

# SESSION 6 BREAKOUT REPORT – APPLICATIONS

---

Participants: Vishwas Rao, Chris Geoga, Qin Li, Daniel Bielich, Kevin Miller, Rob Webber, Shambhavi Suryanarayanan, Kate Pearce, D. Adrian Maldonado, Riley Murray  
Facilitator: Yifu Wang

# Data Assimilation

- Large-scale electronic structure simulation
- Hartree–Fock (HF) and post-HF methods for molecules & materials
- Massive Hermitian eigenvalue problems from discretized electronic Hamiltonians

# Data Assimilation

- Large-scale electronic structure simulation
- Hartree–Fock (HF) and post-HF methods for molecules & materials
- Massive Hermitian eigenvalue problems from discretized electronic Hamiltonians

~> Cost grows rapidly when amount of orbitals and electrons increase.

~> Deterministic sparsification can introduce systematic bias and degrade spectral accuracy

# Data Assimilation

- Large-scale electronic structure simulation
- Hartree–Fock (HF) and post-HF methods for molecules & materials
- Massive Hermitian eigenvalue problems from discretized electronic Hamiltonians

~> Cost grows rapidly when amount of orbitals and electrons increase.

~> Deterministic sparsification can introduce systematic bias and degrade spectral accuracy

Potential randomized tools:

- Unbiased sampling
- Randomized sparsification
- Randomized subspace iteration

## Gaussian and structured kernel methods

- massive covariance/kernel matrices (lattice structures, DFT-related operators etc.)
- specific structure for the kernel matrix
- matrix entries themselves inaccessible; often interact with the operator through matrix-vector products

## Gaussian and structured kernel methods

- massive covariance/kernel matrices (lattice structures, DFT-related operators etc.)
- specific structure for the kernel matrix
- matrix entries themselves inaccessible; often interact with the operator through matrix-vector products

Randomized sketches using Gaussian test vectors, peeling algorithms, and hierarchical low-rank compression can efficiently recover the dominant structure of covariance operators using only matrix-vector interactions.

- near-linear complexity
- better scaling

## PDE-constrained Inverse Problems

- first turn into linearized problem
- linear systems arising from dis-cretized PDEs
- matrix usually being very tall and structured

# Large-Scale Eigensolvers

- Hard to parallelize (sequential norm updating)
- Memory constraint

# Large-Scale Eigensolvers

- Hard to parallelize (sequential norm updating)
- Memory constraint

## **Barriers to adoption – the Verification Challenge**

Engineering applications often require bitwise reproducibility across runs for certification, debugging, and validation.

- How to reproduce the same randomness?

# Tensor Sketching/Decomposition

- In heavy-tailed or noisy settings, ideas from robust statistics are needed because outliers can destroy sketch quality.

~> Robustness to heavy-tailed data

- Deciding when an existing sketch is still reliable and when the system has changed enough to require resketching.

## Active Learning / Adaptive Data Selection for Machine Learning

- remote sensing
- synthetic aperture radar
- Judiciously select "more useful" data (leverage score etc.)

# Open Problems

## Benchmark for RandNLA

- Need to reflect the performance (flops, I/O)
- cannot be cheated (grading the BLAS)
- Independent benchmarking beyond vendor-developed software
- Encourage community submission of real-world matrices

## Extend RandNLA to

- matrix completion
- low-rank approximation/truncation

COMMENTS WELCOME!