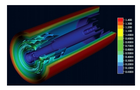
Algorithmic Developments for UQ in Extreme Scale Simulations

Michael Eldred, John Jakeman, Khachik Sargsyan, Cosmin Safta, Jaideep Ray, Kenny Chowdhary, Bert Debusschere*, Habib Najm

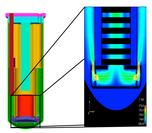
Sandia National Laboratories

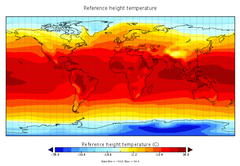
* bjdebus@sandia.gov

QUEST algorithmic developments focus on key challenges in extreme scale computing



- High Dimensionality
  - Gradient-enhanced interpolants
  - Local error estimation with hierarchical value/gradient surpluses
  - Compressive sensing for sparsity detection
- Data characterization:
  - Data-Free Inference to handle missing data
  - Random field representations
- Predictive model calibration:
  - Subgrid model upscaling

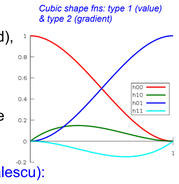




Gradient-enhanced interpolants: global h-refinement

Dimension-adaptive h-refinement for SC:

- *Local spline interpolants:* linear Lagrange (value-based), cubic Hermite (gradient-enhanced)
- *Global grids:* iso/aniso tensor, iso/aniso/generalized sparse
- *h-refinement:* uniform, adaptive, goal-oriented adaptive
- *Basis formulations:* nodal, hierarchical



Multivariate tensor product to arbitrary derivative order (Lalescu):

$$s^{(n)}(x_1, x_2, \ldots, x_D) = \sum_{l_1,\ldots,l_D=0}^{m} \sum_{i_1,\ldots,i_D=0,1} f^{(l_1,\ldots,l_D)}(i_1,\ldots,i_D) \prod_{k=1}^{D} \alpha_{i_k}^{(n,l_k)}(x_k)$$

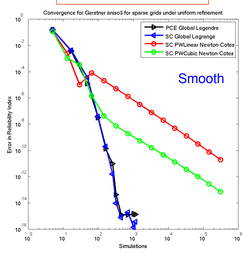
$$f = \sum_{i=1}^{N} f_i H_i^{(1)}(x_1) H_i^{(1)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_1} H_i^{(2)}(x_1) H_i^{(1)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_2} H_i^{(1)}(x_1) H_i^{(2)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_3} H_i^{(1)}(x_1) H_i^{(1)}(x_2) H_i^{(2)}(x_3)$$

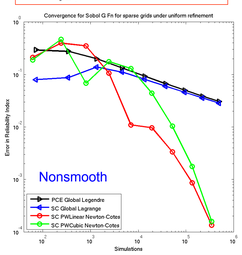
$$\mu = \sum_{i=1}^{N} f_i w_i^{(1)} w_i^{(1)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_1} w_i^{(2)} w_i^{(1)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_2} w_i^{(1)} w_i^{(2)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_3} w_i^{(1)} w_i^{(1)} w_i^{(2)}$$

and similar for higher-order moments

$$f(\mathbf{x}) = \exp\left(-10x_1^2 - 5x_2^2\right)$$

$$f(\mathbf{x}) = 2\prod_{j=1}^{5} \frac{|4x_j - 2| + a_j}{1 + a_j}; \quad a = [0,1,2,4,8]$$



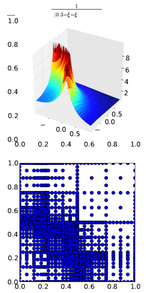


Local Error Estimation with Hierarchical Value/Gradient Surpluses

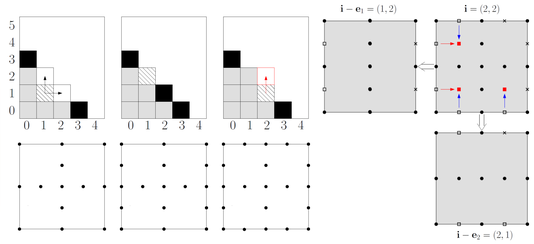
Hierarchical basis:

- *Improved precision* in QoI (mean and higher order moments) increments
- Surpluses provide error estimates for *local refinement* using local/global hierarchical interpolants
- New QoI error indicators naturally leverage *both value and gradient surpluses*

$$\Delta\Sigma_{ij} = \Delta E[R_i R_j] - \mu_{R_i}\,\Delta E[R_j] - \mu_{R_j}\,\Delta E[R_i] - \Delta E[R_i]\Delta E[R_j]$$

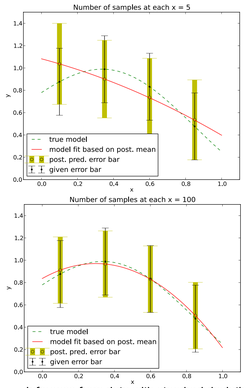


From J. Jakeman, July 2010



Data-Free Inference (DFI)

- Inputs often specified incompletely in the literature
  - Mean and confidence interval provided, but no covariance. Original data missing.
  - Data specified as mean and standard deviation, but number of samples missing
- Overall output uncertainty depends strongly on joint distribution of input parameters
- DFI infers joint distribution of parameters based on available information
  - Pool family of posteriors that are consistent with given information [Berry *et al.*, JCP, 2012]
- Application to
  - Inference of chemical kinetics
  - Inference from data products (processed raw data)

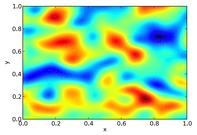


Inference from data with standard deviations but unknown number of samples

Random Field Characterization

- Random fields are conveniently represented with Karhunen-Loève Expansions (KLEs)
  - Enables dimensionality reduction
- KLEs require a covariance matrix
  - Sample covariance matrix requires large ensemble of realizations before eigenfunctions are smooth
  - Sometimes very few samples are available
- Ongoing work on estimation of covariance matrices from few realizations

$$F(x,\theta) = \langle F(x,\theta)\rangle_\theta + \sum_{k=1}^{\infty} \sqrt{\lambda_k} f_k(x) \xi_k$$



2D Gaussian Process realization

Stochastic Expansions on Unstructured Grids: Compressive Sensing

$$\mathbf{\Phi c} = \mathbf{y} \qquad (1)$$

where the elements of the measurement matrix are given by

$$\Phi_{ij} = \Phi_j(\boldsymbol{\xi}^{(i)}), \quad i = 1,\ldots,N,\ j = 1,\ldots,P$$

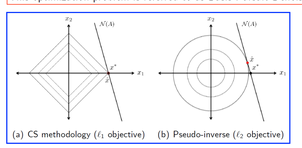
When $N \ll P$ (1) is ill-posed and additional constraints must be introduced to find a unique solution. The traditional way of estimating c is to find the minimum norm solution, which is the optimal set of coefficients

$$\mathbf{c} = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{\Phi c} - \mathbf{y}\|_2^2$$

If the solution is sparse we can do better by seeking a solution

$$\mathbf{c} = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{c}\|_1 \quad \text{such that} \quad \|\mathbf{\Phi c} - \mathbf{y}\|_2 \le \varepsilon \qquad (2)$$

This optimization problem is referred to as Basis Pursuit Denoising

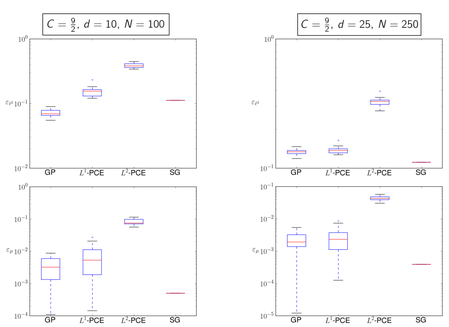


Courtesy of G. Tang (Stanford)

Solvers:

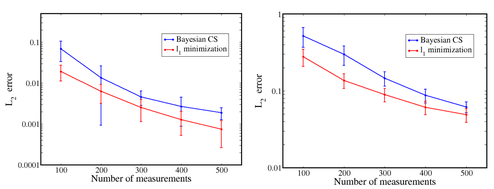
- l_0 denoising with Orthogonal Matching Pursuit
- l_1 Basis Pursuit with Linear Programming solvers
- l_1 Basis Pursuit Denoising with Interior Point solvers

$$f_{OS}(\mathbf{x}) = \cos\left(-\sum_{i=1}^{d} c_i x_i\right), \quad \mathbf{x} \in [0,1]^d \qquad c_i = \frac{\left(i - \frac{1}{2}\right)}{d}\frac{C}{W} \qquad \sum_{i=1}^{d} c_i = C$$



- With high-dimensional input, spectral representations require too many terms, therefore building sparse bases is important.
- Borrowing from the machine learning community, applying sparse learning methods to polynomial bases.
  - Evaluated and/or extended various Compressive Sensing (CS) and Bayesian Compressive Sensing (BCS) approaches
  - Analyzed the dependence of performance on the effective dimensionality (sparsity) of the model and the number of model simulations available.
  - K-folds cross validation is applied to determine the polynomial degree and tolerance ε that produce a PCE that generalizes well to new data [Doostan *et al.*, JCP 2011]
  - An iterative BCS methodology developed and applied to Community Land Model (CLM) in the context of CSSEF (Climate Science for a Sustainable Energy Future).
- Future directions:
  - Adjoint-enhanced CS
  - Least interpolation

Comparison of l_1-minimization CS and BCS algorithms



BCS/CS comparison for 5d, order 10(left) and 10d, order 5(right) Genz corner-peak functions

- CS generally lead to smaller errors
- BCS runs faster with the default settings and stores only relevant bases
- BCS computes uncertainty estimates enabling a posterior predictive distribution

- Dimensionality reduction by using hierarchical priors

$$p(c_k|\sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{c_k^2}{2\sigma_k^2}} \qquad p(\sigma_k^2|\alpha) = \frac{\alpha}{2} e^{-\frac{\alpha\sigma_k^2}{2}}$$

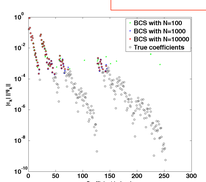
- The parameter α can be further modeled hierarchically, or fixed.
- Evidence maximization dictates values for $\sigma_k^2, \alpha, \sigma^2$ and allows exact Bayesian solution

$$\boldsymbol{c} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

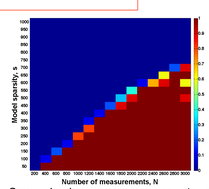
with

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{P}^T\boldsymbol{u} \qquad \boldsymbol{\Sigma} = \sigma^2(\boldsymbol{P}^T\boldsymbol{P} + \operatorname{diag}(\sigma^2/\sigma_k^2))^{-1}$$

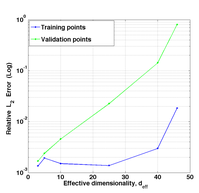
$$f = \cos\left(2\pi e + \sum_{i=1}^{d} a_i x_i\right) \qquad a_i = \frac{1}{i^2}$$

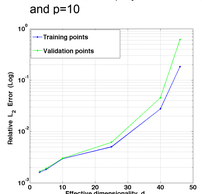


The coefficient recovery for a 5-dimensional Genz oscillatory function up to total order 5. True coefficients are computed by a high-level sparse quadrature.



Sparse basis recovery success rate, out of 10 replicas, vs. the number of measurements and the model sparsity, i.e. the number of non-zero polynomial coefficients in a polynomial model of d=5 and p=10





Relative L2 error measures for both training and validation datasets as the effective dimensionality increases, for three different values of the number of measurements, N = 1000 (left), N = 5000 (right) for d = 50. Effective dimensionality is the number of dimensions that contain 95% of the total sum of a_i's.

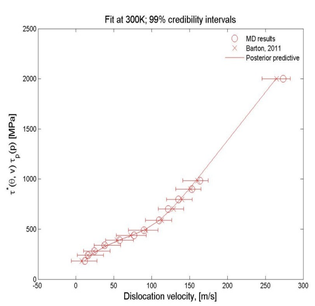
Predictive Model Calibration

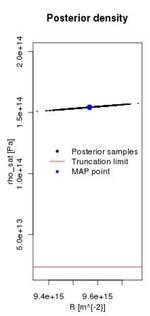
- Multiscale-Multiphysics simulations often rely on subgrid models, based on
  - First principles or other higher-fidelity simulations
  - Observational data
- To asses predictive fidelity of overall simulation, need to
  - Properly calibrate subgrid models
  - Propagate uncertainty in subgrid models across scales (upscaling)

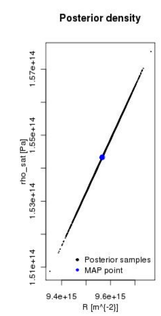
- Application to ParaDiS
  - a LLNL dislocation density (DD) model
  - used to create "sub-grid" models of crystal dislocation dynamics for use in continuum codes (ALE3D)
- Inputs:
  - Dislocation velocity model, by upscaling MD experiments
  - Various operational conditions (P, T, strain rate etc)
- Output
  - Simple ODE models, fitted to ParaDiS simulation of dislocation density and stress-strain evolution

- Analyses done only for Ta
- MD -> DD upscaling [Big source of uncertainty, for Ta]
  - Fitted the dislocation density model to MD simulation data, using MCMC
  - Forward UQ propagation shows significant spread in dislocation density predictions
    - A big source of input uncertainty
- DD -> continuum upscaling [Smaller source of uncertainty, for Ta]
  - ODE models fits ParaDiS outputs very well
    - Posterior distribution of model parameters estimated

- Left: MD -> DD upscaling (uncertainty in model predictions)
- Right: Posterior distribution of DD -> Continuum upscaling model







Support for this work was provided through the Scientific Discovery through Advanced Computing (SciDAC) project funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories
www.sandia.gov

QUANTIFICATION OF UNCERTAINTY IN EXTREME SCALE COMPUTATIONS
QUEST
www.quest-scidac.org