Overview of DAKOTA Project
(from the software perspective)

Michael S. Eldred
Optimization and Uncertainty Quantification Department (1441)

NREL
December 14, 2010

- Capability overview
- Advanced deployment efforts
DAKOTA Project

Began as optimization LDRD in 1994

Team: 5-10 core personnel in NM/CA + TPL developers

Releases: Major/Interim, Stable/VOTD; 5.1 release due 12/10

DAKOTA Training: 8 sessions (~140 students) since 5.0;
26 sessions (~500 students) total since 2001.

2009 Outreach: Minitutorials at IMAC, SIAM CS&E;
SA/UQ short courses at NASA Langley, AFRL WPAFB.

Modern SQE: Linux/Unix, Mac, Windows; Nightly builds/testing;
subversion, TRAC, Cmake; Top 2008 SQE score

GNU LGPL: free downloads worldwide
(~6500 total ext. registrations, ~3500 distributions last yr.)

Community development: open checkouts now avail (→ PSAAP)
Community support: dakota-users, dakota-developers

Iterative systems analysis
Multilevel parallel computing
Simulation management

http://dakota.sandia.gov

Users/Ref/Dev Manuals + training mats. online
Strategy: control of multiple iterators and models

Coordination:
- Nested
- Surrogate
- Recast
- Sequential/Concurrent
- Adaptive/Interactive

Parallelism:
- Asynchronous local
- Message passing
- Hybrid
- 4 nested levels with Master-slave/dynamic Peer/static

Model:
- Parameters
- Interface
- Responses

Design
- continuous
discrete range/set

Uncertain
- normal/lognorm.
uniform/logunif.
triang/exp/β/Γ
EV: I, II, III
histogram: bin/pt
discrete: p/b/nb/g/hg

State
- continuous
discrete range/set

Application
- system
fork
direct
grid

Approximation
- global
polynomial 1/2/3, NN, kriging, MARS, RBF
multipoint – TANA3
local – Taylor series
multifidelity
ROM

Strategy
- Optimization
- Uncertainty
- LeastSq

Optimization
- Hybrid
- OptUnderUnc
- UncOfOptima
- Pareto/MStart
- Branch&Bound/PICO

Uncertainty
- ModelCalUnderUnc

LeastSq
- Mixed A-E UQ

Functions
- objectives
- nonlin constraints
- least sq. terms
- generic

Gradients
- numerical
- analytic

Hessians
- numerical
- analytic
- quasi
Core Methods

**Optimization:** minimize/maximize objective(s) subject to constraints

Karush-Kuhn-Tucker conditions:
\[ \nabla f - \sum_i \lambda_i \nabla g_i = 0 \]

Achieve vector balance: objective fn grad contained within feasibility cone

**Model Calibration/Parameter Estimation:** use nonlinear least squares to minimize errors between model and data

\[ f(x) = \sum_{i=1}^{n} (s_i(x) - d_i)^2 \]

Simulation output that depends on \( x \)

Given data

**Sensitivity Analysis:** identify most influential set of parameters for key response metrics

**Uncertainty Quantification:** quantify effect of random variables on key response metrics
| Uncertainty Quantification Algorithms @ SNL: New methods bridge robustness/efficiency gap |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| **Production**                  | **New**                         | **Under dev.**                  | **Planned**                     | **Collabs.**                    |
| **Sampling**                    | Latin Hypercube, Monte Carlo    | Importance, Incremental         | Bootstrap, Jackknife            | FSU                            |
| **Stochastic expansion**        | **Adv. Deployment**             | Tailored polynomial chaos & stochastic collocation with extended basis selections | p-adaptive, adjoint gradient-enhanced | Stanford, Purdue, Austr. Natl., FSU |
| **Other probabilistic**         |                                  | Random fields/stochastic proc.  | Dimension reduction             | Cornell, Maryland               |
| **Epistemic**                   | Interval-valued/Second-order prob. (nested sampling) | Opt-based interval estimation, Dempster-Shafer | Bayesian | LANL, Applied Biometrics |
| **Metrics & Global SA**         | Importance factors, Partial correlations | Main effects, Variance-based decomposition | Stepwise regression             | UNM                            |
Generalized Polynomial Chaos Expansions

Approximate response w/ spectral proj. using orthogonal polynomial basis fns

\[ R = \sum_{j=0}^{P} \alpha_j \Psi_j(\xi) \]

i.e. using

\[ \Psi_0(\xi) = \psi_0(\xi_1) \psi_0(\xi_2) = 1 \]
\[ \Psi_1(\xi) = \psi_1(\xi_1) \psi_0(\xi_2) = \xi_1 \]
\[ \Psi_2(\xi) = \psi_0(\xi_1) \psi_1(\xi_2) = \xi_2 \]
\[ \Psi_3(\xi) = \psi_2(\xi_1) \psi_0(\xi_2) = \xi_1^2 - 1 \]
\[ \Psi_4(\xi) = \psi_1(\xi_1) \psi_1(\xi_2) = \xi_1 \xi_2 \]
\[ \Psi_5(\xi) = \psi_0(\xi_1) \psi_2(\xi_2) = \xi_2^2 - 1 \]

• Nonintrusive: estimate \( \alpha_j \) using sampling (expectation), pt collocation (regression), tensor-product quadrature, Smolyak sparse grids, or cubature (numerical integration)

Generalized PCE (Wiener-Askey + numerically-generated)

• Tailor basis: optimal basis selection leads to exponential convergence rates

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density function</th>
<th>Polynomial</th>
<th>Weight function</th>
<th>Support range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} )</td>
<td>Hermite ( H_n(x) )</td>
<td>( e^{-\frac{x^2}{2}} )</td>
<td>([-\infty, \infty])</td>
</tr>
<tr>
<td>Uniform</td>
<td>( \frac{1}{2} )</td>
<td>Legendre ( P_n(x) )</td>
<td>1</td>
<td>([-1, 1])</td>
</tr>
<tr>
<td>Beta</td>
<td>( \frac{(1-x)^\alpha(1+x)^\beta}{2^{\alpha+\beta+1}B(\alpha+1, \beta+1)} )</td>
<td>Jacobi ( P_n^{(\alpha,\beta)}(x) )</td>
<td>( (1-x)^\alpha(1+x)^\beta )</td>
<td>([-1, 1])</td>
</tr>
<tr>
<td>Exponential</td>
<td>( e^{-x} )</td>
<td>Laguerre ( L_n(x) )</td>
<td>( e^{-x} )</td>
<td>([0, \infty])</td>
</tr>
<tr>
<td>Gamma</td>
<td>( \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} )</td>
<td>Generalized Laguerre ( L_n^{(\alpha)}(x) )</td>
<td>( x^\alpha e^{-x} )</td>
<td>([0, \infty])</td>
</tr>
</tbody>
</table>

Additional bases generated numerically via Golub-Welsch

• Tailor expansion type/order/range:
  – Total order \( \rightarrow \) tensor and sum of tensor expansions
  – Dimension p-refinement: anisotropic tensor/sparse grids
  – Domain h-refinement: discretization of random domain

1/sqrt(N) for LHS

super-algebraic for num. integration & regression
New DOE ASCR Project (Office of Science): FY2010-2012

Short term:
- MATLAB management of NREL design tool ensemble ("EOLO", Sandia wind group)

Longer term:
- CFD with Joe (Stanford) and FSI with SIERRA/Aria (Sandia)

UQ Research goals
- Inner core: probabilistic UQ
  - Adjoint EE \(\rightarrow\) balance of determ/stoch errors
  - Adaptive, adjoint-enhanced expansions
- Aggregation & Data Fusion
  - Mixed Aleatory/Epistemic UQ
  - Model Form: multifidelity UQ
  - Data fusion: Bayesian inference \(\rightarrow\) BMA
Simulation Management (Black Box case)

**DAKOTA Input File**
- Commands
- Options
- Parameter definitions
- File names

**DAKOTA Parameters File**

\{
  x_1 = 123.4 \\
  x_2 = -33.3 \\
\}, etc.

**DAKOTA Executable**

Sensitivity Analysis, Optimization, Uncertainty Quantification, Parameter Estimation

**DAKOTA Executable**

sim_code_script

**DAKOTA executes**

sim_code_script to launch a simulation job

**DAKOTA Results File**

999.888 \( f_1 \)  \\
777.666 \( f_2 \), etc.

**DAKOTA Output Files**

- Raw data (all x- and f-values)
- Sensitivity info
- Statistics on f-values
- Optimality info

Use APREPRO/DPREPRO to cut-and-paste x-values into code input file

**Code Input**

**Code Output**

**User-supplied** automatic post-processing of code output data into f-values

**Code**

- CALORE: thermal analysis
- ALEGRA: shock physics
- SALINAS: structural dynamics
- Premo: high speed flow
  (your code here)

**Sandia National Laboratories**
Parallelism Options: Multicore Desktops to MPP

1. Algorithmic coarse-grained: concurrency in data requests:
   - Iterators: Gradient-based, Nongradient-based, Surrogate-based
   - Strategies with concurrent Iterators: Multi-start, Pareto, Hybrid
   - Nested Models: OUU/MCUU, Mixed UQ
2. Algorithmic fine-grained: computing the internal linear algebra of an opt. algorithm in parallel
3. Fn eval coarse-grained: concurrent execution of separable simulations within each fn. eval.
4. Fn eval fine-grained: parallelization of the solution steps within a single analysis code
Deployment

**Impact Sandia missions**
- Technology insertion
  - ASC milestones
  - Early adopters
  Jan/Feb 2010: 92% of DAKOTA invocations on SNL clusters were UQ or param studies, but new methods starting to reduce LHS dominance

**Partnerships**
- Government: LLNL, LANL, ORNL, INL, NASA, DOD
- Industry: Lockheed Martin, Goodyear, Exxon Mobil
- University: MIT, Cornell, CU Boulder, Vanderbilt, USC, FSU, Notre Dame, VPISU, UNM
  - CSRI students/postdocs, faculty sabbaticals
  - ASC PSAAP: UT Austin (Bayesian), Purdue (cubature), UIUC (adaptive collocation), Caltech (global opt.), Michigan (gradient-enhanced interpolation), Stanford (adaptive collocation)

**Address core usability barriers**
- JAGUAR
- Library embedding
Deployment Initiative: JAGUAR User Interface

- Eclipse-based rendering of full DAKOTA input spec.
- Automatic syntax updates
- Tool tips, Web links, help
- Symbolics, sim. interfacing

- Flat text editor for experienced users
- Keyword completion
- Automatically synchronized with GUI widgets

- Simplified views for high-use applications ("Wizards")

Impact: streamline problem set-up for user base, spanning novices to experts
Deployment Initiative: Embedding

Make DAKOTA natively available within application codes

- Streamline problem set-up, reduce complexity, and lower barriers
  - A few additional commands within existing simulation input spec.
  - Eliminate analysis driver creation & streamline analysis (e.g., file I/O)
  - Simplify parallel execution
- Integrated options for algorithm intrusion

SNL Embedding

- Existing: Xyce, Sage, Albany (TriKOTA)
- New: ALEGRA, SIERRA (TriKOTA) → STK

External Embedding

- Existing: ModelCenter, university applications
- New: QUESO (UT Austin), R7 (INL)
- Expanding our external focus:
  - GPL → LGPL; svn restricted → open network

ModelEvaluator Levels

Non-intrusive

ModelEvaluator: systems analysis
- All residuals eliminated, coupling satisfied
- DAKOTA optimization & UQ

Intrusive to coupling

ModelEvaluator: multiphysics
- Individual physics residuals eliminated; coupling enforced by opt/UQ
- DAKOTA opt/UQ & MOOCHO opt.

Intrusive to physics

ModelEvaluator: single physics
- MOOCHO opt., Stokhos UQ, NOX, LOCA

Impact: eliminate custom set-up and support fully integrated opt. and UQ studies
Concluding Remarks

**DAKOTA provides a variety of core algorithms for iterative analysis:**

- Optimization
- Calibration

**As well as advanced capabilities for**

- Multilevel parallel computing
- Manage multiple iterative methods, models of varying fidelity, nesting, recasting, etc.
- Emerging UQ methods: adaptive, adjoint-enhanced, multi-{fidelity, physics, scale}, mixed UQ
- Emerging algos. in other areas: OUU, SBO, MINLP, SA w/ PCE/SC, Nond. calibration

**Advanced deployment initiatives will “lower the bar” for adoption**

- JAGUAR

**Expanding from NNSA to include energy missions: Wind, NE**

**Some lessons learned in open source framework development**

- Bound your mission space and manage scope creep
  - Focus on your core strengths and provide flexible APIs for others to use
  - Be selective on strategic partnerships
- Establish a support hierarchy and manage it effectively
  - Small teams may need to rely on community support for bottom tier
- Utilize modern CS tools (svn/git, cmake/scons, Trac) to simplify collaborative development
- Manage quality through sponsorship and review of external contributions