DDDAS Workshop:
WG2 - Mathematical and Statistical Algorithms

Craig C. Douglas, George Biros,
and many friends

More Funding, Less Filing

Charge++

- What is the state-of-the-art and what are the challenges in the applications algorithms to enable such capabilities?
- What advances are needed to enable application algorithms that are tolerant to perturbations from “on-line” input data and have stability properties?
- How can one select and incorporate dynamically appropriate algorithms as the application requirements and data sets change in the course of simulation?
- What kinds of approaches, such as knowledge-based systems, can be employed, and what interfaces and applications assists are needed to enable such capabilities?
- What systems support is required to develop such environments?
- Other topics…

Important algorithm classes and issues

- Dynamic model changing, Model reduction, Scale up, Multiscale methods
- Inexpensive (nonlinear) update methods that can be corrected easily
- Dynamic error analysis for calculations and data assimilation
- Inverse problems to dynamically update parameters
- Bayesian methods, Kalman filtering
- Ensembles
- Dynamic optimization
- Network assessment, quality of service
- Artificial intelligence
- Tight, power inexpensive imbedded algorithms in sensors
- Data discovery: quality, structure, hidden information, anomalies
- Mathematical proofs: local versus global in time

What is the state-of-the-art and what are the challenges in the applications' algorithms to enable such capabilities?

- State of the art: algorithms
  - Auto industry: ABS, climate control, wipers, lights
  - Chemical plants
  - Autopilot of airplane
  - 4DVAR (next is 4DVAR+Ensemble)
  - Activation mechanism for earthquakes and tall buildings
  - Epidemic behavior or individual based social behavior
  - Data mining or online learning
- State of the art: methodology
  - Ensemble Kalman filter

What is the state-of-the-art and what are the challenges in the applications' algorithms to enable such capabilities?

- Challenges
  - Convergence of algorithms that run slower than data delivery
  - Adding physical correctness to the models dynamically
  - System identification (force verification of all math algorithms in the process) including model changing and calibration
  - Reduced scale models
  - Can static methods be extended to dynamic ones (data mining)
  - Unified way to treat hard and soft data conditions
  - Sensor steering
- Grand challenges
  - Nonlinearity
  - Dimensionality (huuuge)
  - Uncertainty (initial and parameter states)
  - Chaotic problems

What advances are needed to enable application algorithms that are tolerant to perturbations from “on-line” input data and have stability properties?

- Errors
  - Dynamic validation versus verification
  - Better dynamic error estimators
  - Sensor error catching and treating
  - Model errors
  - Minimize error subject to uncertain inputs
- Mathematically and statistically consistent theory for boundary and initial conditions
- Guaranteed convergence
- More supercomputing cycles provided by the NSF + FUNDING
- Visualization software to decipher results
- Determine the limits between automatic decision making versus bringing in a person to make a decision

More Funding, Less Filing

How can one select and incorporate dynamically appropriate algorithms as the application requirements and data sets change in the course of simulation?

- Based on the data and the application area, the sensor may force a different physical model or even application
  - Laminar flow versus turbulence
  - Gas flow around a well: Florsheimer versus Darcy
  - Contaminant identification: what to look for next
- Multi-objective decision problems
- Change of discrete method and algorithms

What kinds of approaches, such as knowledge-based systems, can be employed, and what interfaces and applications assists are needed to enable such capabilities?

- Neural networks
- Information retrieval, e.g., Google search/earth
- Decision analysis for when to bring human into the loop
- Machine learning
- Visualization on anything from a cell phone or PDA to a CAVE
- Agent based
- Empirical model

What systems support is required to develop such environments?

- Real time for
  - Managing large dataware
  - Managing processing
  - Parallel prototyping
  - Guaranteed and consistent resources, including networks
  - Dynamic scheduling of processors, networks, grids, and change of allocation
  - Process migration and checkpointing
  - On demand computing, right NOW
- Interfaces and community based software tools
- Standards for compile time source code generation
- Fault tolerance at the operating system level
- More FUNDING

Other topics

- Mathematical infrastructure of data assimilation of ensemble Kalman, Monte Carlo, or stochastic types
- Opportunity for DDDAS for verification and validation of mesoscale problems or facilities
- http://www.mgnet.org/~douglas/math-stat-algs.ppt
- Merciless plug: http://www.dddas.org
  - Papers
  - Proceedings
  - Software
  - Announcements

Ensemble estimation

- Synthetic data (observation function)
- Subset of more general stochastic methodology that can handle non-normal based distributions
- Techniques
  - Kalman filter related
  - Hybrid deterministic / stochastic
- Real-time deterministic inverse problems
  - Uncertainty estimation and propagation
- Robust dynamic optimization methods
- Secure transmission of data

Multiscales

- Pico, Nano, Micro, Meso, Macro, … scales
- Sensor data grids versus computational grids
  - Size
  - Relevance of grids
  - Data grid completion
- Reduced models
- Multiscale and sparse grid boundary values
- Artificial shocks
- Subgrid dynamics

Dynamic Data/Model-Driven Data Collection

- Data relevance to the problem
- Data noise quantification (per sample point possibly)
- Data integrity
- Data cleaning
- Targeted sampling over space and time and scale
  - Very different than for static data sources
- Data reduction
- Sensor steering
- Data stream models (predicting in advance)
- Data fusion and completion

Control

- System identification
  - Leads to a huge system that needs to be reduced
- Model reduction and reduced order controller
- Offline
  - High end computing
  - Real-time: specialized computational elements and models
- Online
  - Nonlinear closed loop, open loop control algorithms for large scale systems

Integration

- Coupling of data and models and control
  - Uncertainties
  - Estimates
  - Stability
- Efficiency of total package requires integration
  - Software
  - Algorithms
  - Data interfaces
  - Data acquisition
- Different scales lead to different techniques - how do you integrate this?
- Human versus computer decisions

Dynamic Sampling

- Actual data
  - What to save and what to ignore
  - How to save whatever is saved
  - Model reduction to the correct size based on the saved data
  - Adaptive stratification on space and time
  - Data re-broadcasting
- Model reduction to the correct size based on the saved data

Decision Making Algorithms

- Why are we doing this application?
- How do you take this knowledge and use it?
- Advising role of DDDAS
  - Medical apps
    - Patient specific health care ($$$)
  - Disaster management
  - Economic decisions
  - Optimal use of resources
  - Engineering and Scientific apps
- Education of decision makers
- Energy and environmental decisions
- Sensor steering