SAND2023-07903C

Sandia National Laboratories

Exceptional service in the national interest

CONSTRAINED TENSOR DECOMPOSITIONS AND CONSERVATION PRINCIPLES FOR DIRECT NUMERICAL SIMULATION DATA COMPRESSION

Danny Dunlavy, Hemanth Kolla, Eric Phipps, John Shadid, Edward Phillips

ICIAM 2023 -- Tokyo, Japan -- August 23, 2023

Minisymposium: [00353] Interpretable constrained tensor decompositions: models, algorithms, efficient implementations and applications

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

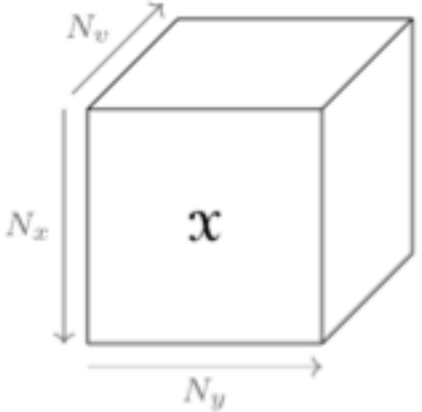
Sandia National Laboratories U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

U.S. DEPARTMENT OF ENERGY NNSA

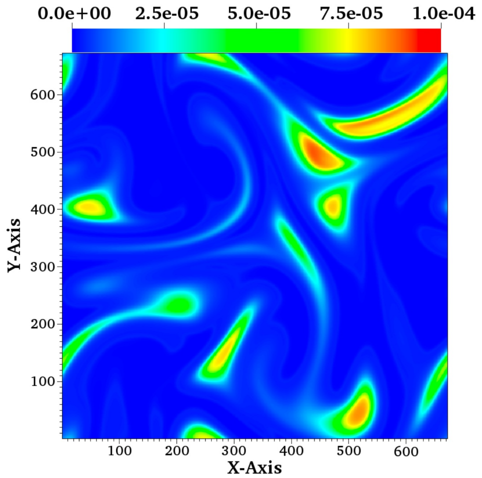
OUTLINE

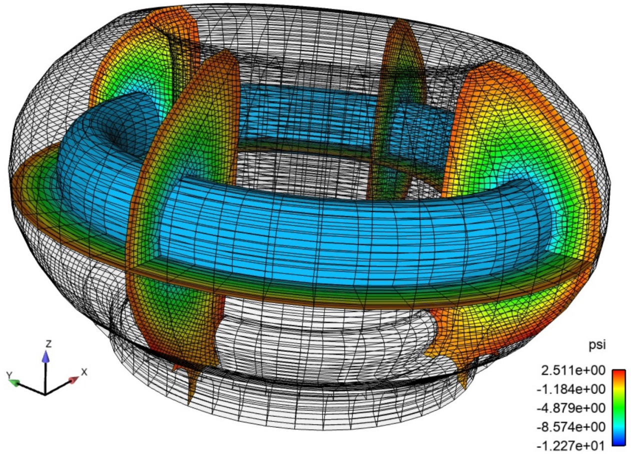
- **Motivating Problem**
  - Compression of numerical simulation data
- **Existing Method**
  - Low-rank decompositions of **tensor data** (i.e., multidimensional arrays)
- **New Method**
  - Goal-oriented tensor decompositions (Tucker/Canonical Polyadic Models)
  - Better modeling of quantities of interests (i.e., functions of simulation data)
- **Demonstration of New Approach on Multiple Applications**



Tensor Data

Combustion
compressible Navier-Stokes equations





Plasma Physics
compressible visco-resistive magnetohydrodynamics equations

MOTIVATING PROBLEM

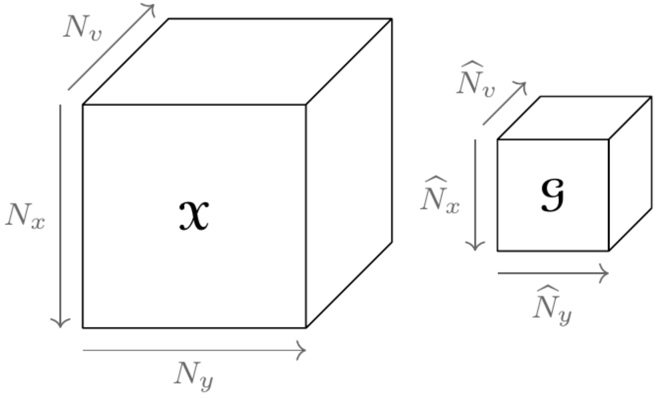
Physics Problem → Numerical Simulation → Data Compression → Scientific & Engineering Analysis

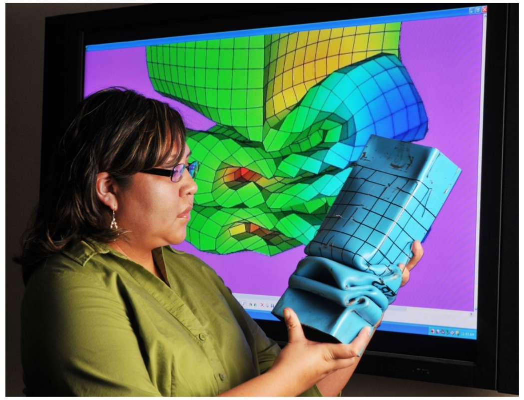
$$\frac{\partial \rho}{\partial t} = -\frac{\partial \rho u_i}{\partial x_i}$$

$$\frac{\partial \rho u_i}{\partial t} = -\frac{\partial \rho u_i u_j}{\partial x_j} - \frac{\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j}$$

$$\frac{\partial \rho e_t}{\partial t} = -\frac{\partial \rho e_t u_j}{\partial x_j} - \frac{\partial P u_j}{\partial x_j} + \frac{\partial (\tau_{ij} \cdot u_i)}{\partial x_j} - \frac{\partial q_j}{\partial x_j}$$

$$\frac{\partial \rho Y_k}{\partial t} = -\frac{\partial \rho Y_k u_j}{\partial x_j} - \frac{\partial J_{k,j}}{\partial x_j} + \omega_k$$





How should we compress the data to best support downstream analysis?

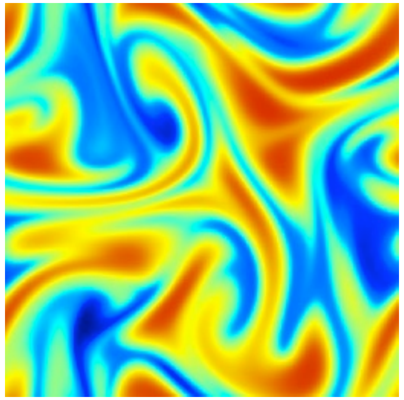
HOMOGENEOUS CHARGE COMPRESSION IGNITION (HCCI)

- Ignition processes in ethanol-air mixture under conditions similar to piston compression in **combustion** engine with exhaust gas recirculation [1]
- 2D simulation data in S3D, a compressible reacting flow solver [2]
  - 672 x 672 spatial grid
  - 50 time steps (snapshot from 626 time steps)
  - 32 variables: 28 chemical species, temperature, pressure, 2 velocities
  - Data: 672 x 672 x 32 x 50 (x x y x $variable$ x $time$) tensor

Problem: Compress data while capturing combustion dynamics

$$\frac{\partial \rho}{\partial t} = -\frac{\partial \rho u_i}{\partial x_i}$$
$$\frac{\partial \rho u_i}{\partial t} = -\frac{\partial \rho u_i u_j}{\partial x_j} - \frac{\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j}$$
$$\frac{\partial \rho e_t}{\partial t} = -\frac{\partial \rho e_t u_j}{\partial x_j} - \frac{\partial P u_j}{\partial x_j} + \frac{\partial (\tau_{ij} \cdot u_i)}{\partial x_j} - \frac{\partial q_j}{\partial x_j}$$
$$\frac{\partial \rho Y_k}{\partial t} = -\frac{\partial \rho Y_k u_j}{\partial x_j} - \frac{\partial J_{k,j}}{\partial x_j} + \omega_k$$

Compressible Navier-Stokes Equations



Temperature after 2e-3 seconds

[1] Bhagatwala, Chen, Lu, Direct numerical simulations of HCCI/SACI with ethanol. *Combustion and Flame*, 161(7):826–1841, 2014.
[2] Chen et al., Terascale direct numerical simulations of turbulent combustion using S3D, *Computational Science & Discovery*, 2(1):015001, 2009.

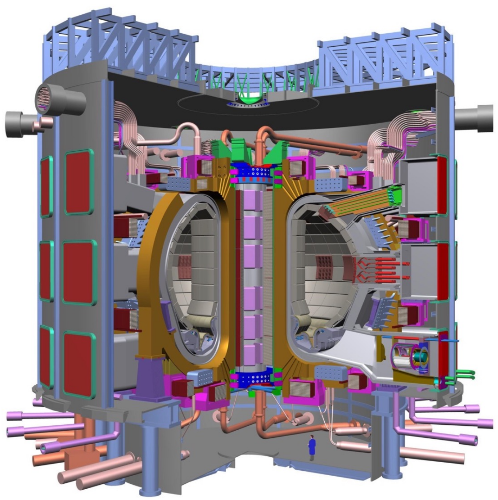
TOKAMAK FUSION REACTOR DESIGN AND DISRUPTION MITIGATION

- Develop and evaluate **plasma physics** models and scalable solution methods to understand disruption physics and explore mitigation strategies to avoid damage to tokamak fusion reactors. [1]
- 2D simulation data in Drekar, a finite element code for magnetohydrodynamics (MHD) [2]
  - 101 x 51 spatial grid
  - 13 variables: magnetic field (3), density, pressure, velocity (3), momentum (3), temperature, constraint Lagrange multiplier
  - 410 time steps
  - Data: 101 x 51 x 13 x 410 (x x y x $variable$ x $time$) tensor

Problem: Compress data while capturing plasma physics dynamics

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0,$$
$$\frac{\partial (\rho \mathbf{u})}{\partial t} + \nabla \cdot [(\rho \mathbf{u} \otimes \mathbf{u}) + pI + \boldsymbol{\pi}] - \mathbf{j} \times \mathbf{B} = \mathbf{0},$$
$$\frac{n}{\gamma - 1}\frac{\partial T}{\partial t} + \frac{n}{\gamma - 1}\mathbf{u} \cdot \nabla T + p(\nabla \cdot \mathbf{u}) + \nabla \cdot \mathbf{q} - \eta \|\mathbf{j}\|^2 - \boldsymbol{\pi} : \nabla \mathbf{u} = 0,$$
$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \cdot \left[\mathbf{u} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{u} - \frac{\eta}{\mu_0}(\nabla \mathbf{B} - (\nabla \mathbf{B})^T)\right] = \mathbf{0},$$
$$\nabla \cdot \mathbf{B} = 0,$$

Compressible Visco-resistive MHD Equations



International Thermonuclear Experimental Reactor (ITER) [under construction, France]

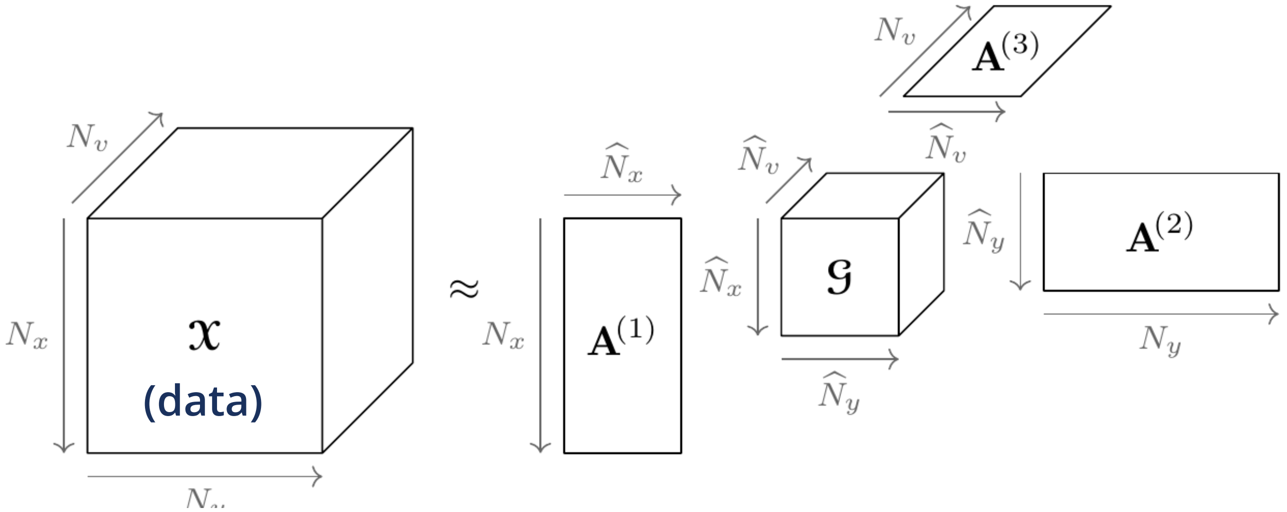
[1] U.S. Department of Energy, Tokamak Disruption Simulation (TDS) SciDAC Center, https://tds-scidac.github.io/.
[2] Shadid, et al., Scalable Implicit Incompressible Resistive MHD with Stabilized FE and Fully-coupled Newton-Krylov-AMG, Comput. Methods Appl. Mech. Engrg. 304, 1-25, 2016.

EXISTING METHOD: LOW-RANK TENSOR DATA COMPRESSION

Sequentially-Truncated Higher-Order Singular Value Decomposition (ST-HOSVD):
Low-rank Tucker tensor model with specified bound (ϵ) on relative root mean squared error (i.e., model error) [1,2]:

$$\min_{\mathcal{G},\{\mathbf{A}^{(k)}\}} \widehat{N}_x \times \widehat{N}_y \times \widehat{N}_v \times \widehat{N}_t$$

$$\text{subject to} \quad \underbrace{\frac{\|\mathcal{X} - \widehat{\mathcal{X}}\|}{\|\mathcal{X}\|}}_{\text{model error}} \leq \epsilon$$



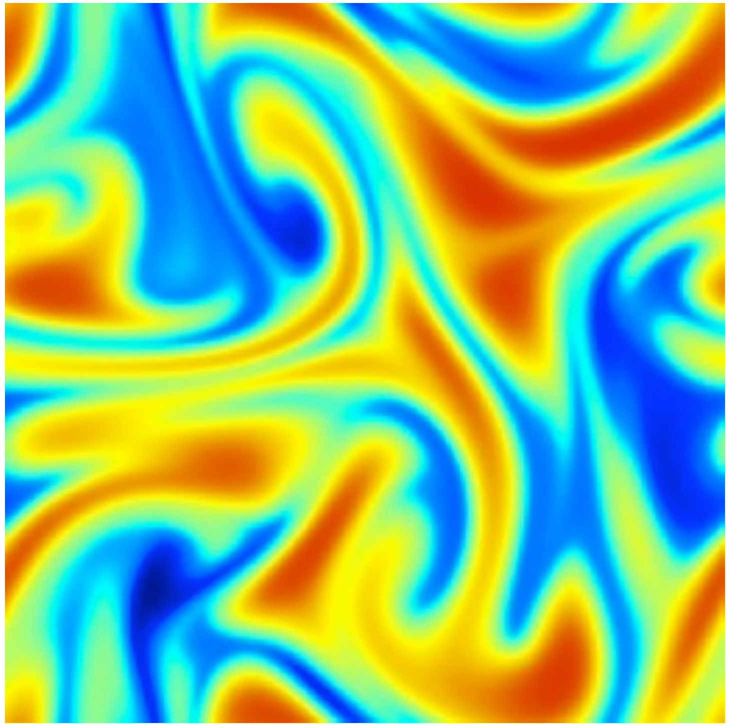
$$\widehat{\mathcal{X}}(x,y,v,t) = \sum_{i_x=1}^{\widehat{N}_x} \sum_{i_y=1}^{\widehat{N}_y} \sum_{i_v=1}^{\widehat{N}_v} \sum_{i_t=1}^{\widehat{N}_t} \underbrace{\mathcal{G}(i_x,i_y,i_v,i_t)}_{\text{core tensor}} \underbrace{\mathbf{A}^{(1)}(x,i_x)\mathbf{A}^{(2)}(y,i_y)\mathbf{A}^{(3)}(v,i_v)\mathbf{A}^{(4)}(t,i_t)}_{\text{factor matrices}}$$

Alternatively, we can minimize model error given a fixed core tensor size

[1] Kolla, et al., Higher Order Tensors for DNS Data Analysis and Compression. In *Data Analysis for Direct Numerical Simulations of Turbulent Combustion*. Springer, 2020.
[2] Vannieuwenhoven, Vandebril, Meerbergen, A new truncation strategy for the higher-order singular value decomposition. *SISC*, 34(2), 2012.

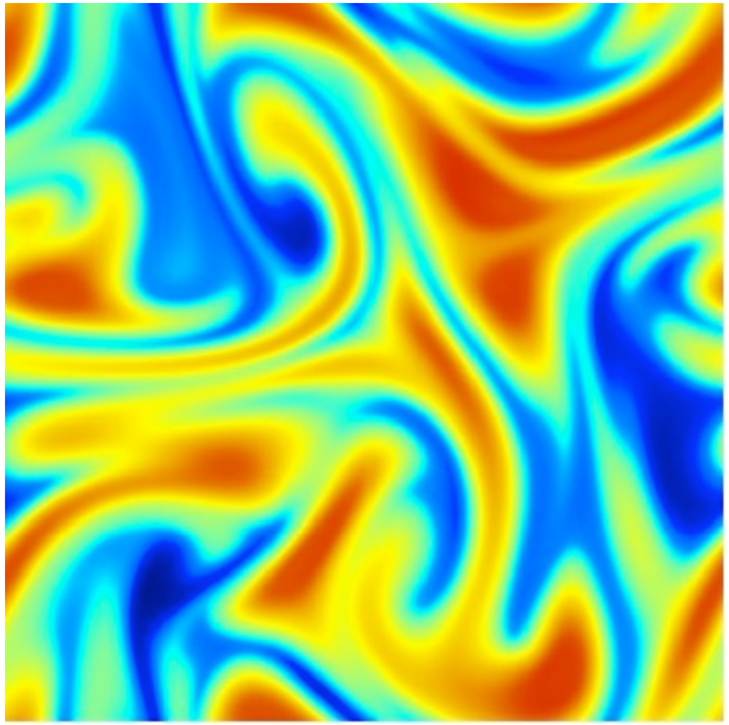
LOW-RANK TENSOR DATA COMPRESSION: COMBUSTION EXAMPLE

$\mathcal{X}$: 672 x 672 x 32 x 50



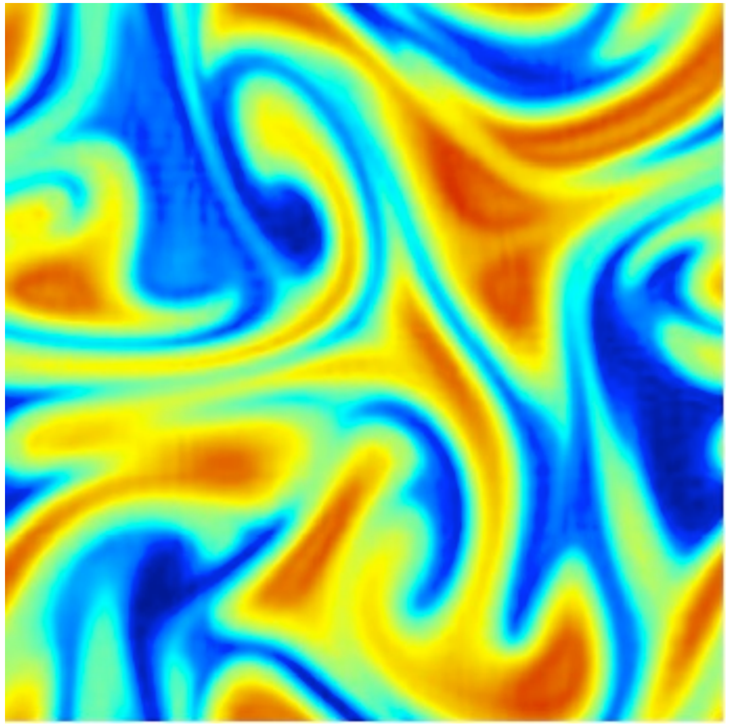
Original Combustion Simulation Data

$\mathcal{G}$: 134 x 130 x 19 x 10



Compression: ~213X
Model Error: ϵ = 1e-2

$\mathcal{G}$: 46 x 40 x 7 x 3



Compression: ~7700X
Model Error: ϵ = 1e-1

Temperature after 2e-3 seconds

QUANTITIES OF INTEREST: COMBUSTION EXAMPLE

Mass (linear)

$$\mathcal{M}_{\mathcal{X}}(x,y,t) = \sum_{v=1}^{28} \mathcal{X}(x,y,v,t)$$

$$G_{1,t}(\mathcal{X}) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \mathcal{M}_{\mathcal{X}}(x,y,t)$$

Kinetic Energy (nonlinear)

$$\mathcal{V}_{\mathcal{X}}^{(x)}(x,y,t) = \mathcal{X}(x,y,31,t) \qquad \mathcal{V}_{\mathcal{X}}^{(y)}(x,y,t) = \mathcal{X}(x,y,32,t)$$

$$\mathcal{K}_{\mathcal{X}}(x,y,t) = \mathcal{M}_{\mathcal{X}}(x,y,t) \left[\left(\mathcal{V}_{\mathcal{X}}^{(x)}(x,y,t) \right)^2 + \left(\mathcal{V}_{\mathcal{X}}^{(y)}(x,y,t) \right)^2 \right]$$

$$G_{2,t}(\mathcal{X}) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \mathcal{K}_{\mathcal{X}}(x,y,t)$$

Goal: Preserve quantities of interest (QoIs) at each time step between the simulation data and low-rank Tucker tensor model data

NEW METHOD: GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION

$$\min_{\widehat{\mathcal{X}}} \quad \underbrace{\alpha_0 \frac{\|\mathcal{X} - \widehat{\mathcal{X}}\|}{\|\mathcal{X}\|}}_{\text{model error}} + \sum_{q=1}^{N_q} \sum_{t=1}^{N_t} \alpha_{q,t} \underbrace{\left(\frac{G_{q,t}(\mathcal{X}) - G_{q,t}(\widehat{\mathcal{X}})}{G_{q,t}(\mathcal{X})} \right)^2}_{\text{Quantity of Interest (QoI) errors}}$$

- **Tucker model:** $\widehat{\mathcal{X}}(x,y,v,t) = \sum_{i_x=1}^{\widehat{N}_x} \sum_{i_y=1}^{\widehat{N}_y} \sum_{i_v=1}^{\widehat{N}_v} \sum_{i_t=1}^{\widehat{N}_t} \mathcal{G}(i_x,i_y,i_v,i_t)\mathbf{A}^{(1)}(x,i_x)\mathbf{A}^{(2)}(y,i_y)\mathbf{A}^{(3)}(v,i_v)\mathbf{A}^{(4)}(t,i_t)$

- **Low-rank tensor modeling framework:** Tensor Toolbox for MATLAB [1]
- **Loss function:** model error + sums of N_q QoI per-time constraints
- **Derivatives:** MATLAB Deep Learning Toolbox (`dlfeval/dlgradient`) [2]
- **Minimization:** limited-memory quasi-Newton (L-BFGS-B) [3,4]

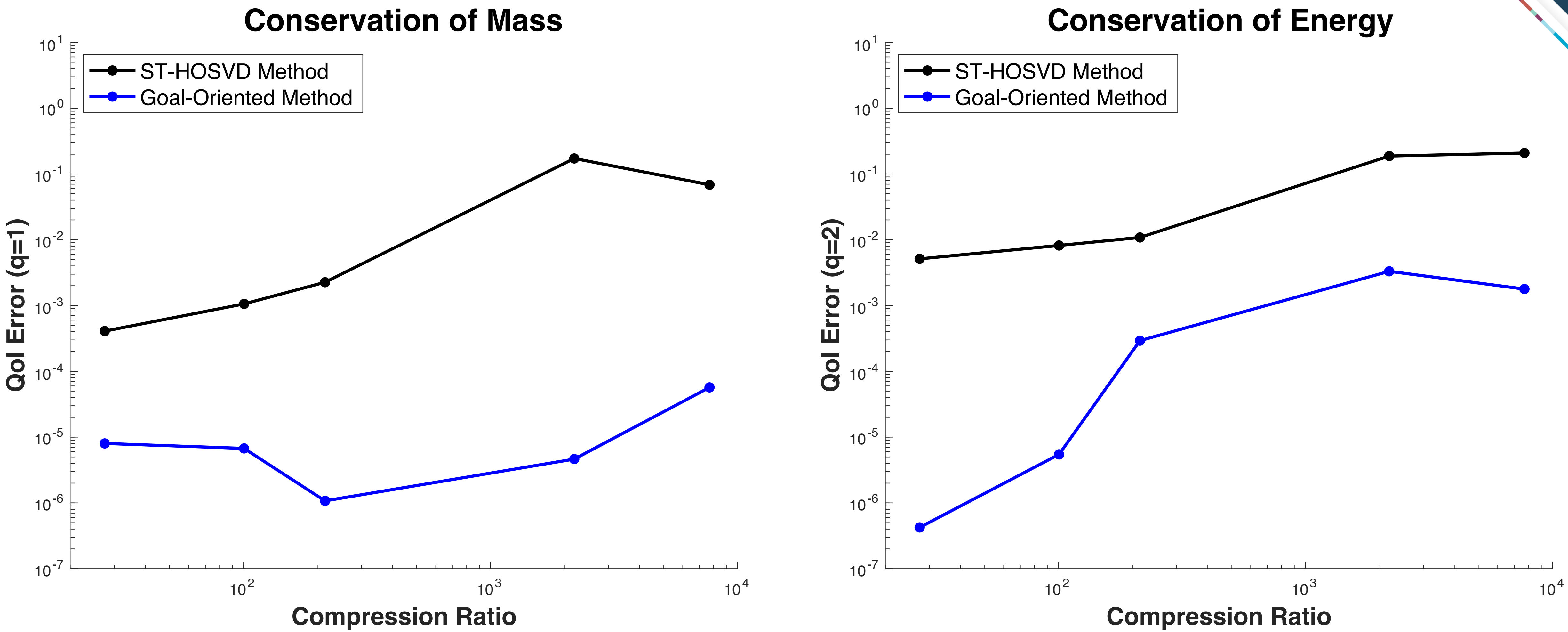
[1] Bader, Kolda, Dunlavy and others, Tensor Toolbox for MATLAB, Version 3.4, http://www.tensortoolbox.org, September 21, 2022.
[2] MathWorks, https://www.mathworks.com/help/deeplearning/ug/deep-learning-with-automatic-differentiation-in-matlab.html, accessed February 14, 2023.
[3] R. H. Byrd, P. Lu and J. Nocedal. A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific and Statistical Computing, 16(5):1190-1208, 1995.
[4] Becker, https://github.com/stephenbeckr/L-BFGS-B-C, accessed February 14, 2023.

GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION: EXPERIMENTS

$$\min_{\widehat{\mathcal{X}}} \quad \underbrace{\alpha_0 \frac{\|\mathcal{X} - \widehat{\mathcal{X}}\|}{\|\mathcal{X}\|}}_{\text{model error}} + \sum_{q=1}^{N_q} \sum_{t=1}^{N_t} \alpha_{q,t} \underbrace{\left(\frac{G_{q,t}(\mathcal{X}) - G_{q,t}(\widehat{\mathcal{X}})}{G_{q,t}(\mathcal{X})} \right)^2}_{\text{Quantity of Interest (QoI) errors}}$$

- **Initialization:** ST-HOSVD solution (uses model error only)
- **Model errors:** ϵ={1e-3, 5e-3, 1e-2, 5e-2, 1e-1}
- **Weights:** $\alpha_{q,t}$ = 1.0 (q={1,2}; t={1..50})

GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION: RESULTS



Goal-oriented tensor data compression significantly reduces QoI error

GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION: RESULTS

Model Error: $$\frac{\|\mathcal{X} - \hat{\mathcal{X}}\|}{\|\mathcal{X}\|}$$

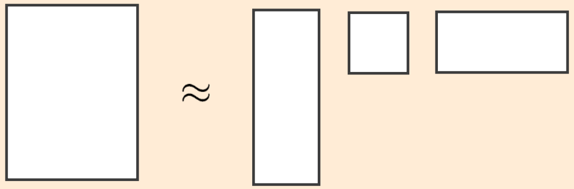
Compression Ratio	Model Core Tensor Size	ST-HOSVD	Goal-Oriented (Mass)	Goal-Oriented (Energy)
2.74E+01	233 x 229 x 28 x 18	9.32E-04	9.32E-04	9.33E-04
1.00E+02	163 x 160 x 23 x 12	4.57E-03	4.57E-03	4.57E-03
2.14E+02	134 x 130 x 19 x 10	9.05E-03	9.06E-03	9.06E-03
2.18E+03	70 x 65 x 11 x 5	4.47E-02	4.50E-02	4.48E-02
7.70E+03	46 x 40 x 7 x 3	9.35E-02	9.39E-02	9.39E-02

Goal-oriented tensor data compression leads to negligible model error

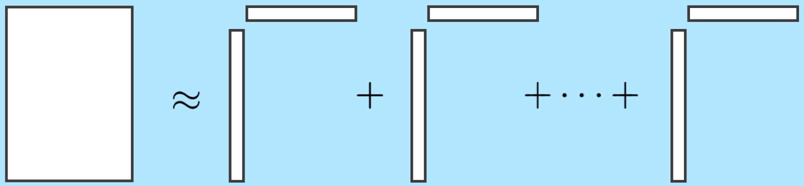
LOW-RANK DECOMPOSITIONS: TWO POINTS OF VIEW

Matrix Decompositions

Viewpoint 1: High-variance subspaces, useful for compression



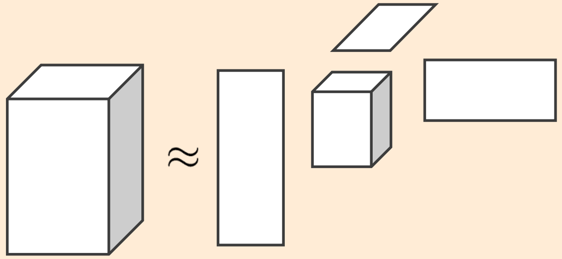
Viewpoint 2: Sum of vector outer products, useful for interpretation



Singular value decomposition (SVD), eigendecomposition (EVD), nonnegative matrix factorization (NMF), etc.

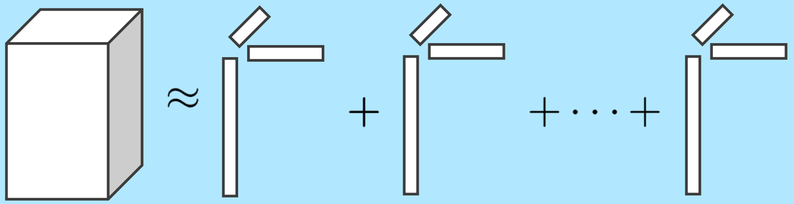
Tensor Decompositions

Tucker Model: Project onto high-variance subspaces to reduce dimensionality



HOSVD, Best Rank-($R_1, R_2, \ldots, R_d$) decomposition

CP Model: Sum of d-way vector outer products, useful for interpretation



Canonical Polyadic, CANDECOMP, PARAFAC, CP

Other models for compression include hierarchical Tucker, tensor train, tensor ring, tensor network, etc.

Kolda and Bader (2009), Tensor Decompositions and Applications, https://doi.org/10.1137/07070111X

QUANTITIES OF INTEREST: PLASMA PHYSICS EXAMPLE

Momentum	Divergence of Magnetic Field (B)

$$G_{1,t}(\mathcal{X}) = \left(\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \sum_{v=7}^{9} \mathcal{X}(x,y,v,t)^2 \right)^{\frac{1}{2}}$$

$$G_{2,t}(\mathcal{X}) = \left(\iint_D (\nabla \cdot \mathbf{B}(x,y,t))^2 \, dx dy \right)^{\frac{1}{2}}$$

Goal: Preserve quantities of interest (QoIs) at each time step between the simulation data and low-rank CP tensor model data

GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION (CP MODEL)

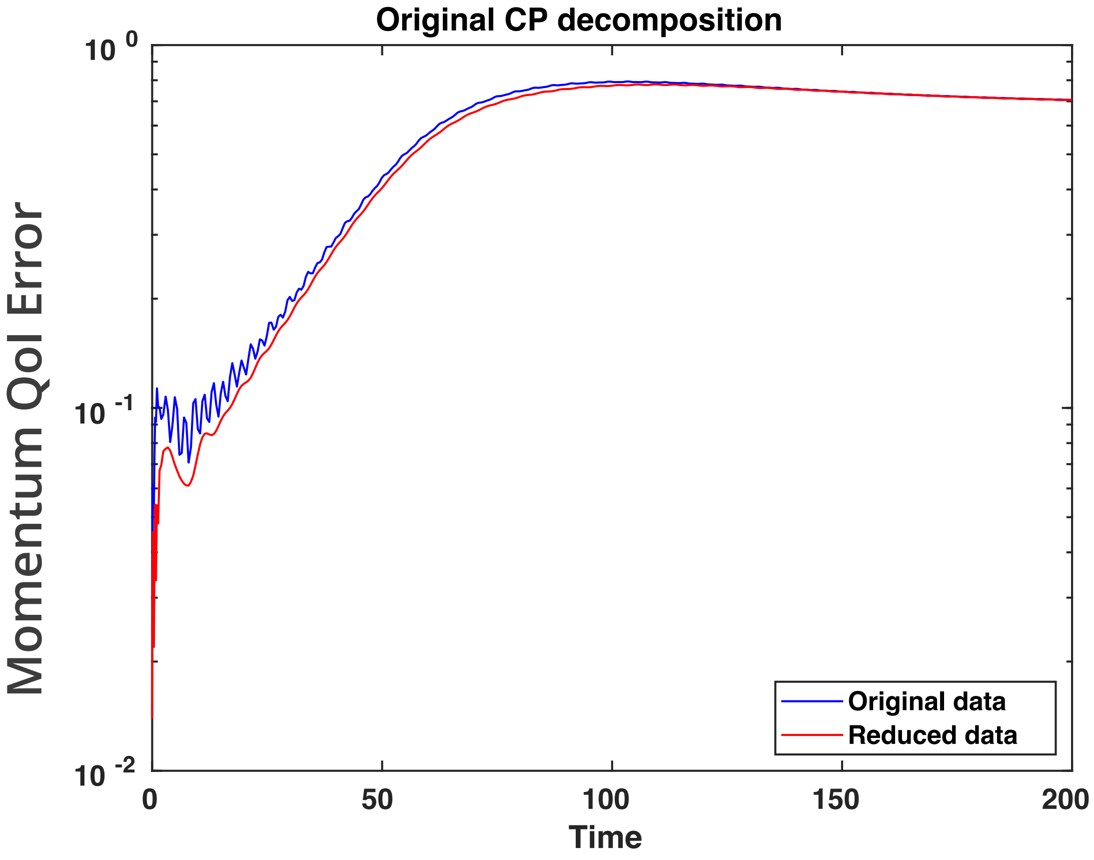
$$\min_{\widehat{\mathcal{X}}} \quad \underbrace{\alpha_0 \frac{\|\mathcal{X} - \widehat{\mathcal{X}}\|}{\|\mathcal{X}\|}}_{\text{model error}} + \sum_{q=1}^{N_q} \sum_{t=1}^{N_t} \alpha_{q,t} \underbrace{\left(\frac{G_{q,t}(\mathcal{X}) - G_{q,t}(\widehat{\mathcal{X}})}{G_{q,t}(\mathcal{X})} \right)^2}_{\text{Quantity of Interest (QoI) errors}}$$

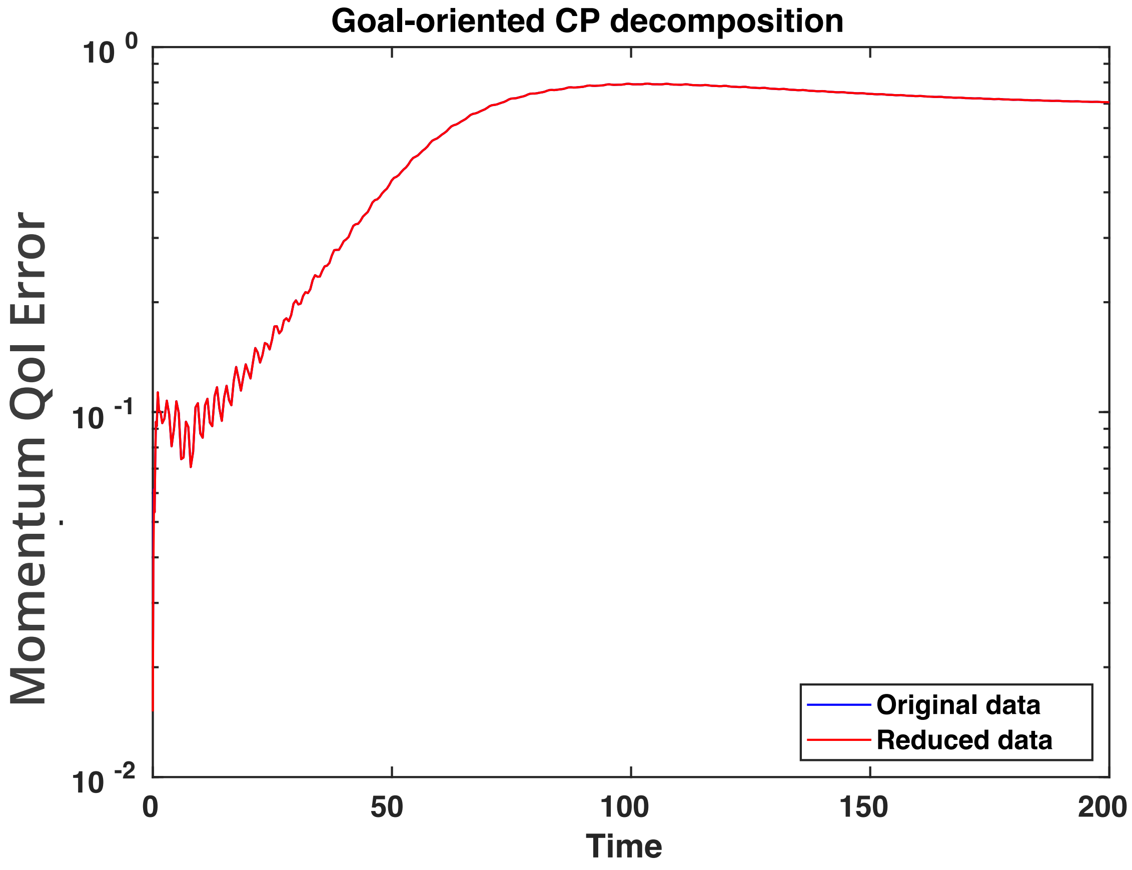
- **Rank-*R* CP Model:** $\widehat{\mathcal{X}}(x, y, v, t) = \sum_{r=1}^{R} \mathbf{A}^{(1)}(x, r)\mathbf{A}^{(2)}(y, r)\mathbf{A}^{(3)}(v, r)\mathbf{A}^{(4)}(t, r)$

- **Low-rank tensor modeling framework:** Tensor Toolbox for MATLAB
- **Loss function:** model error + sums of N_q QoI per-time constraints
- **Derivatives:** MATLAB Deep Learning Toolbox (`dlfeval`/`dlgradient`)
- **Minimization:** generalized CP (GCP) using L-BFGS-B [1] and ADAM [2]
- **Weights:** manual tuned for improved model fitting

[1] D. Hong, T. G. Kolda, and J. A. Duersch. Generalized Canonical Polyadic Tensor Decomposition. *SIAM Review*, 62(1):133–163, January 2020.
[2] T. G. Kolda and D. Hong, Stochastic Gradients for Large-Scale Tensor Decomposition, *SIAM Journal on Mathematics of Data Science*, 2 (2020), pp. 1066–1095.

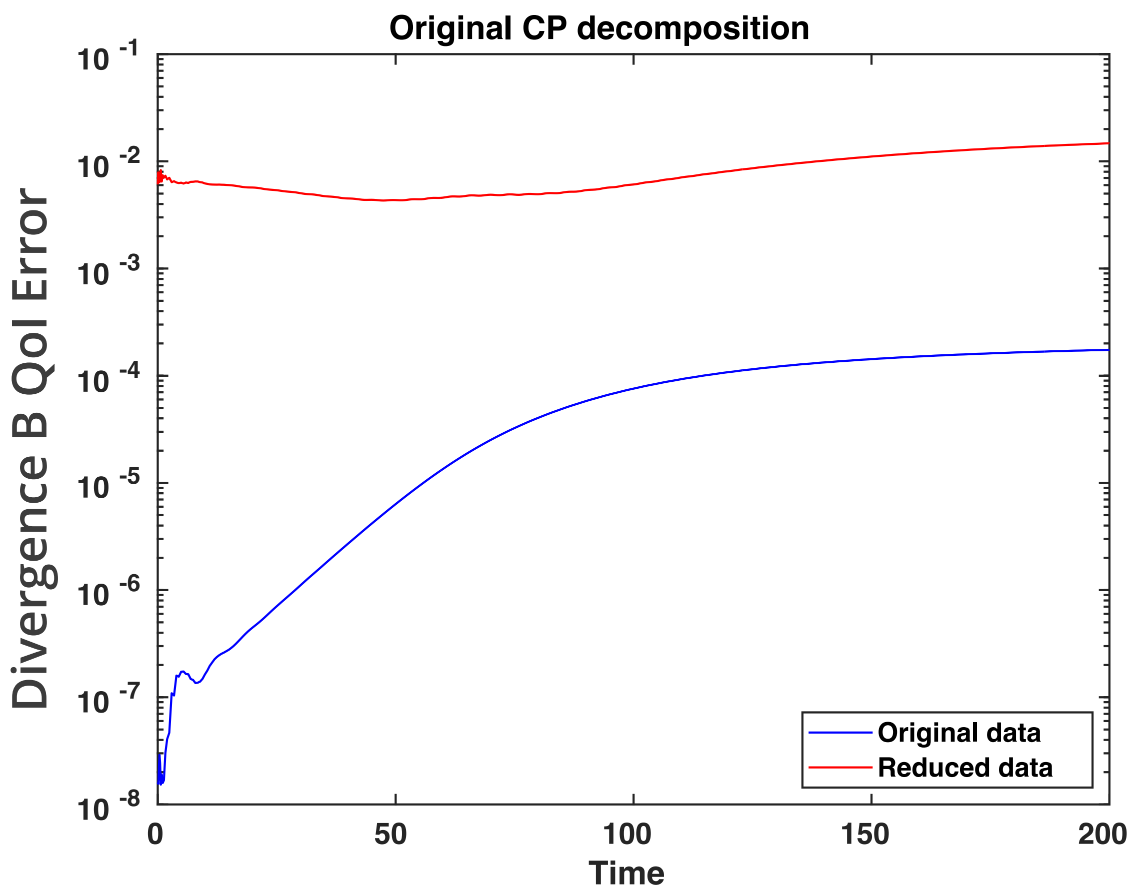
GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION: RESULTS

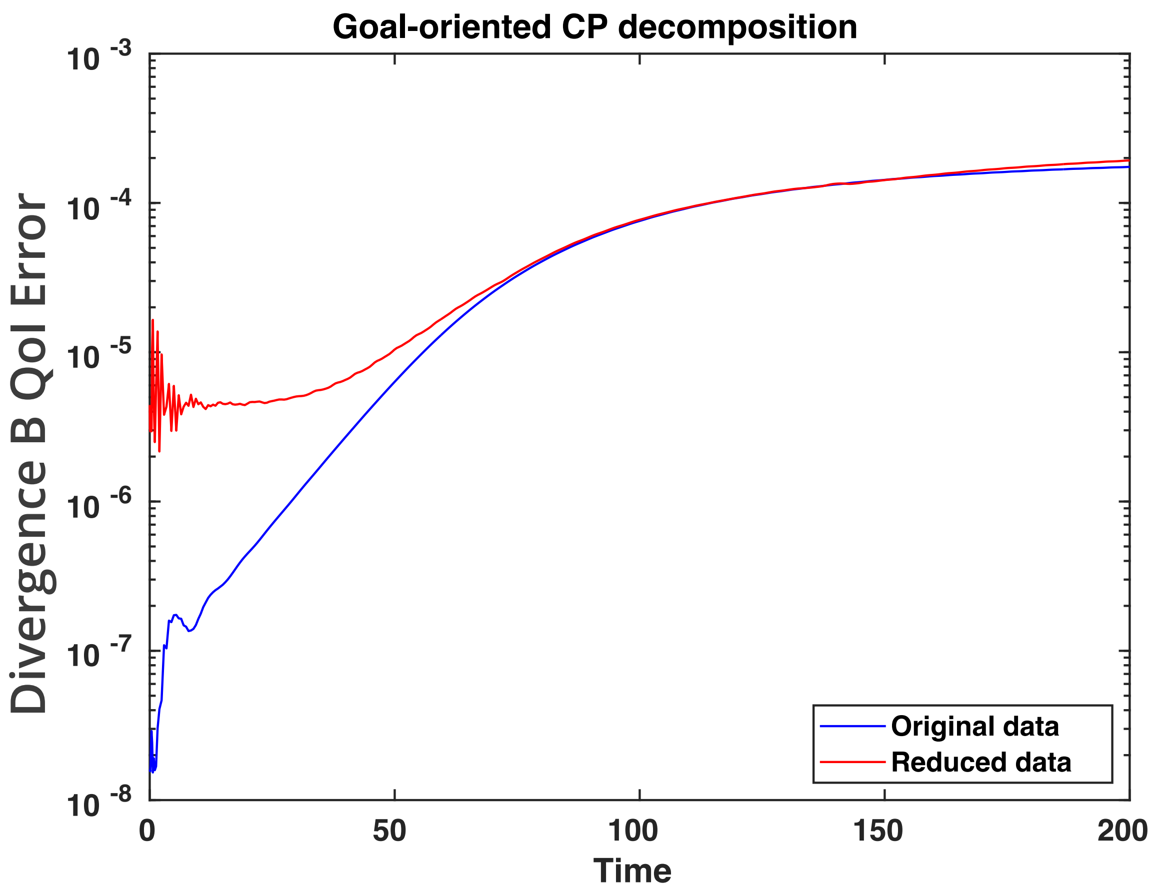




Goal-oriented tensor data compression significantly reduces QoI error

GOAL-ORIENTED LOW-RANK TENSOR DATA COMPRESSION: RESULTS





Much improved QoI error with negligible increase in model error

SUMMARY: GOAL-ORIENTED TENSOR DECOMPOSITIONS

- **Conclusions**
  - Goal-oriented low-rank tensor decompositions reduce errors for quantities of interest without much increase in model error
  - L-BFGS-B minimization can be slow for current goal-oriented model fitting formulation; initializing with fast methods for fitting model error alone is critical
  - Weights for model error and goal terms in loss functions may require manual tuning

- **Next Steps**
  - Comparison of minimization methods (ADMM, interior point method, etc.)
  - Scalable goal-oriented tensor modeling leveraging stochastic methods (SGD, ADAM)
  - *In-situ* goal-oriented tensor modeling to provide adaptive compression during simulations
  - More challenging application problems (3D, more complex/realistic physics, etc.)

Goal-Oriented Tensor Modeling

QUESTIONS?

Contact:

Danny Dunlavy

dmdunla@sandia.gov