Exascale Algorithms for Large-Scale Solvers and Uncertainty Quantification

George Biros, Roger Ghanem, Omar Ghattas

The University of Texas at Austin

January 15, 2013
“Moore’s Law” for MHD simulations

A brief history of parallel algorithms research by our team

As measured in the “SC’XY awards norm”

Team members have published 18 SC’XY papers on scalable parallel algorithms since late 90s, receiving a number of SC honors:

- SC02: Inexact Newton-Krylov for inverse problems (Best Paper Award)
- SC03: Inverse problems in wave propagation (Gordon Bell Prize)
- SC03: Kernel independent fast multipole method (Gordon Bell Finalist, Best Paper Finalist, Best Student Paper Award)
- SC06: Integrated simulation and visualization (Best Student Paper Finalist)
- SC06: Online meshing, simulation, and visualization (HPC Analytics Award)
- SC07: Non-uniform multigrid (Best Paper Finalist)
- SC08: AMR on octrees (Gordon Bell Finalist)
- SC09: KIFMM for heterogeneous systems (Best Student Paper Finalist)
- SC09: High-order AMR on complex geometry (Best Poster Award)
- SC10: High-order AMR on complex geometry (Gordon Bell Finalist)
- SC10: Fast multipole for complex fluids (Gordon Bell Prize)
- SC12: UQ for inverse problems (Gordon Bell Finalist)
A hybrid geometric-algebraic multigrid method

- Multigrid is the gold standard for linear solvers
- AMG: ideal for unstructured meshes, but difficulty scaling to extreme core counts (e.g., ML from Trilinos, BoomerAMG from hypre)
- GMG: demonstrated good scaling to \( O(10^5) \) cores, but challenges for unstructured meshes
- Hybrid AMG-GMG:
  - Hexahedral coarse mesh to resolve geometry
  - GMG using forest of octrees adaptivity to define finer meshes and prolongation & restriction operators
  - AMG as the coarse mesh solver
  - Weak-scales to 262K cores with 71% efficiency

Weak scalability of hybrid geometric-algebraic multigrid based on forest-of-octrees adaptivity

<table>
<thead>
<tr>
<th></th>
<th>64</th>
<th>512</th>
<th>4096</th>
<th>32,768</th>
<th>262,144</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup</td>
<td>2.97</td>
<td>2.64</td>
<td>3.1</td>
<td>3.76</td>
<td>8.6</td>
</tr>
<tr>
<td>Smoother</td>
<td>289.7</td>
<td>301.5</td>
<td>336.3</td>
<td>391.3</td>
<td>409.1</td>
</tr>
<tr>
<td>Transfer</td>
<td>7.45</td>
<td>8.47</td>
<td>11.5</td>
<td>11.35</td>
<td>15.88</td>
</tr>
<tr>
<td>Coarse Setup</td>
<td>1.85</td>
<td>2.13</td>
<td>0.82</td>
<td>1.27</td>
<td>1.63</td>
</tr>
<tr>
<td>Coarse Solve</td>
<td>24.3</td>
<td>30.8</td>
<td>18.47</td>
<td>30.1</td>
<td>26.01</td>
</tr>
<tr>
<td>Total Time</td>
<td>326.3</td>
<td>345.5</td>
<td>370.2</td>
<td>437.8</td>
<td>461.2</td>
</tr>
</tbody>
</table>

- **weak scaling** of Poisson solve with 400K elements per core (largest problem = 100 billion DOF)
- **45K octrees coarse mesh**
- **4 pre- and post-smoothing steps**
- **ML AMG solver** (from Trilinos) used as **coarse grid solver**
- **71% parallel efficiency** for 4000\(\times\) increase in problem size & core count from 64 to 262,144 cores
Research issues for exascale multigrid solvers

- Extending our method to \textit{anisotropic} and \textit{rough} operators
- Extending our method to \textit{high-order discretizations}
- Fault tolerance
- \textit{Performance} tuning, particularly for heterogeneous architectures
- \textit{deployment} within \textit{rvdDNS} and \textit{GRINS}
- Alternative to Newton-MG: \textit{Nonlinear multigrid}
**p4est: Parallel forest-of-octrees AMR library**


Weak scalability of *p4est*-only operations on full Jaguar

Excellent scalability of pure AMR operations over 18,360X range of core count

Left: Runtime dominated by Balance and Nodes while Partition and Ghost take less than 10% (*New* and *Refine* are negligible and not shown).

Right: **Weak scaling** for 2.3 million elements/core; ideal scaling would result in bars of constant height. Largest mesh created contains over **513 billion elements** and is balanced in **21 s**.

Scalable methods for polynomial chaos expansions

- Conventional PCEs suffer from the curse of dimensionality
- Stems from attempting to approximate entire spatio-temporal field of solution
- Yet QoIs are generally low-dimensional functionals of solution
- Coordinate rotation of stochastic space results in most of the probabilistic content of QoI being concentrated about a single dimension
- Work to compute transformation scales linearly in parameter space dimension
- These ideas will be developed and applied to target combustion problem

Bayesian framework for inverse problems: Quest for knowledge from data and models

Uncertainty is a fundamental feature of ill-posed inverse problems:

- **Deterministic** approach to ill-posedness: employ regularization to penalize unwanted solution features, guarantee unique solution
- **Bayesian** approach to ill-posedness: describe probability of all parameters that are consistent with the data, the model, and any prior knowledge of the parameters
- Unfortunately, solution of Bayesian inverse problems via MCMC (method of choice) is intractable for high dimensional parameter spaces and expensive forward models!
Stochastic Newton MCMC sampling

Goal: exploit problem structure in form of Hessian of parameter-to-observable map

Sample posterior probability density

\[
\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2} \left\| f(m) - d_{\text{obs}} \right\|_{\Gamma_{\text{noise}}}^{-1} - \frac{1}{2} \left\| m - m_{\text{pr}} \right\|_{\Gamma_{\text{pr}}}^{-1}\right)
\]

MCMC: propose from distribution \(q(m_k, \cdot)\); accept with probability

\[
\alpha = \min\left(1, \frac{\pi(y)q(y, m_k)}{\pi(m_k)q(m_k, y)}\right)
\]

Convergence comparison: Stochastic Newton vs. DRAM

Random walk proposal: isotropic Gaussian

Stochastic Newton proposal: local Hessian-tailored Gaussian
Million-dimensional example

- 1.07 million uncertain acoustic wave speed parameters
- 630 million state variables, 2400 time steps
- Up to 100K cores on Jaguar XK6 (single forward solve is 1 minute on 64K cores)
- $2000 \times$ reduction in problem dimension (488 dominant eigenvectors)
- Top row: Samples from prior
- Bottom row: Samples from the posterior
- Right: “true” earth model (black dots=5 sources, white dots=100 receivers)


Research challenges for intrusive MCMC sampling

- Scalable prior operators
- Reuse of Hessian information to improve Gaussian proposals
- Devise problem-specific Hessian approximations when even low rank approximation is too expensive
- Develop trust region methods to enhance robustness of stochastic Newton for strongly non-Gaussian distributions
- All of the above in extreme-scale setting
Synergistic projects

- **QUEST**: Quantification of Uncertainty in Extreme Scale Computations, DOE ASCR SciDAC Institutes program, 2011–2016. (SNL, LANL, Duke, MIT, USC, UT Austin)
- **DiaMonD**: An Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions, DOE ASCR MMICCs program, 2012–2017. (UT Austin, MIT, FSU, CSU, Stanford, ORNL, LANL)
Extra slides