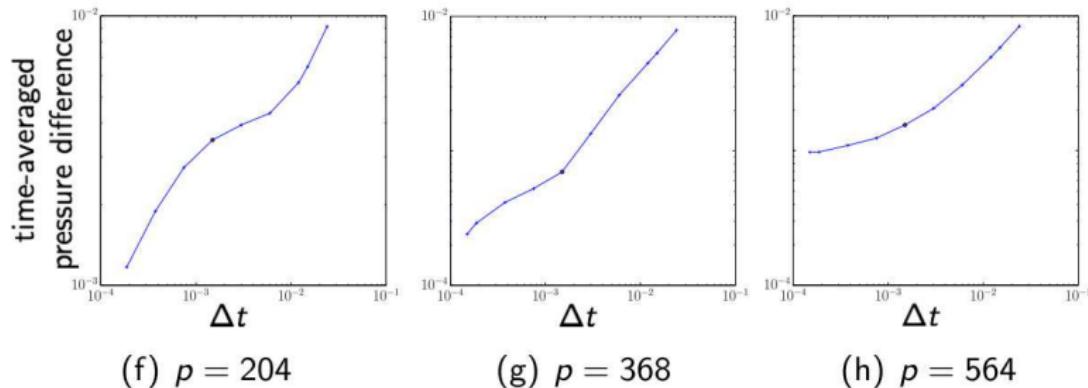
(d) $0 \leq t \leq 1.1$



(e) $0 \leq t \leq 1.54$

- ✓ Discrete-optimal ROM yields a **smaller error** for all time intervals and time steps.

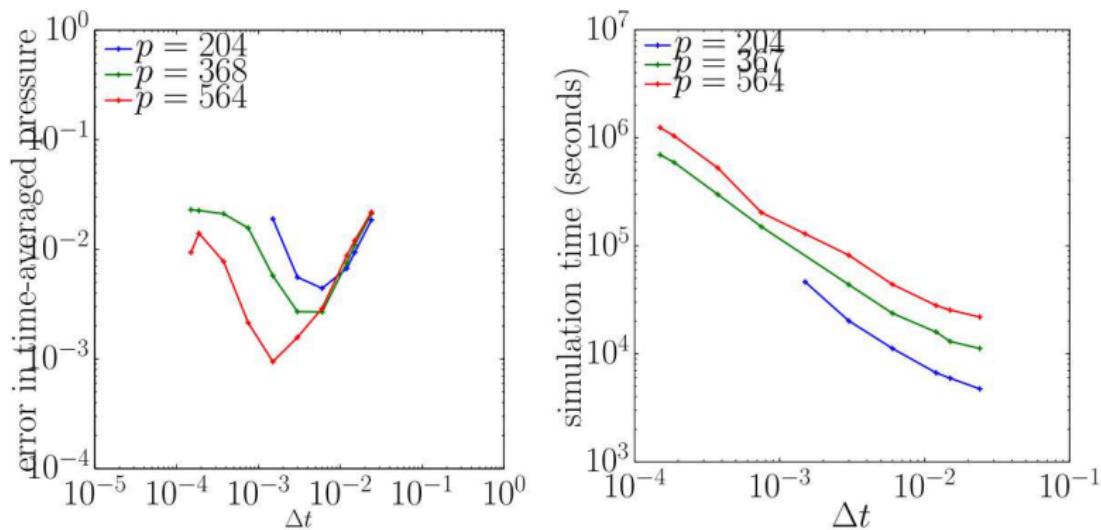
Limiting equivalence



Galerkin/discrete-optimal difference in the stable Galerkin interval
 $0 \leq t \leq 1.1$.

- ✓ The discrete-optimal ROM converges to Galerkin as $\delta t \rightarrow 0$.

Discrete-optimal performance in $0 \leq t \leq 2.5$



- ✓ An intermediate δt produces the **lowest error** and **better speedup**.

$p = 564$ case:

- $\delta t = 1.875 \times 10^{-4}$ sec: relative error = **1.40%**, time = **289 hrs**
- $\delta t = 1.5 \times 10^{-3}$ sec: relative error = **0.095%**, time = **35.8 hrs**

KTC: Still slow!!!

Our research goal

Nonlinear model-reduction methods that are accurate, low cost, certified, and reliable.

+ Accuracy

- Improve projection technique [?, C. et al., 2015a]
- Preserve problem structure [C. et al., 2012, C. et al., 2015c]

+ Low cost

- Sample-mesh approach [C. et al., 2011, C. et al., 2013]
- Leverage time-domain data [C. et al., 2015b]

+ Certification

- Error bounds [C. et al., 2015a]
- Statistical error modeling [Drohmann and C., 2015]

+ Reliability

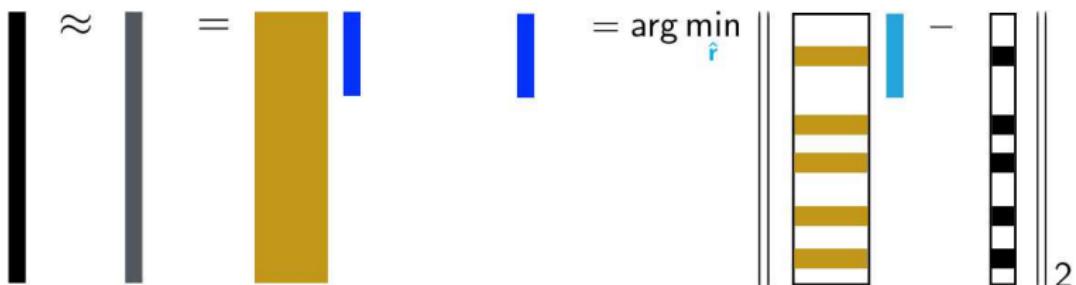
- *A posteriori* h -refinement [C., 2015]

Hyper-reduction via Gappy POD [Everson and Sirovich, 1995]

$$\hat{\mathbf{x}}^n = \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \|\mathbf{A} \mathbf{r}^n(\Phi \hat{\mathbf{z}})\|_2^2.$$

Can we select \mathbf{A} to make this inexpensive?

1. $\mathbf{r}^n(\mathbf{x}) \approx \tilde{\mathbf{r}}^n(\mathbf{x}) = \Phi_R \hat{\mathbf{r}}^n(\mathbf{x})$ 2. $\hat{\mathbf{r}}^n(\mathbf{x}) = \arg \min_{\hat{\mathbf{r}}} \|\mathbf{P} \Phi_R \hat{\mathbf{r}} - \mathbf{P} \mathbf{r}^n(\mathbf{x})\|_2$



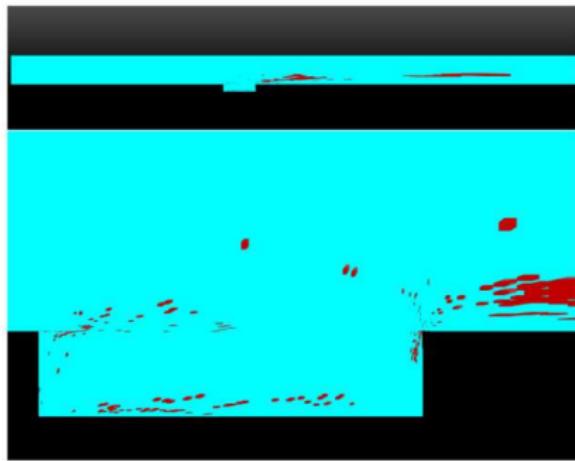
$$\begin{aligned} \hat{\mathbf{x}}^n &= \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \|\tilde{\mathbf{r}}^n(\Phi \hat{\mathbf{z}})\|_2^2 = \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \|\Phi_R \hat{\mathbf{r}}^n(\Phi \hat{\mathbf{z}})\|_2^2 = \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \|\hat{\mathbf{r}}^n(\Phi \hat{\mathbf{z}})\|_2^2 \\ &= \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \| \underbrace{(\mathbf{P} \Phi_R)^+ \mathbf{P}}_A \mathbf{r}^n(\Phi \hat{\mathbf{z}}) \|_2^2. \end{aligned}$$

+ $\mathbf{A} = (\mathbf{P} \Phi_R)^+ \mathbf{P}$ leads to low-cost

■ Offline: Construct Φ_R (POD) and \mathbf{P} (greedy method)

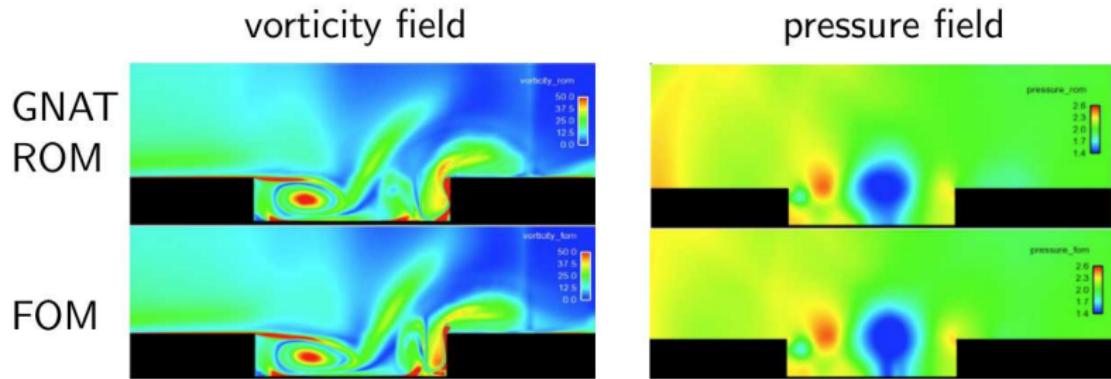
Cavity-flow problem: GNAT

$$\hat{\mathbf{x}}^n = \arg \min_{\hat{\mathbf{z}} \in \mathbb{R}^p} \| (\mathbf{P} \Phi_R)^+ \mathbf{P} \mathbf{r}^n (\Phi \hat{\mathbf{z}}) \|_2^2$$



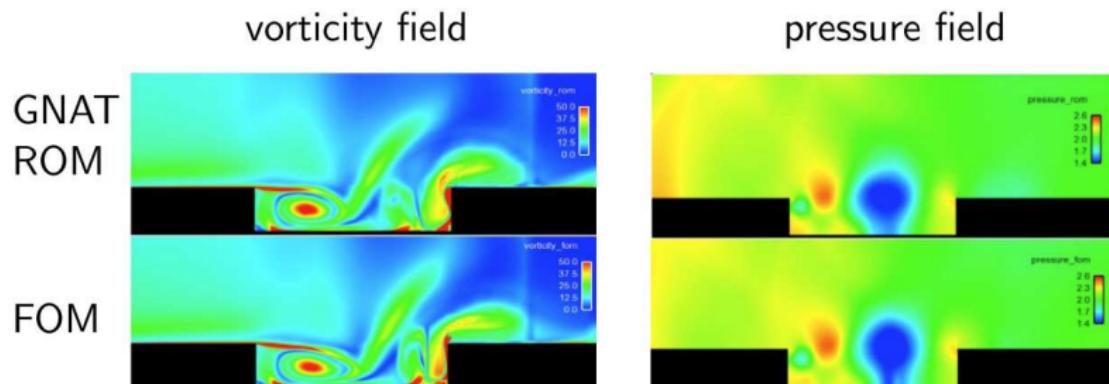
- Sample mesh: 4.1% nodes, 3.0% cells
- + Small problem size: can run on many fewer cores

GNAT performance



- + $< 1\%$ error in time-averaged drag
- FOM: 5 hour \times 48 CPU
- GNAT ROM: 32 min \times 2 CPU
- + 229x CPU-hour savings

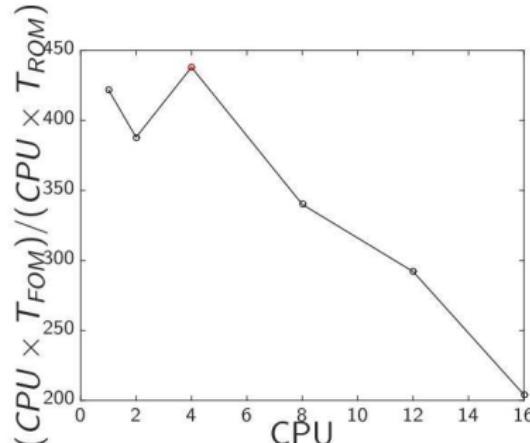
GNAT performance



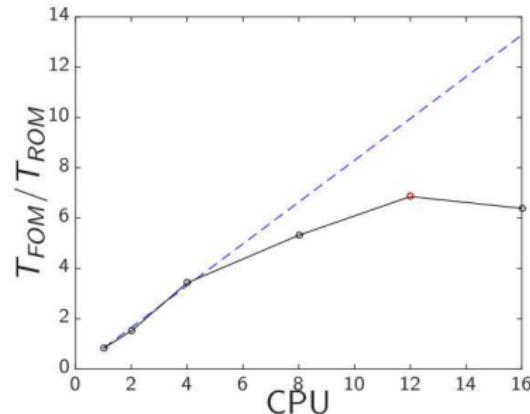
- + < 1% error in time-averaged drag
- FOM: 5 hour x 48 CPU
- GNAT ROM: 32 min x 2 CPU
- + 229x CPU-hour savings. Good for many query.
- 9.4x walltime savings. Bad for **real time**.

Why?

GNAT: strong scaling (Ahmed body) [C., 2011]



(c) CPU-hour savings



(d) Walltime savings

- + Significant CPU-hour savings (max: 438 for 4 CPU)
- Modest walltime savings (max: 7 for 12 CPU)

Spatial parallelism is quickly saturated!

Our research goal

Nonlinear model-reduction methods that are accurate, low cost, certified, and reliable.

+ Accuracy

- Improve projection technique [?, C. et al., 2015a]
- Preserve problem structure [C. et al., 2012, C. et al., 2015c]

+ Low cost

- Sample-mesh approach [C. et al., 2011, C. et al., 2013]
- Leverage time-domain data [C. et al., 2015b]

+ Certification

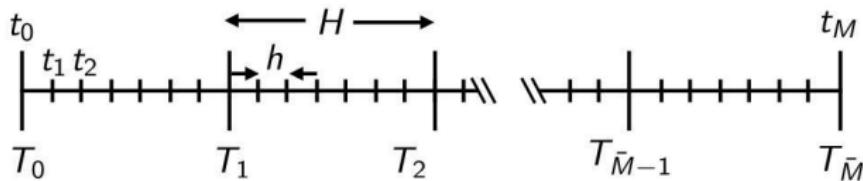
- Error bounds [C. et al., 2015a]
- Statistical error modeling [Drohmann and C., 2015]

+ Reliability

- *A posteriori* h -refinement [C., 2015]

Time-parallel algorithms [Lions et al., 2001a, Farhat and Chandesris, 2003]

Goal: expose more parallelism to reduce walltime



- Fine propagator: time step h

$$\mathcal{F}(\mathbf{x}; \tau_1, \tau_2)$$

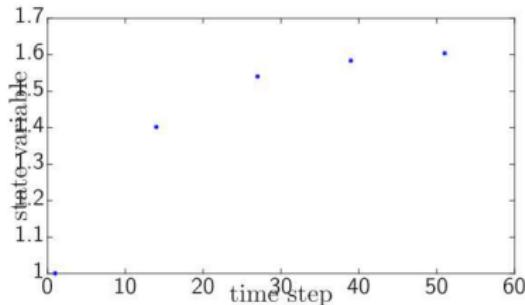
- Coarse propagator: time step H

$$\mathcal{G}(\mathbf{x}; \tau_1, \tau_2)$$

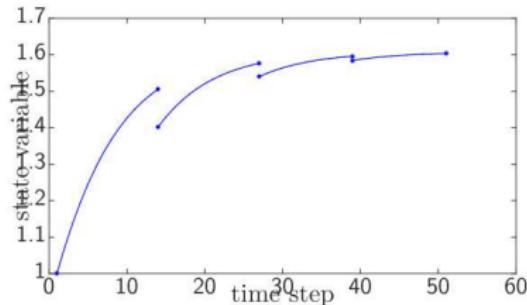
- Parareal iteration k (**sequential** and **parallel** steps):

$$\mathbf{x}_{k+1}^{m+1} = \mathcal{G}(\mathbf{x}_{k+1}^m; T_m, T_{m+1}) + \mathcal{F}(\mathbf{x}_k^m; T_m, T_{m+1}) - \mathcal{G}(\mathbf{x}_k^m; T_m, T_{m+1})$$

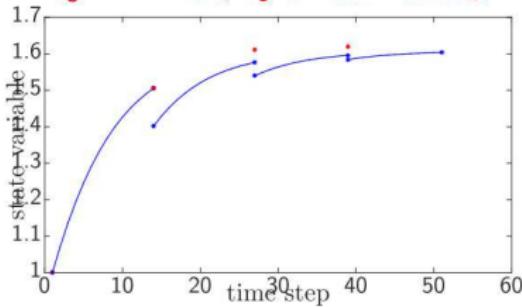
Illustration: sequential and parallel steps



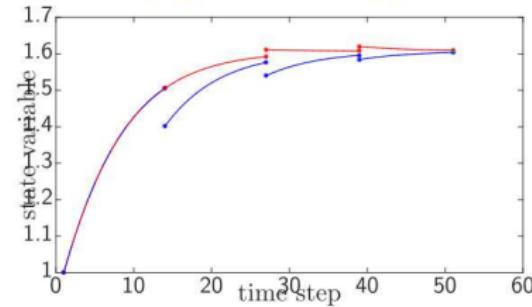
$$\mathbf{x}_0^{m+1} = \mathcal{G}(\mathbf{x}_0^m; T_m, T_{m+1})$$



$$\mathcal{F}(\mathbf{x}_0^m; T_m, T_{m+1})$$



$$\mathbf{x}_1^{m+1} = \mathcal{F}(\mathbf{x}_0^m; T_m, T_{m+1}) + \mathcal{G}(\mathbf{x}_1^m; T_m, T_{m+1}) - \mathcal{G}(\mathbf{x}_0^m; T_m, T_{m+1})$$



$$\mathcal{F}(\mathbf{x}_1^m; T_m, T_{m+1})$$

Coarse propagator

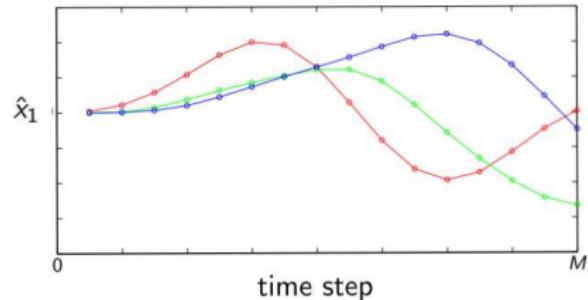
Critical: coarse propagator should be **fast, accurate, stable**

- Existing coarse propagators
 - Same integrator [Lions et al., 2001b, Bal and Maday, 2002]
 - Coarse spatial discretization
[Fischer et al., 2005, Farhat et al., 2006, Cortial and Farhat, 2009]
 - Simplified physics model [Baffico et al., 2002, Maday and Turinici, 2003, Blouza et al., 2011, Engblom, 2009, Maday, 2007]
 - Relaxed solver tolerance [Guibert and Tromeur-Dervout, 2007]
 - Reduced-order model (on the fly) [Farhat et al., 2006, Cortial and Farhat, 2009, Ruprecht and Krause, 2012, Chen et al., 2014]

Can we leverage offline data to improve the coarse propagator?

Revisit the SVD

$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \mathbf{U} \ \boldsymbol{\Sigma} \ \mathbf{v}^T$$



First row of \mathbf{v}^T

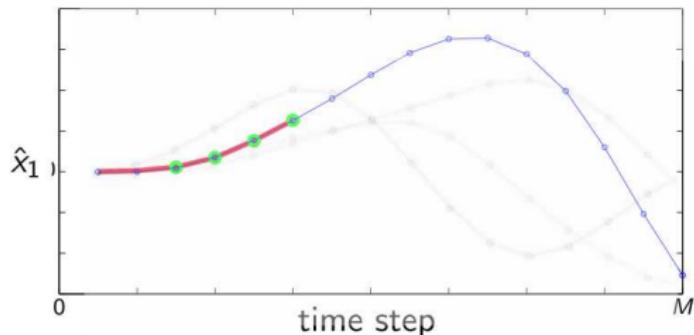
jth row of \mathbf{v}^T contains a basis for time evolution of \hat{x}_j

- Construct Ξ_j : basis for time evolution of \hat{x}_j

$$\Xi_j := [\xi_j^1 \ \dots \ \xi_j^{n_{\text{train}}}], \quad \xi_j^i := [v_{M(i-1)+1,j} \ \dots \ v_{Mi,j}]^T$$

Previous method [C. et al., 2015b]

- 1 compute forecast by gappy POD in time domain:



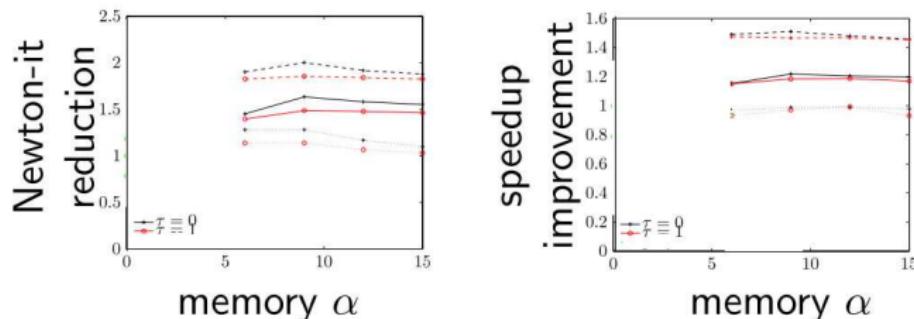
\hat{x}_1 so far; memory $\alpha = 4$; forecast; temporal basis

$$z_j = \arg \min_{z \in \mathbb{R}^{a_j}} \|Z(m-1, \alpha) \Xi_j z - Z(m-1, \alpha) g(\hat{x}_j)\|_2$$

- Time sampling: $Z(k, \beta) := [\mathbf{e}_{k-\beta} \ \cdots \ \mathbf{e}_k]^T$
- Time unrolling: $g(\hat{x}_j) : \hat{x}_j \mapsto [\hat{x}_j(t_0) \ \cdots \ \hat{x}_j(t_M)]^T$

- 2 use $\mathbf{e}_m^T \Xi_j z_j$ as *initial guess* for $\hat{x}_j(t_m)$ in Newton solver

Previous results: structural dynamics [C. et al., 2015b]

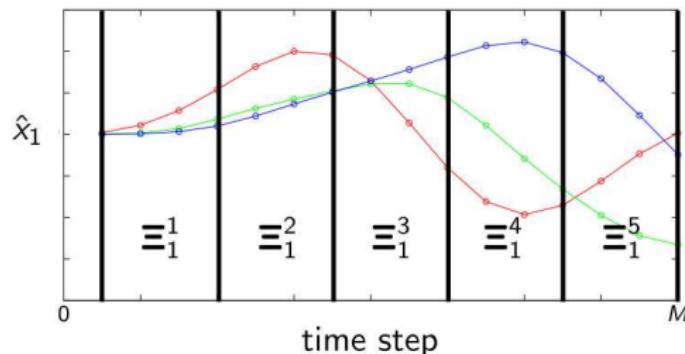


- + Newton iterations reduced by up to $\sim 2x$
- + Speedup improved by up to $\sim 1.5x$
- + No accuracy loss
- + Applicable to any nonlinear ROM
- Insufficient for real-time computation

Can we apply the same idea for the coarse propagator?

Coarse propagator for coordinate j and time interval m

- **Offline:** Construct time-evolution basis Ξ_j^m



- **Online:** Coarse propagator \mathcal{G}_j^m defined via forecasting:
 - 1 Compute α time steps with fine propagator
 - 2 Compute forecast via gappy POD
 - 3 Select last timestep of forecast

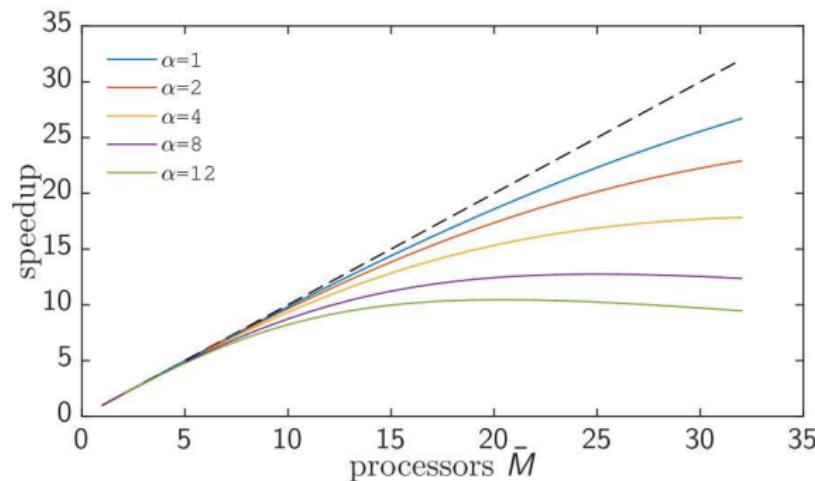
$$\mathcal{G}_j^m : (\hat{\mathbf{x}}_j; T_m, T_{m+1}) \mapsto \mathbf{e}_{H/h}^T \Xi_j^m [\mathbf{Z}(\alpha+1, \alpha) \Xi_j^m]^+ \begin{bmatrix} \mathcal{F}(\hat{\mathbf{x}}_j; T_m, T_m + h) \\ \vdots \\ \mathcal{F}(\hat{\mathbf{x}}_j; T_m, T_m + h\alpha) \end{bmatrix}$$

Ideal-conditions speedup

Theorem

If $g(\hat{x}_j) \in \text{range}(\Xi_j)$, $j = 1, \dots, p$, then the proposed method converges in one parareal iteration and realizes a theoretical speedup of

$$\frac{\bar{M}}{\bar{M}(\bar{M} - 1)\alpha/M + 1}.$$



Ideal-conditions speedup for $M = 5000$

Ideal-conditions speedup with initial guesses

Corollary

If f is nonlinear, $g(\hat{x}_j) \in \text{range}(\Xi_j)$, $j = 1, \dots, p$, and the forecasting method also provides Newton-solver initial guesses, then

- 1 the method converges in **one parareal iteration**, and
- 2 only α nonlinear systems of algebraic equations are solved in each time interval.

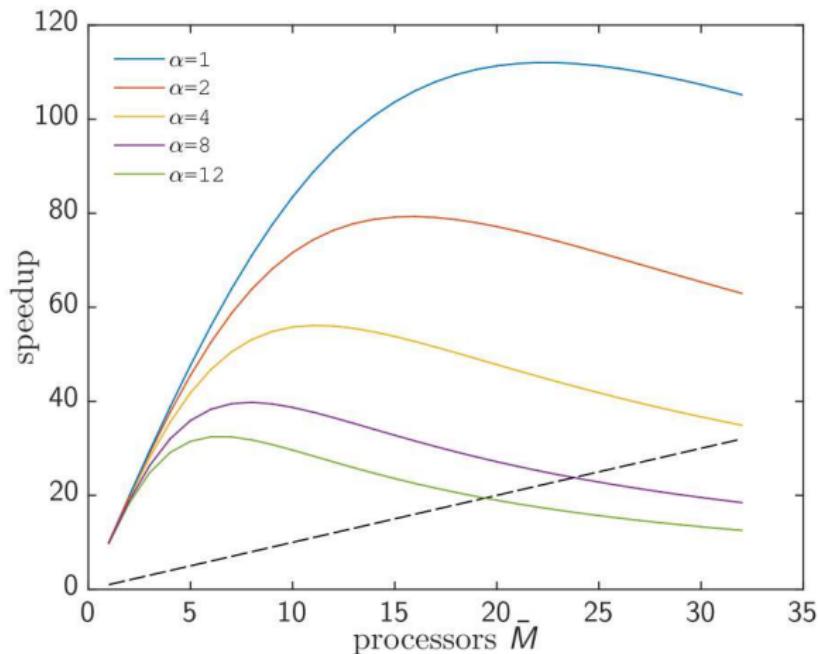
The method then realizes a theoretical speedup of

$$\frac{M}{(\bar{M}\alpha) + (M/\bar{M} - \alpha)\tau_r}$$

relative to the sequential algorithm without forecasting. Here,

$$\tau_r = \frac{\text{residual computation time}}{\text{nonlinear-system solution time}}.$$

Ideal-conditions speedup with initial-guesses



Ideal-condition speedup for $M = 5000$, $\tau_r = 1/10$

Significant speedups possible by leveraging time-domain data!

Stability

Theorem

If the fine propagator is stable, i.e.,

$$\|\mathcal{F}(\mathbf{x}; \tau_1, \tau_2)\| \leq (1 + C_{\mathcal{F}} H) \|\mathbf{x}\|,$$

then the proposed method is also stable, i.e.,

$$\|\hat{\mathbf{x}}_{k+1}^m\| \leq C_m \exp(C_F m H) \|\hat{\mathbf{x}}^0\|.$$

- $C_m := \sum_{k=1}^m \binom{k}{m} \beta_k \gamma^m \alpha^k (H/h)^{m-k}$
- $\beta_k := \exp(-C_{\mathcal{F}} k (H - h\alpha)) \leq 1$
- $\gamma := \max(\max_{m,j} 1/\|\mathbf{Z}(\alpha+1, \alpha) \Xi_j^m\|, 1/\sigma_{\min}(\mathbf{Z}(\alpha+1, \alpha) \Xi_j^m))$

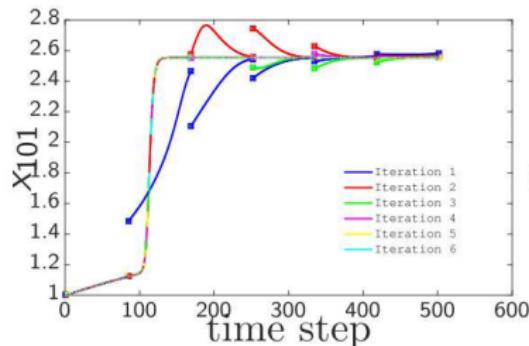
Example: inviscid Burgers equation [Rewienski, 2003]

$$\frac{\partial u(x, \tau)}{\partial \tau} + \frac{1}{2} \frac{\partial (u^2(x, \tau))}{\partial x} = 0.02e^{\mu_2 x}$$
$$u(0, \tau) = \mu_1, \quad \forall \tau \in [0, 25]$$
$$u(x, 0) = 1, \quad \forall x \in [0, 100],$$

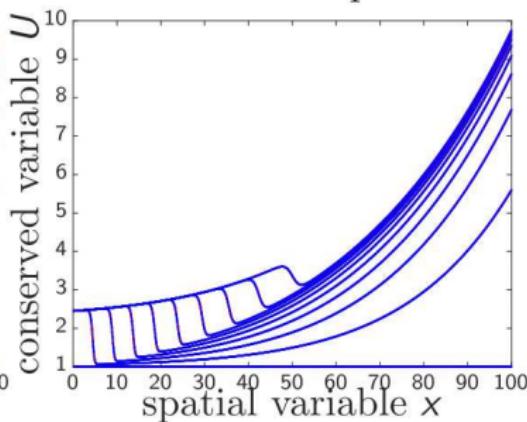
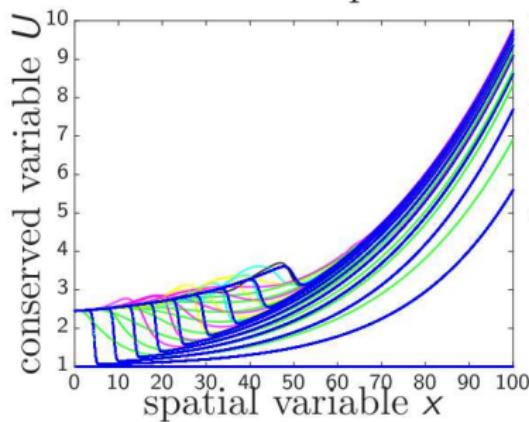
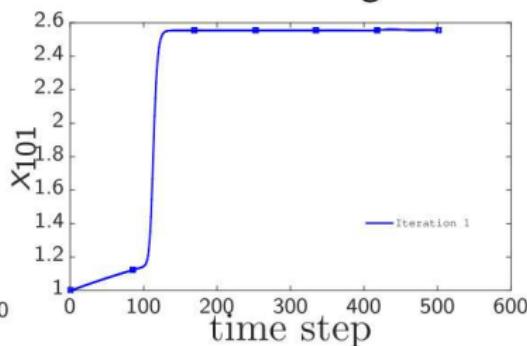
- Discretization: Godunov's scheme
- $(\mu_1, \mu_2) \in [2.5, 3.5] \times [0.02, 0.075]$
- $h = 0.1, M = 250$ fine time steps
- FOM: $N = 500$ degrees of freedom
- ROM: LSPG [?] with POD basis dimension $p = 100$
- $n_{\text{train}} = 4$ training points (LHS sampling); random online point
- Two coarse propagators: Backward Euler and forecasting

Forecasting outperforms backward Euler

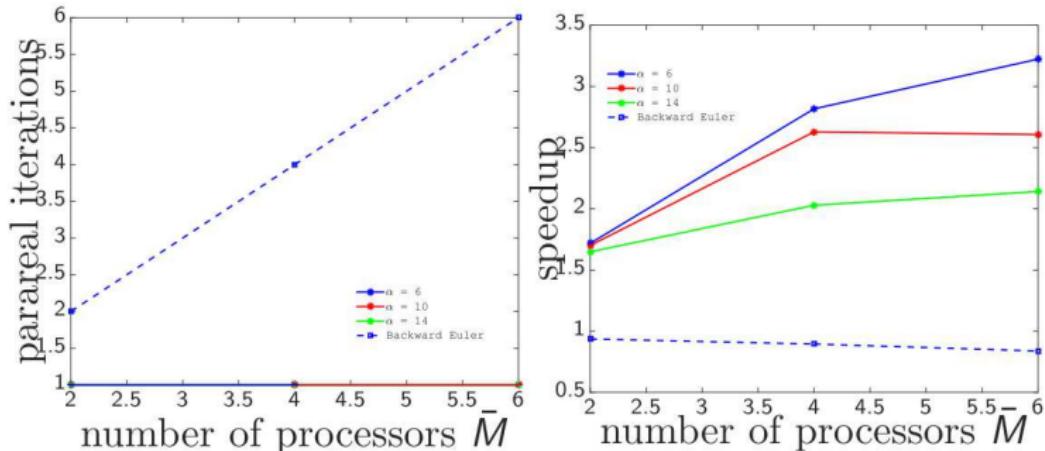
Backward Euler



Forecasting



Parareal performance



- + *Forecasting*: minimum possible iterations
- *Backward Euler*: maximum possible iterations

More parallelism successfully exposed!

Conclusions

Use temporal data to reduce ROM simulation time

- **offline:** time-evolution bases from right singular vectors
- **online:** use as coarse propagator

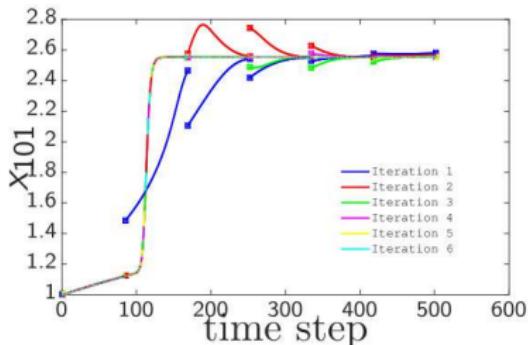
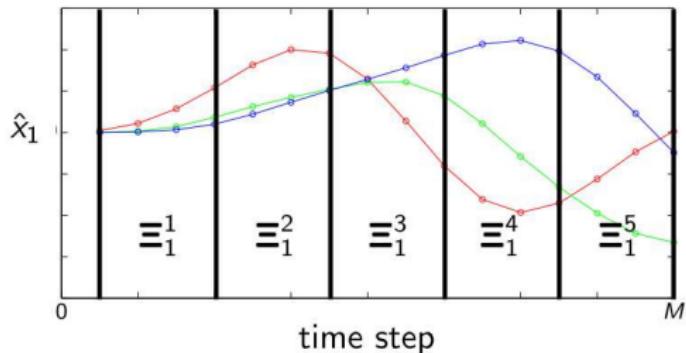
- 1 compute α time steps with fine propagator
- 2 use gappy POD to forecast

- + theory: excellent speedup and stability
- + ideal parareal performance observed
- + significant improvement over Backward Euler
- + no additional error introduced
- + generally applicable

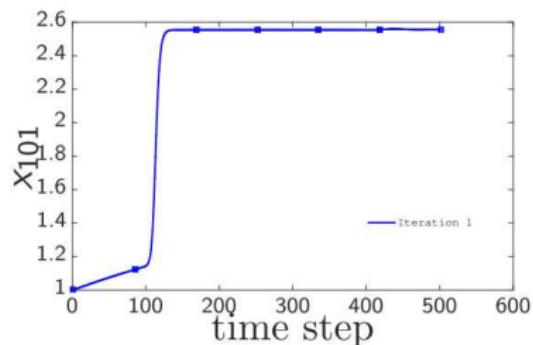
- **References:**

- K. Carlberg, L. Brenner, B. Haasdonk, and A. Barth.
"Time-parallel reduced-order models via forecasting," *in preparation*.
- K. Carlberg, J. Ray, and B. van Bloemen Waanders.
"Decreasing the temporal complexity for nonlinear, implicit reduced-order models by forecasting," CMAME, Vol. 289, p. 79–103 (2015).

Questions?



Backward Euler



Forecasting

Acknowledgments

- This research was supported in part by an appointment to the Sandia National Laboratories Truman Fellowship in National Security Science and Engineering, sponsored by Sandia Corporation (a wholly owned subsidiary of Lockheed Martin Corporation) as Operator of Sandia National Laboratories under its U.S. Department of Energy Contract No. DE-AC04-94AL85000.

-  National power grid simulation capability: Needs and issues.
-  Abgrall, R. and Amsallem, D. (2015).
Robust model reduction by l^1 -norm minimization and approximation via dictionaries: Application to linear and nonlinear hyperbolic problems.
Stanford University Preprint.
-  Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003).
An introduction to mcmc for machine learning.
Machine learning, 50(1-2):5–43.
-  Antoulas, A. C. (2005).
Approximation of Large-Scale Dynamical Systems.
Society for Industrial and Applied Mathematics, Philadelphia, PA.
-  Baffico, L., Bernard, S., Maday, Y., Turinici, G., and Zérah, G. (2002).
Parallel-in-time molecular-dynamics simulations.

 Bai, Z. (2002).

Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems.

Applied Numerical Mathematics, 43(1):9–44.

 Bal, G. and Maday, Y. (2002).

A “parareal” time discretization for non-linear pdes with application to the pricing of an american put.

In *Recent developments in domain decomposition methods*, pages 189–202. Springer Berlin Heidelberg.

 Barrault, M., Maday, Y., Nguyen, N. C., and Patera, A. T. (2004).

An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations.

Comptes Rendus Mathématique Académie des Sciences, 339(9):667–672.

 Baur, U., Beattie, C., Benner, P., and Gugercin, S. (2011).

Interpolatory projection methods for parameterized model reduction.

SIAM Journal on Scientific Computing, 33(5):2489–2518.

-  Benner, P., Gugercin, S., and Willcox, K. (2015).
A survey of projection-based model reduction methods for parametric dynamical systems.
SIAM Review, 57(4):483–531.
-  Blouza, A., Boudin, L., and Kaber, S. M. (2011).
Parallel in time algorithms with reduction methods for solving chemical kinetics.
Communications in Applied Mathematics and Computational Science, 5(2):241–263.
-  Bui-Thanh, T., Willcox, K., and Ghattas, O. (2008).
Model reduction for large-scale systems with high-dimensional parametric input space.
SIAM Journal on Scientific Computing, 30(6):3270–3288.
-  C., K. (2011).

*Model Reduction of Nonlinear Mechanical Systems via
Optimal Projection and Tensor Approximation.*
PhD thesis, Stanford University.

-  C., K. (2015).
Adaptive h -refinement for reduced-order models.
International Journal for Numerical Methods in Engineering,
102(5):1192–1210.
-  C., K., Barone, M., and Antil, H. (2015a).
Galerkin v. discrete-optimal projection in nonlinear model
reduction.
arXiv e-print, (1504.03749).
-  C., K., Cortial, J., Amsallem, D., Zahr, M., and Farhat, C. (2011).
The GNAT nonlinear model reduction method and its
application to fluid dynamics problems.
*AIAA Paper 2011-3112, 6th AIAA Theoretical Fluid Mechanics
Conference, Honolulu, HI.*

-  C., K., Farhat, C., Cortial, J., and Amsallem, D. (2013).
The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows.
Journal of Computational Physics, 242:623–647.
-  C., K., Ray, J., and van Bloemen Waanders, B. (2015b).
Decreasing the temporal complexity for nonlinear, implicit reduced-order models by forecasting.
Computer Methods in Applied Mechanics and Engineering, 289:79–103.
-  C., K., Tuminaro, R., and Boggs, P. (2012).
Efficient structure-preserving model reduction for nonlinear mechanical systems with application to structural dynamics.
In *AIAA Paper 2012-1969, 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Honolulu, Hawaii*.
-  C., K., Tuminaro, R., and Boggs, P. (2015c).

Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics.

SIAM J. Sci. Comput., 37(2):B153—B184.

-  Chen, F., Hesthaven, J. S., and Zhu, X. (2014).
On the use of reduced basis methods to accelerate and stabilize the parareal method.
In *Reduced Order Methods for Modeling and Computational Reduction*, pages 187–214. Springer.
-  Cortial, J. and Farhat, C. (2009).
A time-parallel implicit method for accelerating the solution of non-linear structural dynamics problems.
International Journal for Numerical Methods in Engineering, 77(4):451.
-  Drohmann, M. and C., K. (2015).
The romes method for reduced-order-model uncertainty quantification.

SIAM/ASA Journal on Uncertainty Quantification,
3(1):116–145.

-  Eftang, J. L. and Patera, A. T. (2013).
Port reduction in parametrized component static condensation: approximation and a posteriori error estimation.
International Journal for Numerical Methods in Engineering, 96(5):269–302.
-  Engblom, S. (2009).
Parallel in time simulation of multiscale stochastic chemical kinetics.
Multiscale Modeling & Simulation, 8(1):46–68.
-  Everson, R. and Sirovich, L. (1995).
Karhunen–Loève procedure for gappy data.
Journal of the Optical Society of America A, 12(8):1657–1664.
-  Farhat, C. and Chandresris, M. (2003).

Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications.

International Journal for Numerical Methods in Engineering, 58(9):1397–1434.

 Farhat, C., Cortial, J., Dastillung, C., and Bavestrello, H. (2006).

Time-parallel implicit integrators for the near-real-time prediction of linear structural dynamic responses.

International Journal for Numerical Methods in Engineering, 67:697–724.

 Farhat, C., Geuzaine, P., and Brown, G. (2003).

Application of a three-field nonlinear fluid-structure formulation to the prediction of the aeroelastic parameters of an F-16 fighter.

Computers & Fluids, 32(1):3–29.

 Fischer, P. F., Hecht, F., and Maday, Y. (2005).

A parareal in time semi-implicit approximation of the navier-stokes equations.

In *Domain decomposition methods in science and engineering*, pages 433–440. Springer.

 Freund, R. (2003).

Model reduction methods based on Krylov subspaces.
Acta Numerica, 12:267–319.

 Gallivan, K., Vandendorpe, A., and Van Dooren, P. (2004).

Model reduction of mimo systems via tangential interpolation.
SIAM Journal on Matrix Analysis and Applications, 26(2):328–349.

 Guibert, D. and Tromeur-Dervout, D. (2007).

Adaptive parareal for systems of odes.
In *Domain decomposition methods in science and engineering XVI*, pages 587–594. Springer.

 Ionita, A. and Antoulas, A. (2014).

Data-driven parametrized model reduction in the loewner framework.

SIAM Journal on Scientific Computing, 36(3):A984–A1007.

-  Lefteriu, S. and Antoulas, A. C. (2010).
A new approach to modeling multiport systems from frequency-domain data.
Computer-Aided Design of Integrated Circuits and Systems,
IEEE Transactions on, 29(1):14–27.
-  LeGresley, P. A. (2006).
Application of Proper Orthogonal Decomposition (POD) to Design Decomposition Methods.
PhD thesis, Stanford University.
-  Lions, J., Maday, Y., and Turinici, G. (2001a).
A “parareal” in time discretization of pdes.
Comptes Rendus de l'Academie des Sciences Series I Mathematics, 332(7):661–668.
-  Lions, J.-L., Maday, Y., and Turinici, G. (2001b).

-  Maday, Y. (2007).
Parareal in time algorithm for kinetic systems based on model reduction.
High-dimensional partial differential equations in science and engineering, 41:183–194.
-  Maday, Y. and Rønquist, E. M. (2002).
A reduced-basis element method.
J. Sci Comput, 17(1-4):447–459.
-  Maday, Y. and Turinici, G. (2003).
Parallel in time algorithms for quantum control: Parareal time discretization scheme.
International journal of quantum chemistry, 93(3):223–228.
-  Moore, B. (1981).

Principal component analysis in linear systems: Controllability, observability, and model reduction.

Automatic Control, IEEE Transactions on, 26(1):17–32.

 Phuong Huynh, D. B., Knezevic, D. J., and Patera, A. T. (2013).

A static condensation reduced basis element method: approximation and a posteriori error estimation.

ESAIM: Mathematical Modelling and Numerical Analysis, 47(01):213–251.

 Prud'Homme, C., Rovas, D. V., Veroy, K., Machiels, L., Maday, Y., Patera, A. T., Turinici, G., et al. (2001).

Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods.

Journal of Fluids Engineering, 124(1):70–80.

 Rewienski, M. J. (2003).

A Trajectory Piecewise-Linear Approach to Model Order Reduction of Nonlinear Dynamical Systems.

-  Rowley, C. W. (2005).
Model reduction for fluids, using balanced proper orthogonal decomposition.
Int. J. on Bifurcation and Chaos, 15(3):997–1013.
-  Rozza, G., Huynh, D., and Patera, A. T. (2008).
Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations.
Archives of Computational Methods in Engineering, 15(3):229–275.
-  Ruprecht, D. and Krause, R. (2012).
Explicit parallel-in-time integration of a linear acoustic-advection system.
Computers & Fluids, 59:72–83.
-  Veroy, K., Prud'homme, C., Rovas, D. V., and Patera, A. T. (2003).

A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations.

AIAA Paper 2003-3847, 16th AIAA Computational Fluid Dynamics Conference, Orlando, FL.



Willcox, K. and Peraire, J. (2002).

Balanced model reduction via the proper orthogonal decomposition.

AIAA Journal, 40(11):2323–2330.