Final Report

DOE Award Number:	DE-SC0007099
Recipient:	Massachusetts Institute of Technology
Project Title:	Quantification of Uncertainty in Extreme Scale Computations (QUEST)
Principal Investigator:	Prof. Youssef M. Marzouk Department of Aeronautics and Astronautics Massachusetts Institute of Technology 77 Massachusetts Ave, Room 37-451 Cambridge, MA 02139
Date:	9 June 2017
Period Covered:	1 September 2011 – 30 August 2016

1 Summary

QUEST (`www.quest-scidac.org`) is a SciDAC Institute that is focused on uncertainty quantification (UQ) in large-scale scientific computations. Our goals are to (1) advance the state of the art in UQ mathematics, algorithms, and software; and (2) provide modeling, algorithmic, and general UQ expertise, together with software tools, to other SciDAC projects, thereby enabling and guiding a broad range of UQ activities in their respective contexts. QUEST is a collaboration among six institutions (Sandia National Laboratories, Los Alamos National Laboratory, the University of Southern California, Massachusetts Institute of Technology, the University of Texas at Austin, and Duke University) with a history of joint work in leading UQ research.

Our vision encompasses all aspects of UQ in leadership-class computing. This includes the well-founded setup of UQ problems; characterization of the input space given available data/information; local and global sensitivity analysis; adaptive dimensionality and order reduction; forward and inverse propagation of uncertainty; handling of application code failures, missing data, and hardware/software fault tolerance; and model inadequacy, comparison, validation, selection, and averaging. The nature of the UQ problem requires the seamless combination of data, models, and information across this landscape in a manner that provides a self-consistent quantification of requisite uncertainties in predictions from computational models. Accordingly, our UQ methods and tools span an interdisciplinary space across applied math, information theory, and statistics.

The MIT QUEST effort centers on statistical inference and methods for surrogate or reduced-order modeling. MIT personnel have been responsible for the development of adaptive sampling methods, methods for approximating computationally intensive models, and software for both forward uncertainty propagation and statistical inverse problems.

A key software product of the MIT QUEST effort is the MIT Uncertainty Quantification library, called MUQ (`muq.mit.edu`). A detailed description of MUQ is provided below.

2 Key technical contributions

Below we review some of the main technical achievements of the MIT QUEST effort. Most of these technical capabilities have been implemented in the MUQ software library, and others have also been adapted into other QUEST software products (e.g., UQTk).

The MIT PI has also had significant interactions with participants in the Center for Edge Physics Simulation (EPSI), one of the SciDAC Application Partnerships in Fusion Energy Sciences; these interactions focused on UQ methodologies for analyzing data from fusion experiments and for the solution of inverse problems, and have resulted in multiple talks and several journal publications already (J4 and J10).

2.1 Adaptive Smolyak pseudospectral approximations

Polynomial approximations of computationally intensive models are central to uncertainty quantification. We have developed an adaptive method for non-intrusive pseudospectral approximation, based on Smolyak's algorithm with generalized sparse grids. We rigorously analyzed and extended the non-adaptive method proposed in [Constantine, Phipps, & Eldred *CMAME* 2012] and compared it to a common alternative approach for using sparse grids to construct polynomial approximations, direct quadrature. Analysis of direct quadrature shows that $O(1)$ errors are an intrinsic property of some configurations of the method, as a consequence of internal aliasing. We provided precise conditions, based on the chosen polynomial basis and quadrature rules, under which this aliasing error occurs. We then established theoretical results on the accuracy of Smolyak pseudospectral approximation, and showed that the Smolyak approximation avoids internal aliasing and makes far more effective use of sparse function evaluations. These results are applicable to broad choices of quadrature rule and generalized sparse grids. Exploiting this flexibility, we introduce a greedy heuristic for adaptive refinement of the pseudospectral approximation. We numerically demonstrated convergence of the algorithm on the Genz test functions, and illustrate the accuracy and efficiency of the adaptive approach on a realistic chemical kinetics problem.

In collaboration with QUEST researchers at Duke University, we connected our adaptive Smolyak scheme to an extreme-scale database of realizations of an ocean general circulation model, using all available realizations to generate a sparse grid surrogate. Our results demonstrated that the adaptive algorithm leads to order-of-magnitude computational savings over sparse isotropic sampling and substantial accuracy improvements over naïve application of dimension-adaptive sparse quadrature.

2.2 Local approximation MCMC for inference with computationally intensive models

We construct a new framework for accelerating Markov chain Monte Carlo (MCMC) in posterior sampling problems where standard methods are limited by the computational cost of the likelihood, or of numerical models embedded therein. Our approach introduces local approximations of these models into the Metropolis-Hastings kernel, borrowing ideas from deterministic approximation theory, optimization, and experimental design. Previous efforts at integrating approximate models into inference typically sacrifice either the sampler's exactness or efficiency; our work addresses these limitations by exploiting useful convergence characteristics of local approximations. We prove the ergodicity of our approximate Markov chain, showing that it samples asymptotically from the *exact* posterior distribution of interest. We describe variations of the algorithm that employ either local polynomial approximations or local Gaussian process regressors. Our theoretical results reinforce the key observation underlying this paper: when the likelihood has some *local* regularity, the number of model evaluations per MCMC step can be greatly reduced without biasing the Monte Carlo average. Numerical experiments demonstrate multiple order-of-magnitude reductions in the number of forward model evaluations used in representative ODE and PDE inference problems, with both synthetic and real data.

2.3 Parallel local approximation MCMC

Performing Bayesian inference via MCMC can be exceedingly expensive when posterior evaluations invoke the evaluation of a computationally expensive model, such as a system of partial differential equations. Above we described a framework for constructing and refining local approximations of such models during an MCMC simulation. These posterior–adapted approximations harness regularity of the model to reduce the computational cost of inference while preserving asymptotic exactness of the Markov chain. Here we pursue two extensions of that work. First, focusing on the Metropolis–adjusted Langevin algorithm, we describe how a proposal distribution can successfully employ gradients and other relevant information extracted from the approximation. Second, we prove that samplers running in parallel can collaboratively construct a shared posterior approximation while ensuring ergodicity of each associated chain, providing a novel opportunity for exploiting parallel computation in MCMC. We investigate the practical performance of our strategies using two challenging inference problems, the first in subsurface hydrology and the second in glaciology. Using local approximations constructed via parallel chains, we successfully reduce the run time needed to characterize the posterior distributions in these problems from days to hours and from months to days, respectively, dramatically improving the tractability of Bayesian inference.

2.4 Spectral tensor-train (STT) surrogates

The accurate approximation of high-dimensional functions is an essential task in uncertainty quantification and many other fields. We propose a new function approximation scheme based on a spectral extension of the tensor-train (TT) decomposition. We first define a functional version of the TT decomposition and analyze its properties. We obtain results on the convergence of the decomposition, revealing links between the regularity of the function, the dimension of the input space, and the TT ranks. We also show that the regularity of the target function is preserved by the univariate functions (i.e., the "cores") comprising the functional TT decomposition. This result motivates an approximation scheme employing polynomial approximations of the cores. For functions with appropriate regularity, the resulting *spectral tensor-train decomposition* combines the favorable dimension-scaling of the TT decomposition with the spectral convergence rate of polynomial approximations, yielding efficient and accurate surrogates for high-dimensional functions. To construct these decompositions, we use the sampling algorithm `TT-DMRG-cross` to obtain the TT decomposition of tensors resulting from suitable discretizations of the target function. We assess the performance of the method on a range of numerical examples: a modifed set of Genz functions with dimension up to 100, and functions with mixed Fourier modes or with local features. We observe significant improvements in performance over an anisotropic adaptive Smolyak approach. The method is also used to approximate the solution of an elliptic PDE with random input data.

Code for STT has been implemented in Python and made publicly available, and in the coming period we plan to incorporate these algorithms into MUQ. Ongoing work also includes order-adaptive and parallelized versions of STT, and further comparison with previous state-of-the-art surrogate construction schemes.

We have also been developing new functional extensions of the tensor-train decomposition, which combine the high-dimensional approximation capabilities of TT with the "numerical computing with functions" philosophy/features of Chebfun and related projects (e.g., ApproxFun, pychebfun).

3 Software products: MUQ

MUQ (MIT Uncertainty Quantification) is a C++/Python library for uncertainty quantification—in particular, for connecting complex models with UQ tools in a way that exposes model structure to the algorithms. MUQ is designed both for use by application scientists and engineers, and as a platform for algorithm developers. It currently includes a wide variety of UQ capabilities: advanced Markov chain Monte Carlo algorithms for inference; approximation methods for computationally intensive likelihoods and forward models; adaptive methods (e.g., sparse polynomial approximations) for uncertainty propagation, global sensitivity analysis, and surrogate construction; and many others. It is intended to be a lightweight and flexible research capability that complements other QUEST tools. MUQ is released under an open-source (BSD) license, with updates continually appearing in the online repository at `http://muq.mit.edu`.

A key feature of MUQ's design is a graph-based approach to defining "joint" statistical and physical models. An enabler for ensemble computing, this approach optimizes forward and inverse UQ workflows through the use of directed acyclic graphs (DAGs) for dependency management. The underlying dependency graph enables structure-exploiting algorithms to work together in a relatively transparent fashion. For example, MUQ automatically caches expensive model evaluations and ensures that any available derivative information from the simulation model is seamlessly propagated through other objects and made available to MCMC samplers, optimization methods, and emulators.

Current MUQ capabilities include:

- Forward UQ and global sensitivity analysis via adaptive Smolyak pseudospectral approximations; adaptive sparse quadratures
- A suite of cutting-edge MCMC algorithms: adaptive Metropolis, Metropolis-adjusted Langevin, simplied manifold MALA, parallel tempering, Hamiltonian MCMC
- MCMC accelerated through the construction of transport maps
- Algorithms for deploying MCMC with computationally intensive forward models, through the construction of local approximation, in serial and in parallel
- Spatial statistical tools, including the construction of Karhunen-Loève expansions for arbitrary meshes and kernels
- Nonparametric regression

Underlying these efforts is the MUQ philosophy of *exposing model structure to algorithms* via a modular specification of each component of a UQ problem: prior or input parameter distributions, hyperpriors, observation operators, observational errors, likelihood functions, PDE or ODE-based forward models, predictive quantities of interest, etc. This structure is formalized by a graphical model that describes how information flows from one component to another. Explicit management of dependencies allows for parallel execution and for appropriate information, such as gradients and Hessians, to be passed to algorithms that can exploit it. MUQ also operates seamlessly with packages such as FEniCS, libMesh, Sacado, SUNDIALS, and NLopt. These tools help users build more sophisticated models using well-established numerical implementations.

Since the initiation of MUQ, we have held several user workshops, beginning with a two-day workshop at MIT in January 2015. We have also used MUQ as a platform for teaching and outreach at several UQ summer schools and external workshops: the two-week New Directions short course held at the Institute for Mathematics and its Applications (IMA) in June 2015; the ICERM IdeaLab on inverse problems in July 2015; and the upcoming Gene Golub SIAM Summer School on inverse problems and data assimilation, to be held in June 2018. In parallel, we are developing a full suite of tutorial problems, and comprehensive documentation for users and developers. We have also integrated MUQ's adaptive sparse sampling capabilities with UQTk, another QUEST software library developed at Sandia National Laboratories.

4 MIT personnel supported

P Conrad; PhD student (Sep 2011 to Apr 2014) then postdoctoral associate (May–Oct 2014)
M Parno; postdoctoral associate (Oct 2014 to Dec 2015)
D Bigoni; postdoctoral associate (June 2015 to Aug 2016)
Y Marzouk, Associate Professor of Aeronautics and Astronautics

5 Publications (cumulative)

5.1 Journal articles

J1 G. Bal, I. Langmore, and Y. M. Marzouk, "Bayesian inverse problems with Monte Carlo forward models," *Inverse Problems and Imaging* **7**, 81–105, 2013.

J2 J. Winokur, P. Conrad, I. Sraj, O.M. Knio, A. Srinivasan, W.C. Thacker, Y. M. Marzouk & M. Iskandarani,"A priori testing of sparse adaptive polynomial chaos expansions using an ocean general circulation model database," *Comput. Geosci.* **17**, 899–911 (2013).

J3 P. Conrad and Y. M. Marzouk, "Adaptive Smolyak pseudospectral approximations," *SIAM Journal on Scientific Computing* **35**, A2643–2670 (2013).

J4 M. A. Chilenski and M. Greenwald and Y. M. Marzouk and N. T. Howard and A. E. White and J. E. Rice and J. R. Walk, "Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression," *Nuclear Fusion* **55**, 023012 (2015).

J5 P. Conrad, Y. M. Marzouk, N. Pillai, A. Smith, "Accelerating asymptotically exact MCMC for computationally intensive models via local approximations," *Journal of the American Statistical Association* **111**, 1591–1607 (2016).

J6 D. Bigoni, A. P. Engsig-Karup, Y. M. Marzouk, "Spectral tensor-train decomposition," *SIAM Journal on Scientific Computing* **38** A2405–A2439 (2016).

J7 A. Gorodetsky, S. Karaman, Y. M. Marzouk, "Function-train: a continuous analogue of the tensor-train decomposition," submitted to *SIAM Journal on Scientific Computing* (2017), arXiv:1510.09088.

J8 P. Conrad, A. Davis, Y. M. Marzouk, N. Pillai, A. Smith, "Parallel local approximation MCMC for expensive models," submitted to *SIAM/ASA Journal on Uncertainty Quantification* (2017), arXiv:1607.02788.

J9 A. Gorodetsky, S. Karaman, Y. M. Marzouk, "High-dimensional stochastic optimal control using continuous tensor decompositions," submitted to *International Journal of Robotics Research* (2017), arXiv:1611.04706.

J10 M. Chilenski, M. Greenwald, A. Hubbard, J. Hughes, J. Lee, Y. M. Marzouk, J. Rice, A. White, "Experimentally testing the dependence of momentum transport on second derivatives using Gaussian process regression," submitted to *Nuclear Fusion* (2017).

5.2 Presentations

The past five years have included over 60 presentations by the PI and/or his graduate students and postdocs on their QUEST-supported research.

5.3 Other dissemination

Research supported by this DOE SciDAC project (in particular, local approximation MCMC described in Section 2.2) was described in an MIT News article, `http://news.mit.edu/2015/shrinking-bulls-eye-algorithm-speeds-complex-modeling-days-hours-1116` that was itself picked up by multiple national news outlets, including Slashdot.

Research supported by this DOE SciDAC project was also featured in the 2015 issue of *AeroAstro*, the annual magazine of the MIT Aeronautics and Astronautics department. Online version at `http://aeroastro.mit.edu/sites/aeroastro.mit.edu/files/aeroastro-magazine-2014-15.pdf`