Gradient Enhanced Universal Kriging Model for Uncertainty Propagation in Nuclear Engineering

Brian A. Lockwood[1] and Mihai Anitescu[2]

[1]Department of Mechanical Engineering
University of Wyoming

[2]Mathematics and Computer Science Division
Argonne National Laboratory

June 28, 2011

Introduction

- Engineering Simulations:
  - Map Analysis Inputs to Design Metrics using relevant Physics (PDE's)
  - Number of Inputs $\ggg$ Number of Design Metrics
  - Many Inputs Uncertain with corresponding distributions or intervals
- No longer sufficient to predict single metric for specific inputs

Goal: Use Uncertainty Quantification to provide simulation confidence bounds and enable robust optimization

- Uncertainty Quantification seeks to characterize distribution of simulation output
- Distribution of metric can predicted by exhaustive sampling

Uncertainty Quantification

Challenges:

- Statistics from exhaustive sampling converge slowly
- Brute Force approach prohibitively expensive for complex simulation
- Effort often increases rapidly with input dimension

Surrogate Model can be constructed based on limited number of simulation results

Requirements:

- Inexpensive to exhaustively sample
- Must capture design space with number of simulations within our budget (even as dimension increases)
- Must have error estimate to account for surrogate model limitations

Gradient Enhanced Universal Kriging Model used to fulfill these requirements

Problem Statement

Overview:

- Consider uncertain inputs, $\vec{x}$, and state s
- Simulation Output: $J(s, \vec{x})$
- State given by solving governing equations: $f(s, \vec{x}) = 0$
- Need to characterize distribution $\widehat{J}(\vec{x}) = J(s(\vec{x}), \vec{x})$
- Construct approximation $\widetilde{J}(\vec{x}, a) \approx \widehat{J}(\vec{x})$
- Sample $\widetilde{J}$ exhaustively

Observations:

- Approximation for $\widehat{J}(\vec{x})$ requires small number of observations
- Derivatives of $\widehat{J}(\vec{x})$ can provide additional training information for $\widetilde{J}$
- All derivatives of $\widehat{J}(\vec{x})$ w.r.t $\vec{x}$ can be calculated efficiently with Automatic differentiation and/or Adjoint methods

Universal Kriging Surrogate Model

- Assume $\widetilde{J}$ obeys Gaussian Process:

$$\widetilde{J} = N(m(\vec{x}), K(\vec{x}, \vec{x}'))$$

- $m(\vec{x})$: mean function
- $K(\vec{x}, \vec{x}')$: Covariance Matrix between Samples
- For Universal Kriging, polynomial regression used for mean
- Parameters determined via Maximum Likelihood
- Derivative observations included by considering joint distribution of function/derivatives

Gradient Enhancement

- Covariance Matrix extended to block matrix

$$\underline{K} = \begin{bmatrix} cov(Y,Y) & cov(Y,\nabla Y) \\ cov(\nabla Y, Y) & cov(\nabla Y, \nabla Y) \end{bmatrix}$$

- Function/Function

$$cov(y, y') = k(\vec{x}, \vec{x}').$$

- Derivative/Function

$$cov(\frac{\partial y}{\partial x_k}, y') = \frac{\partial}{\partial x_k} k(\vec{x}, \vec{x}').$$

- Derivative/Derivative

$$cov(\frac{\partial y}{\partial x_k}, \frac{\partial y'}{\partial x_l'}) = \frac{\partial^2}{\partial x_k \partial x_l'} k(\vec{x}, \vec{x}').$$

- Covariance Function must now be twice differentiable

Covariance Functions

- Covariance Function product of 1D functions

$$k(\vec{x},\vec{x}';\theta)=\sigma^2\prod_{i=1}^{d}k_i(x_i-x_i';\theta_i)$$

- One dimensional Functions
  - Squared Exponential:

$$k_i(x_i-x_i')=e^{-\left(\frac{x_i-x_i'}{\theta_i}\right)^2}$$

  - Matern Function $\nu=\frac{3}{2}$:

$$k_i(x_i-x_i')=\left(1+\sqrt{3}\left|\frac{x_i-x_i'}{\theta_i}\right|\right)e^{-\sqrt{3}\left|\frac{x_i-x_i'}{\theta_i}\right|}$$

  - Cubic Spline 2:

$$k_i(x_i-x_i')=\begin{cases}1-6\left|\frac{x_i-x_i'}{\theta_i}\right|^2+6\left|\frac{x_i-x_i'}{\theta_i}\right|^3 & \text{for } 0\le\left|\frac{x_i-x_i'}{\theta_i}\right|\le 0.5\\ 2\left(1-\left|\frac{x_i-x_i'}{\theta_i}\right|\right)^3 & \text{for } 0.5\le\left|\frac{x_i-x_i'}{\theta_i}\right|\le 1\\ 0 & \text{for } \left|\frac{x_i-x_i'}{\theta_i}\right|\ge 1\end{cases}$$

GEUK Details

- Covariance Parameters determined via Maximum Likelihood:

$$\log(p(y|X;\theta)) = -\frac{1}{2}[Y^T \delta Y^T]\underline{K}^{-1}\begin{bmatrix} Y \\ \delta Y \end{bmatrix} + \frac{1}{2}[Y^T \delta Y^T]\underline{C}\begin{bmatrix} Y \\ \delta Y \end{bmatrix} - \frac{1}{2}log|P| - \frac{1}{2}log|M| - \frac{1}{2}log|\underline{A}| - \frac{nd+n-s}{2}log 2\pi$$

- Optimization carried out with standard tools (L-BFGS/Active Set Algorithms)
- Most Expensive and Problematic part of Surrogate Construction
  - Optimization problem scales with dimension
  - Covariance Matrix inversion $O(n^3 d^3)$ if dense
  - Improvements possible with sparse covariance and better optimization algorithm

GEUK Details

- Previously developed Hermite Polynomials used as basis (Roderick and Anitescu,2010)
- Basis set is truncated based on sensitivity analysis (High order used for most sensitive parameters)
- Derivatives included in Basis to reduce number of required samples
- Parameters assumed to follow GP:

$$\hat{\beta} = \left([H^T G^T] \underline{K}^{-1} \begin{bmatrix} H \\ G \end{bmatrix} \right)^{-1} [H^T G^T] \underline{K}^{-1} \begin{bmatrix} Y \\ \delta Y \end{bmatrix}$$

- Function predictions:

$$y_* | \vec{X}, Y, \delta Y = [k_*^T w_*^T] \underline{K}^{-1} \begin{bmatrix} Y \\ \delta Y \end{bmatrix} + \left(h(\vec{x}_*) - [k_*^T w_*^T] \underline{K}^{-1} \begin{bmatrix} H \\ G \end{bmatrix} \right) \hat{\beta}$$

Comparison with Regression:

- Eliminates discrepancy between Regression model and sample points
- Improved Assimilation of Extra Data
- Stochastic Model allows for prediction distributions

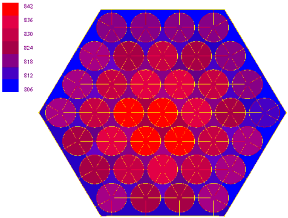
Comparison with Ordinary Kriging:

- Approximately Same Cost
- Beneficial for smooth design space
- Requires no addition samples when truncated basis used

Results for Nuclear Engineering Applications

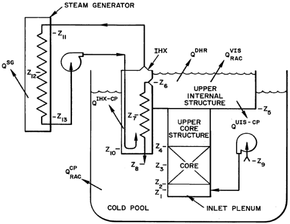
Matlab Model:

- 3D Steady-State FV Reactor Core with heat/neutron diffusion
- Properties of Sodium-cooled fast Reactor
- 12 Input parameters relating to heat transfer and thermodynamics
- Peak Fuel Pin Temperature used as Output
- Written specifically to produce derivatives w.r.t input parameters
- Comparison with 500 validation points



MATWS:

- Subset of Industrial code SAS4A/SASSYS-1 code
- Point Kinetics reactor core model with heat removal model
- Four input parameters (Expansion and Doppler feedback coefficients)
- Peak Fuel Pin Temperature in time used as Output
- Derivatives w.r.t inputs calculated using OpenAD
- Comparison with 1000 validation points

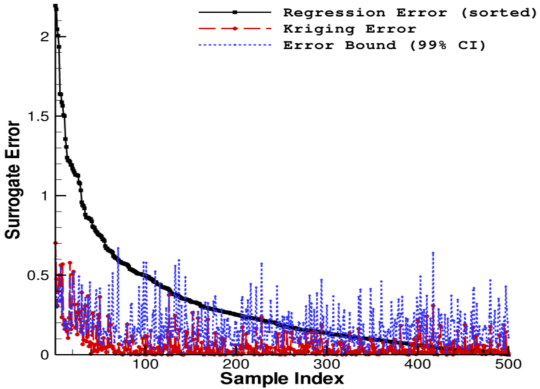


Investigation Points

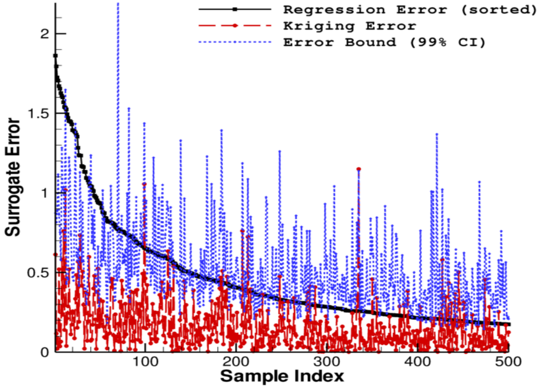
- P1: GEUK results in less error compared with L_2 regression (i.e. assumption of i.i.d. samples)
- P2: GEUK results in less error compared with UK. Decrease in error should be greater than cost required to obtain derivatives.
- P3: GEUK results in less error than Ordinary Kriging for same number of samples (at least for problems well suited for regression).
- P4: The variance of the model provides a good estimate of the distribution of error in the surrogate.
- P5: The choice of covariance function greatly affects mean and variance predictions. Best to pick least differentiable.

P1: Kriging vs. Regression

Matlab Point-wise Error Comparison (N=8)



MATWS Point-wise Error Comparison (N=16)



P1: Kriging vs. Regression

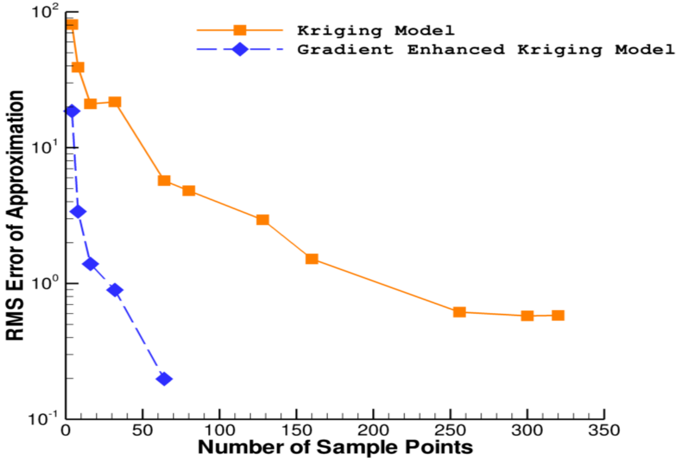
- MATWS Global Error Comparison

Training Points	GEUK RMS	REG RMS	GEUK Max	REG Max
4 (p=2)	3.6433	15.2304	13.7491	56.6632
6 (p=2)	0.5260	3.2833	2.2040	14.0380
8 (p=3,trunc)	0.1841	0.5695	1.1980	3.1272
16	0.0766	0.427	0.747	2.404
24	0.0887	0.405	0.910	1.877
32	0.0995	0.309	1.118	1.959
40	0.0517	0.295	0.437	2.112
50	0.0508	0.251	0.386	1.476
100	0.0337	0.181	0.0998	1.068

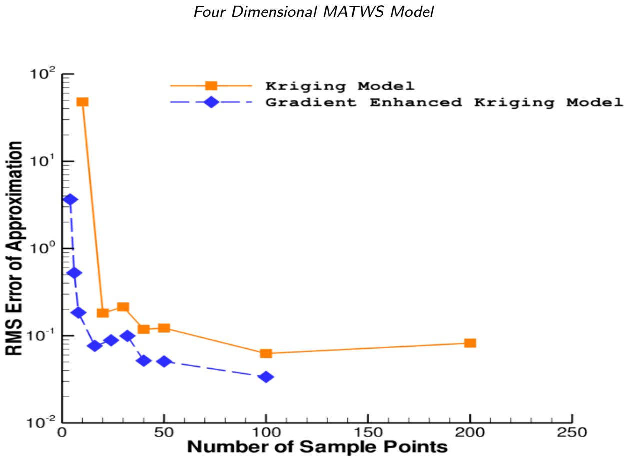
- *P1: GEUK outperforms regression for all sample points*

P2: GEUK vs. UK

Twelve Dimensional MATLAB Model



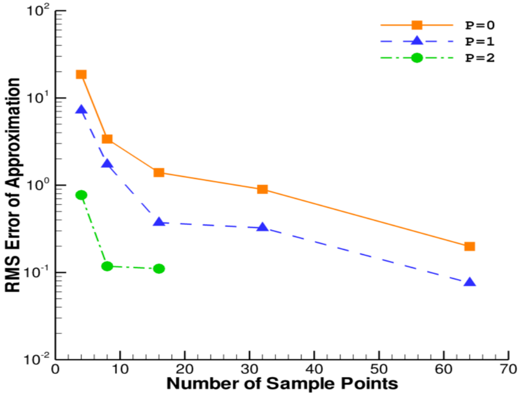
P2: GEUK vs. UK



P2: GEUK requires fewer samples for similar error level

P3: Effect of Regression Order

Regression Order comparison for Matlab Model



P3: Effect of Regression Order

- MATWS Global Error comparison with Ordinary Kriging

Training Points	GEUK RMS	OGEK RMS	GEUK Max	OGEK Max
4 (p=2)	3.6433	0.5729	13.749	3.9739
6 (p=2)	0.5260	0.2585	2.2040	1.8793
8	0.1841	0.1931	1.1980	1.7142
16	0.0766	0.0968	0.7474	1.2039
24	0.0887	0.0845	0.9104	1.0363
32	0.0995	0.0863	1.1181	1.1478
40	0.0517	0.0551	0.4368	0.7108
50	0.0508	0.0498	0.3862	0.6151

- *P3: Higher order regression provides benefit for relatively smooth design space and low sampling density*

P4: Distribution of Validation Data

- Distribution of Validation data used to assess Kriging variance prediction
- For small training sets, fraction of validation data within standard deviation of mean can be used to assess performance:

Matlab Model-Distribution of Validation Data (Independent Data)

Training Points	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
8	0.690	0.882	0.950
6	0.378	0.568	0.668
4	0.362	0.626	0.720

MATWS Model-Distribution of Validation Data (Independent Data)

Data Set	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
16	0.697	0.942	1.000
24	0.533	0.857	0.971
32	0.525	0.817	0.942
40	0.475	0.800	0.937
50	0.366	0.685	0.870
100	0.221	0.424	0.600

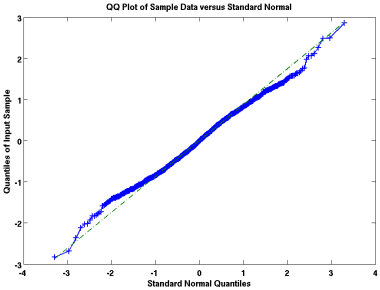
P4: Distribution of Validation Data

- For Larger data sets, samples should be decorrelated:

$$Z = K^{-1/2}(\widehat{J} - m(x)).$$

- Z should follow standard Normal

QQ plot for MATWS Z score (N=50)



P4: Variance provides bound for error in surrogate provided surrogate is "accurate enough"

P5: Choice of Covariance Function

- Matlab Model - Comparison of Error (N=8)

Covariance Function	RMS Error	Max Error
Cubic Spline 1	0.57759	2.2239
Cubic Spline 2	0.26542	1.5721
Squared Exponential	0.11554	0.70207
Matern-3/2	0.33149	1.7272
Matern-5/2	0.20259	1.1576

- Matlab Model - Comparison of Distribution (N=8)

Covariance Function	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
Cubic Spline 1	0.290	0.592	0.758
Cubic Spline 2	0.776	0.878	0.930
Squared Exponential	0.690	0.882	0.95
Matern-3/2	0.676	0.874	0.932
Matern-5/2	0.704	0.884	0.928

- MATWS Model - Comparison of Error (N=50)

Covariance Function	RMS Error	Max Error
Cubic Spline 1	0.0567	0.3963
Cubic Spline 2	0.0528	0.4551
Squared Exponential	0.1487	2.0268
Matern-3/2	0.0398	0.2552
Matern-5/2	0.0749	0.7991

- *P5: Proper Covariance Function dependent on problem and sample density (Matern with $\nu = 3/2$ safe bet)*

Quantile Estimation

- Surrogate must be able to accurately predict quantiles of interest
- For Nuclear Engineering, 95th percentile with 95% confidence particularly important
- "Conservative Estimate" can be made using Uniform distribution and properties of Order Statistics

Table: MATWS – cubic spline 2 – Quantile calculation for MATWS data

Sample Points	Kriging Estimate	Regression Estimate	Training Estimate
4 (p=2)	865.73	864.78	863.55
6 (p=2)	865.86	871.15	863.55
8	866.08	866.60	863.46
16	865.89	866.51	865.45
24	865.83	866.49	865.56
32	865.87	866.32	865.76
40	865.82	866.37	865.86
50	865.83	866.42	865.86

Actual Value = 866.16 Data Range=[859.15, 869.66]

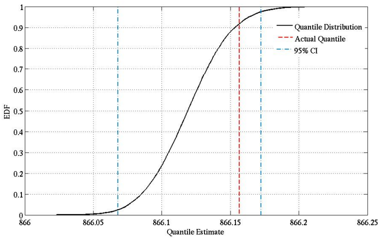
Quantile Estimation

- Because GEUK model is stochastic, samples can be repeatedly extracted:

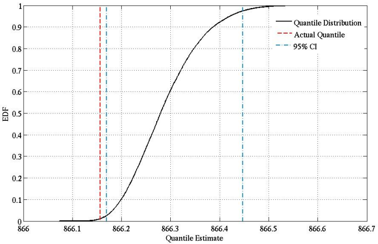
$$\widetilde{J} = m(x) + K^{1/2}u$$

- Leads to a distribution Quantile predictions
- EDF of Quantile prediction can be used to assign confidence interval

MATWS Model N=50



MATWS Model N=8



Summary

Conclusions:

- The Gradient-Enhanced Universal Kriging model consistently created a more accurate surrogate than regression alone.
- For smooth problems, GEUK may provide benefits over Ordinary Kriging.
- Stochastic Nature of Kriging gives distribution for statistic predictions, giving confidence intervals.

Future Work:

- Extension to Higher dimension (and incorporation of dimension reduction techniques)
- Investigation of alternative basis functions
- Application to more complex design space
- Surrogate Construction with imperfect data (Partially converged/Noisy function or derivatives)

Acknowledgments

DOE CSGF

Argonne NATIONAL LABORATORY