SAND2018-1746PE

Scalable Uncertainty Quantification: Exploiting Structure in Models and Data

John Jakeman, Cosmin Safta
(jdjakem,csafta@sandia.gov)
Sandia National Laboratories

Roger Ghanem
(ghanem@usc.edu)
University of Southern California

Overview of the technology

The curse of parameter dimensionality. UQ is computationally demanding. The number of simulations required to explore uncertainties grows exponentially with the number or model uncertainties

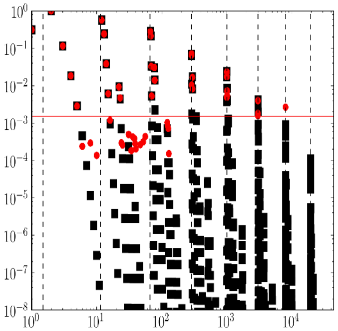
- **Technology Area:**
  - Quantify uncertainty in the output of high-fidelity simulations of realistic multiscale, multiphysics applications
- **FASTMath Tasks**
  - Deploy surrogates that exploit structure in parameter to output map. Utilize sampling and preconditioning schemes that ensure stability and target identification and exploitation of structure.
  - Integrate manifold construction with FASTMath eigensolvers and develop preconditioning schemes to accelerate convergence of basis adaptation schemes.
- **Applications Impacted**
  - Probabilistic Sea Level Projections from Ice Sheet and Earth System Models
  - Optimization of Sensor Networks for Improving Climate Model Predictions (OSCM)
  - Simulation of Fission Gas in Uranium Oxide Nuclear Fuel

FASTMATH

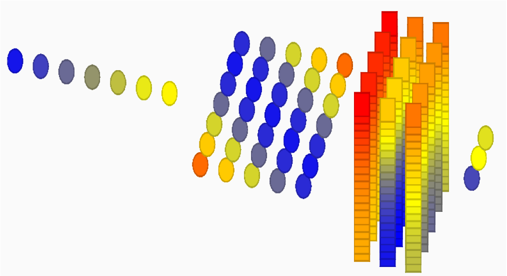
Motivation

Develop efficient methods for approximating high-dimensional functions, which exploit structure in the input-output map

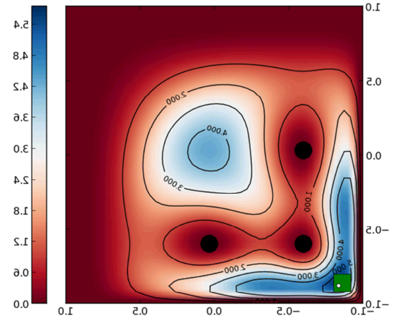
- The number of samples required by structure unaware function approximation grows exponentially with number of uncertainties.



- Identifying and exploiting structure can result in a sample complexity which is only weakly dependent on dimension.



- Goal-oriented sampling and approximation can further increase computational savings



FASTMATH

Sparse approximation

Approximate function with small number of nonzero terms $f_\Lambda(z) = \sum_{\lambda \in \Lambda} \alpha_\lambda \phi_\lambda(z)$

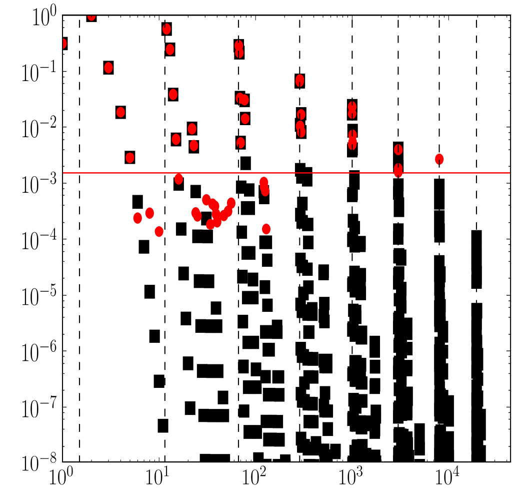
$$s = \#\{\lambda \mid |\alpha_\lambda| > \delta\}$$

l_0 -minimization (NP HARD)

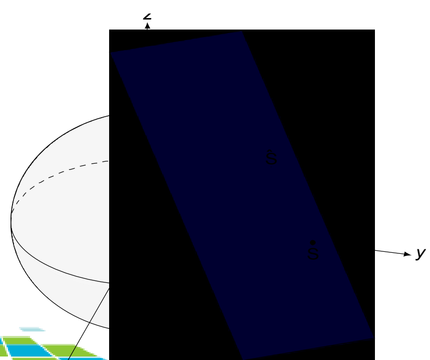
$$\min_\alpha \|\alpha\|_0 \; s.t. \; \|f - f_\Lambda\|_2 \le \epsilon$$

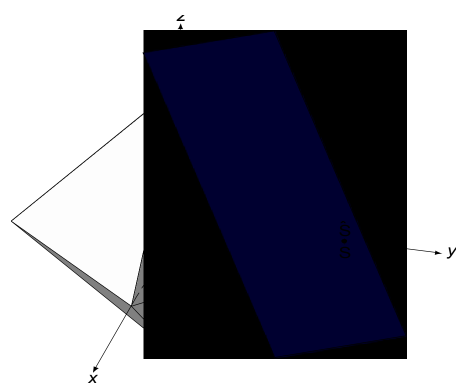
l_1 -minimization (Finds sparse solution under certain conditions)

$$\min_\alpha \|\alpha\|_1 \; s.t. \; \|f - f_\Lambda\|_2 \le \epsilon$$



Sparse approximation of a function





Complexity

l_2-minimization (least squares)

$$M = O(N \log N)$$

l_1-minimization

$$M = O(s \log^3 s \log N)$$

FASTMATH

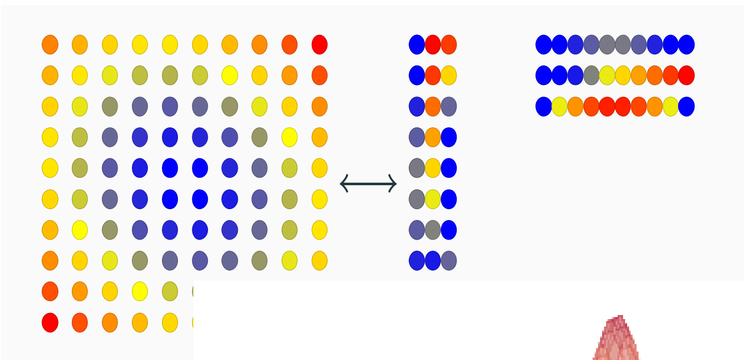
Low-rank tensor decompositions

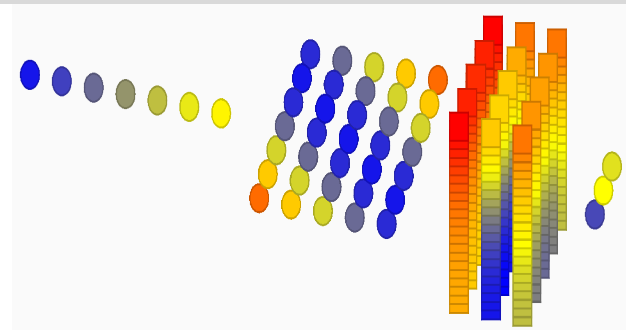
The canonical decomposition

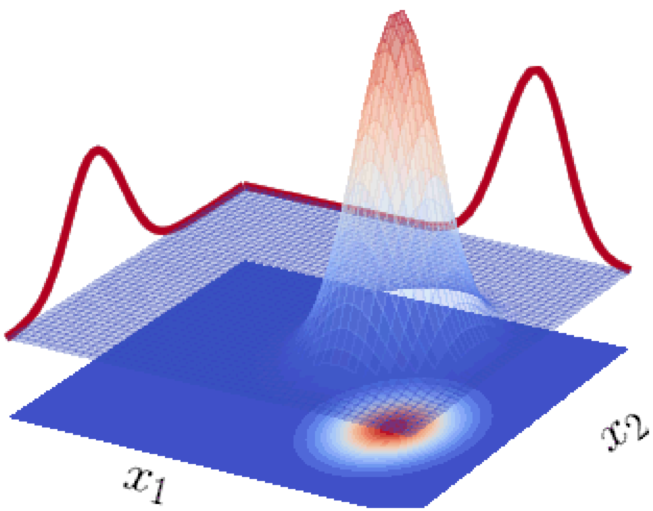
$$A = \sum_{i=1}^{r} v_1 \circ \cdots \circ v_d$$

Tensor-train decomposition

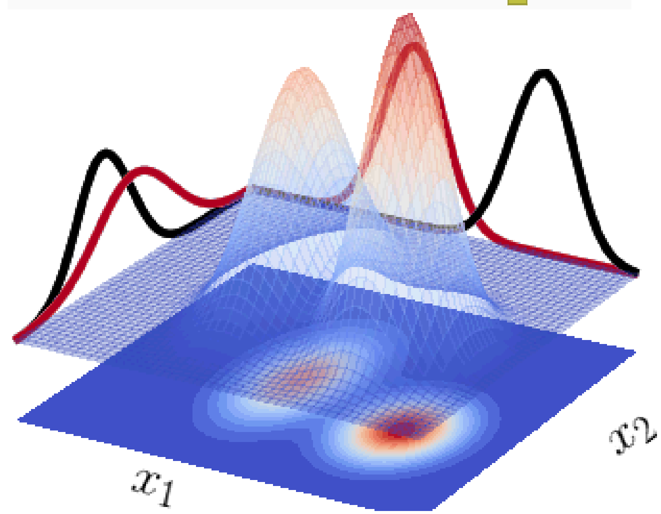
$$f_r(x) = \sum_{i_0=1}^{r_0} \sum_{i_1=0}^{r_1} \cdots \sum_{i_d=1}^{r_d} f_1^{i_0 i_1}(x_1) f_2^{i_1 i_2}(x_2) \cdots f_d^{i_{d-1} i_d}(x_d)$$







$r = 1$



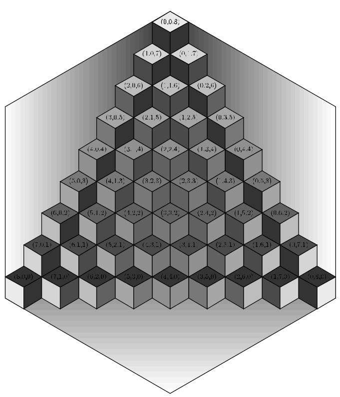
$r = 2$

TT Complexity $M = O(pdr^2)$

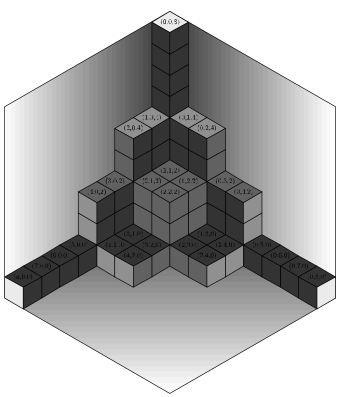
Goal-oriented approximation

The choice of the index set Λ in $f_\Lambda(z) = \sum_{\lambda\in\Lambda} \alpha_\lambda \phi_\lambda(z)$ can significantly influence the accuracy for a fixed number of samples M

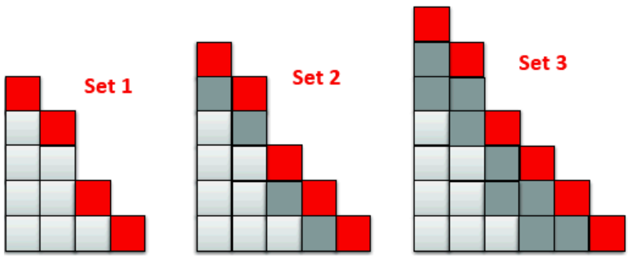
- The best index set is problem dependent
- Any knowledge on nature of variable interactions can be incorporated into index set a priori



Total-degree indices



Sparse grid indices



Index set can be determined adaptively

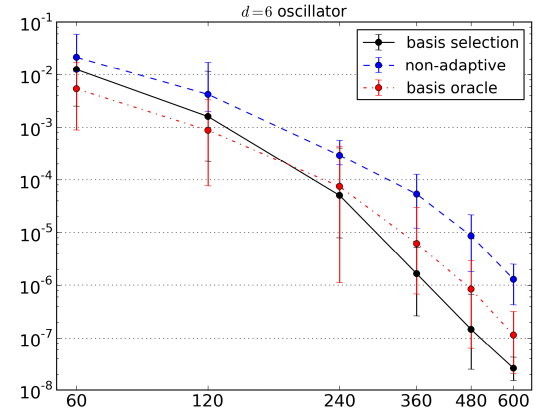
The multi index λ has the form

$$\lambda = (\lambda_1, \ldots, \lambda_d)$$

Each entry specifies the polynomial degree in each dimension of ϕ_λ

$$\phi_{(\lambda_1,\lambda_2\lambda_3)=x_1^{\lambda_1}\cdot x_2^{\lambda_2}\cdot x_3^{\lambda_3}}$$

Error for fixed M can improve significantly when basis adaptivity is used



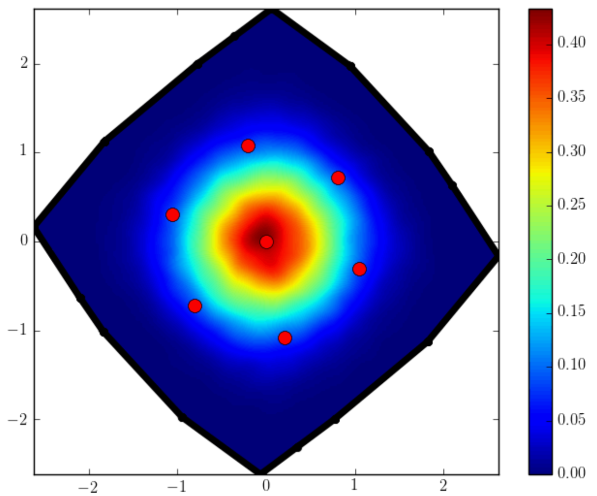
FASTMATH

Goal-oriented sampling

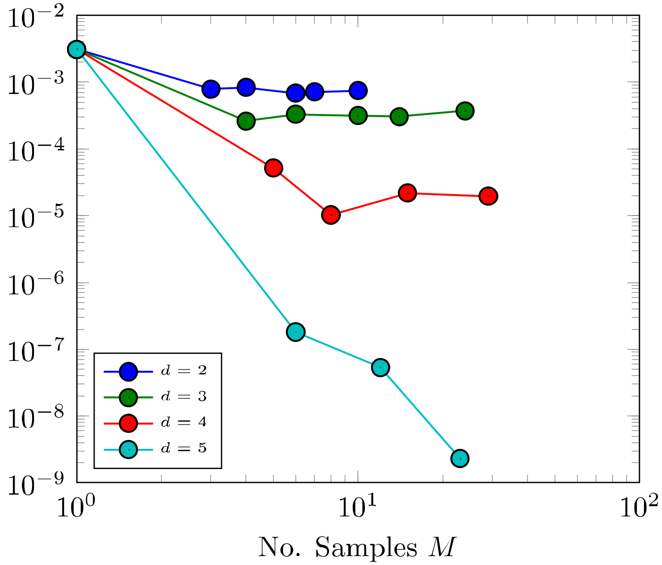
Solve optimization problem to find minimal quadrature rule such that we exactly integrate a set of polynomial moments, i.e.

$$\int_{I_z} \phi_\lambda(z)dp(z) = \sum_{m=1}^{M} \phi_\lambda(z_m)w_m, \forall \lambda \in \Lambda$$

subject to linear inequality constraints defining zonotope vertices



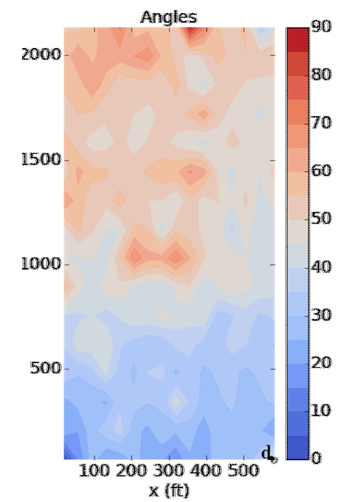
PDF of 2D active subspace of aircraft nozzle model and 5th order minimal cubature rule on the zonotope



Accuracy of moments again tied to dimension of subspace

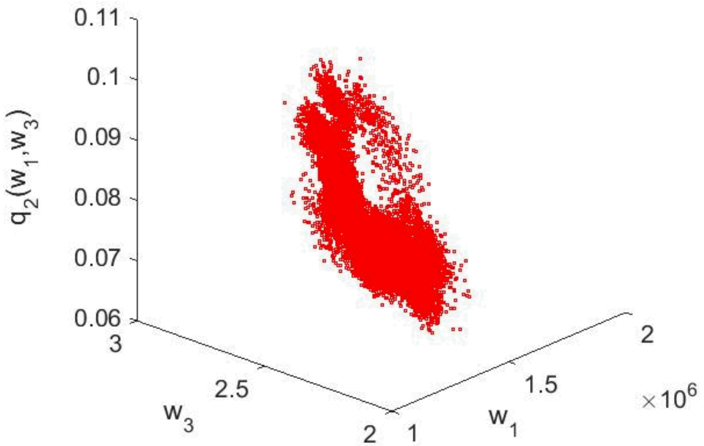
Dimension reduction

- Stochastic basis adaptation
  - Adaptations are judicious rotations of the input uncertainties
  - Learn adaptations as statistics of data
  - Implement adaptations as preconditioners
  - Interpolate adaptations over parameter and design space
  - Multifidelity transitions with adaptations



- Sampling and UQ on manifolds
  - Structure in multi dimensional data helps explains big part of statistical fluctuations.
  - Structure is discovered and characterized as diffusion manifolds.
  - Sampling on these manifolds from target distributions augments expensive and scarce data.

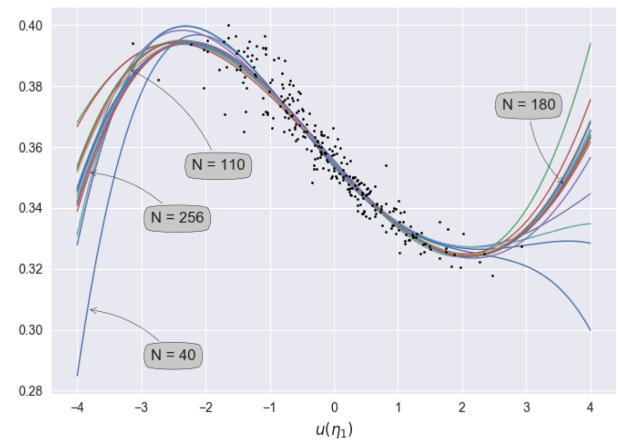
$q_{2,ar}$ with ν_{sim} = 25,800 additional samples

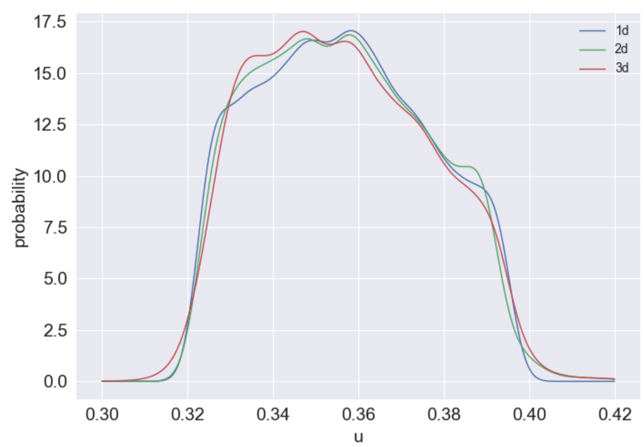


FASTMATH

Dimension reduction: Stochastic basis adaptation

- Stochastic basis adaptation
  - Initial random input described as nonlinear map on standard Gaussian vector ξ
  - Introduce new hyperparameter rotation matrix $\mathcal{A}$ with $\eta = \mathcal{A}\xi$
  - Rewrite system equations in terms of η
  - Find $\mathcal{A}$ so that new system equations have desired features:
    - Sparsity (signature of only the first few η)
    - Accuracy
    - Sufficiency (learned enough from data)
    - Interpolate $\mathcal{A}$ over parameter and design space.

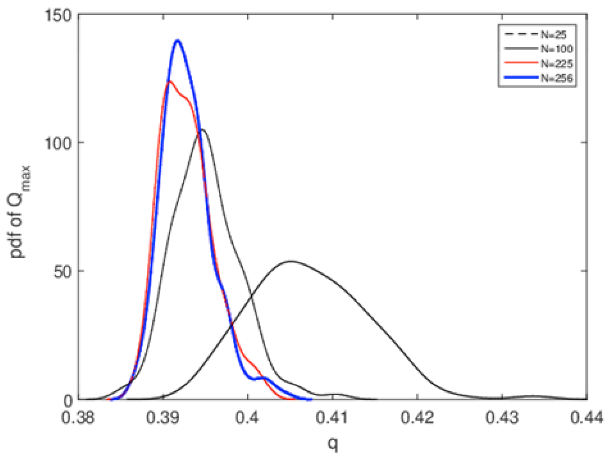


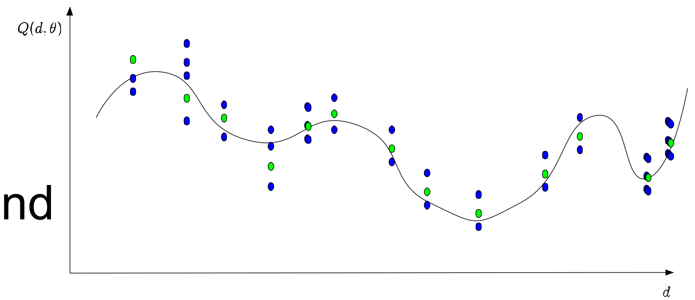


FASTMATH

Dimension reduction: Sampling and UQ on manifolds

- Sampling and UQ on manifolds
  - Using a few expensive or scarce samples
  - Samples consist of QoI, controls, and uncertain parameters
  - Construct a joint probability density of all observables
  - Identify diffusion manifold for observables by solving an **eigenvalue problem**
  - Construct an Ito equation that samples on this manifold from that density function
  - Solve the **Ito equation in parallel** many thousand times
  - Integrate into **UQ-solvers** and **optimization** software

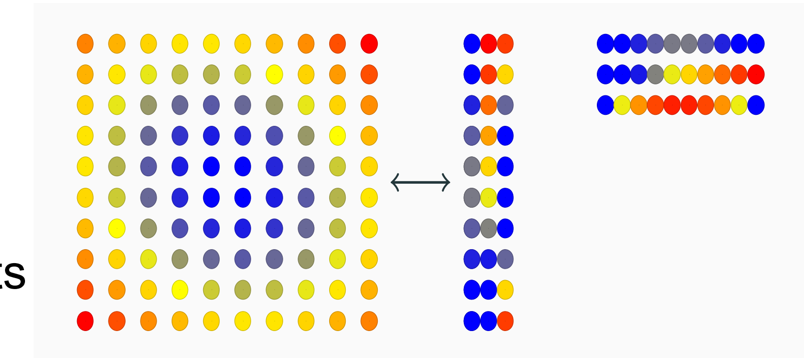




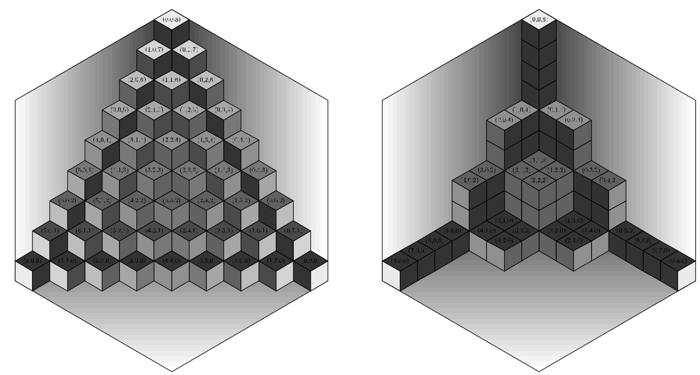
FASTMATH

Key points

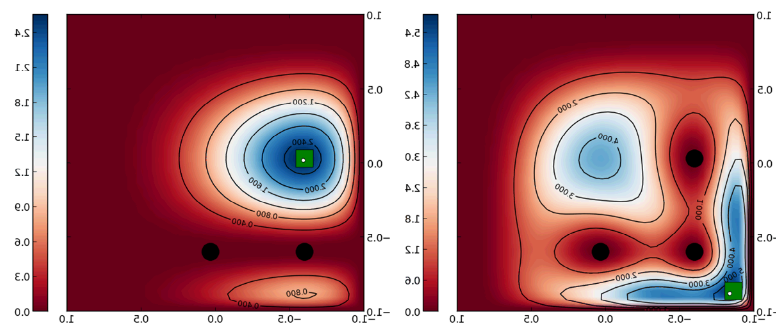
- Low-rank tensor approximations
  - High-dimensional analogy to SVD
  - Can be used to reduce sample complexity of UQ
  - Can also be used to compress large data sets



- Sparse approximation
  - Identify largest contributions to function error
  - Can also compress data sets



- Goal-oriented methods
  - Target dimensions, regions, subspaces that significantly inform quantities of interest
- Sampling schemes
  - Ensure stability, target regions of high probability
  - Support mixed continuous discrete variables



Proposed work in FASTMath

Year 1

- Deploy low-rank decomposition of mixed time, spatial, and stochastic spaces.
- Integrate manifold sampling with FASTMath eigensolvers.
- Develop iterative logic so that as dimensions are added in basis adaptation, previous solutions are used for convergence acceleration.

Year 2

- Develop well-conditioned sampling strategies for high-dimensional approximation of functions, with both continuous and discrete parameters
- Parallelize solutions of the Ito stochastic differential equation.
- Formulate the problem of interpolating basis adaptation isometries: this will involve suitable parameterization of the isometries.

Year 3

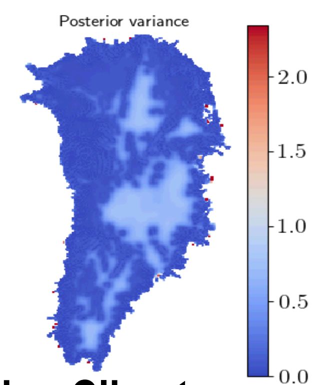
- Develop goal-oriented low-rank and sparse approximation methods
- Develop algorithms to pool information from multiple Ito SDEs for sampling on manifolds.
- Implement interpolation algorithms for adaptations.

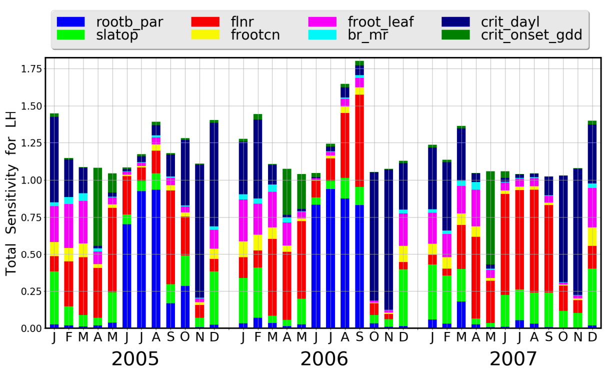
Description of the software tools

- Dakota
  - Optimization, adaptive sparse quadrature, low-rank representations, compressive sensing, Bayesian inference
  - Web: `https://dakota.sandia.gov`
- UQTk
  - Sparse regression, random field representations, global sensitivity analysis
  - Web: `http://www.sandia.gov/UQToolkit`
- Chaos_Toolkit
  - Python toolkit for recursive basis adaptation with convergence checks, with both specialized and generic interfaces

Application Interactions

- **Probabilistic Sea Level Projections from Ice Sheet and Earth System Models (ProSPect)**
  - PIs: Stephen Price (LANL), Esmond Ng (LBNL)
  - Tasks: Goal oriented dimension reduction to infer uncertainties from data and quantify impact on sea level rise.
  - Impact: The first probabilistic estimates of sea-level rise due to ice-sheet mass loss





- **Optimization of Sensor Networks for Improving Climate Model Predictions (OSCM)**
  - PI: Daniel Ricciuto (ORNL)
  - Tasks: Dimension reduction & calibration of land-atmosphere models; optimal design of measurement network
  - Impact: Identify which processes are most responsible for predictive uncertainties and reduce uncertainties with targeted observations

- **NE-ASCR partnership: Simulation of Fission Gas in Uranium Oxide Nuclear Fuel**
  - PI: David Andersson
  - Tasks: Forward and inverse UQ studies
  - Impact: Modeling of degradation in nuclear fuel rods. Improved understanding of gas bubble diffusion and dynamics in nuclear fuel materials.

FASTMATH