Gradient Enhanced Universal Kriging Model for Uncertainty Quantification in Reactor Safety Simulations

Brian Lockwood
Department of Mechanical Engineering
University of Wyoming
Cheyenne, Wyoming

Mihai Anitescu
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL, 60439
Email: anitescu@mcs.anl.gov

***Abstract*—The increased resolution needs of high accuracy codes applied on novel physical phenomena or configurations require very large computing times even on advanced architectures. In turn this makes the use of regular tools for uncertainty quantification very problematic since multiple example runs are needed to construct a statistical model for the response of the system and the total number of runs is limited by the computational budget. While multi-resolution approaches are expected to alleviate this problem, the issue of constructing reasonable models of the uncertainty propagation step out of very few samples remains. To this end, we study the ability of a Universal Gradient Enhanced Kriging Model to provide an accurate model for uncertainty propagation out of very few samples. For this model, the mean behavior of the surrogate is determined by a polynomial regression and deviations from this mean are represented as a gaussian process. Preliminary tests with nuclear engineering data show the Universal Gradient Enhanced Kriging Model provides a more accurate surrogate model when compared to either regression or ordinary Kriging models. We demonstrate that our approach produces good uncertainty models even for the case where less than 10 sample points are considered for nuclear engineering applications with 12 uncertainty parameters.**

I. Introduction

In this work, we are concerned in propagation an uncertainty set, $\mathcal{U} \in \mathbb{R}^{n_u}$ through an expensive software code, which outputs the analysis function $f(u)$. Since a sample is very expensive, we cannot expect to use direct simulation in order to propagate the uncertainty. Instead, we propose to use a few samples u_i and the function and *gradient* information of $f(u)$ in order to reduce the computational effort required by the evaluation of samples outcome, in order to create an approximate model of $\tilde{f}(u)$ and an estimate of its error in approximating $f(u)$ in a statistical sense. The overall uncertainty assessment, that includes both initial uncertainty and small sample uncertainty propagation effects is then obtained by interrogating the inexpensive approximate model $\tilde{f}(u)$. This approach is based on the observation that automatic differentiation tools can provide the adjoint parametric derivative of sophisticated codes with an effort that is a small multiple of the function evaluation [3] and, many times, as we have experienced in our test case here, with an effort that is essentially equal to it. Since the gradient provides n_u time more information about f, this has the potential of drastically reducing the number of samples for the same accuracy level. We have observed this fact in a polynomial regression with derivative information approach that we have recently introduced [7]. On the other hand that approach did not provide an error estimate, something that we remediate with this work. Our main inquiry is this: can uncertainty models of scientific computing codes be constructed out of a very small number of samples, say O(10)? To that end, we investigate an extension of our regression approach, where the polynomial regression model becomes the mean of a Gaussian Process model. This has the potential of identifying correlations in uncertainty space and further reducing the propagation uncertainty, $\tilde{f}(u) - f(u)$.

II. Uncertainty Propagation with Derivative Info using Gaussian Processes

We will use an uncertainty model, where we assume that the response of the system can be represented as a Gaussian process with explicit mean function and specified covariance function governed by a set of parameters (here after known as hyperparameters). This assumption can be represented functionally as: $\hat{y} = N(h(\vec{x})\beta, K(\vec{x}, \vec{x}; \theta))$, where $h(\vec{x})$ is a column vector containing the basis functions of the regression (or, equivalently, mean function) evaluated at the point $\vec{x}$ and β are the regression parameters. The parameters β and θ can be determined from data points by using a maximum likelihood approach [10], [6]. Once the parameters of the model are determined, model predictions with uncertainty throughout the domain are subsequently determined by sampling from the conditional distribution $y_*|\vec{X}, Y$ where $\vec{X}, Y$ are the training data [6]. This approach is called kriging, and it originates in geostatisics applications [1] where it is used as a technique for interpolation of spatial fields. Kriging is a very versatile approach for representing functional responses with uncertainty and has been used in robotics applications [11], climate [4], atmospheric chemistry [2], composite materials [8]. Several other application areas are presented in the monograph [6].

Gradient information may be included in the Gaussian process model and the covariance matrix becomes a block

matrix which includes the covariance between derivative observations and function observations. This block matrix can be represented as [9]:

$$\underline{K} = \begin{bmatrix} cov(Y,Y) & cov(Y,\nabla Y) \\ cov(\nabla Y,Y) & cov(\nabla Y,\nabla Y) \end{bmatrix} \quad (1)$$

where $cov(Y,Y)$ represents a covariance matrix between the function values at the training points, $cov(\nabla Y,Y)$ is a covariance matrix between gradient components and the function values at the training points and $cov(\nabla Y,\nabla Y)$ is a covariance matrix between gradient components. For clarity, if n is the number of training points and d is the dimension of the problem, then $cov(Y,Y)$ is an $n \times n$ matrix, $cov(\nabla Y,Y)$ is size $nd \times n$ and $cov(\nabla Y,\nabla Y)$ is $nd \times nd$. The total size of $\underline{K}$ is thus $(d+1)n \times (d+1)n$. The components of the matrix $cov(Y,Y)$ are given by $cov(y,y') = k(\vec{x},\vec{x}')$ [5].

The covariance function between function values and gradient functions then becomes $cov(\frac{\partial y}{\partial x_k}, y') = \frac{\partial}{\partial x_k} k(\vec{x},\vec{x}')$. Differentiating once more (now w.r.t to the second argument of the covariance function) gives the covariance between gradient components: $cov(\frac{\partial y}{\partial x_k}, \frac{\partial y'}{\partial x_l}) = \frac{\partial^2}{\partial x_k \partial x_l} k(\vec{x},\vec{x}')$. The gradient vector and hessian matrix resulting from equations can then be arranged into the matrices $cov(\nabla Y,Y)$ and $cov(\nabla Y,\nabla Y)$ respectively.

The improved model is extended to include derivative values by using the block-wise definition of the covariance matrix outlined in the previous section and extending the regression matrix to include derivatives of the basis functions. First, the definition for the regression parameters must be updated to include derivatives. This is done by using the block-wise definition of the covariance matrix as well as including basis function derivatives in the regression matrix. The formula for the regression parameters is now given by:

$$\hat{\beta} = \left([H^T G^T] \underline{K}^{-1} \begin{bmatrix} H \\ G \end{bmatrix} \right)^{-1} [H^T G^T] \underline{K}^{-1} \begin{bmatrix} Y \\ \delta Y \end{bmatrix} \quad (2)$$

where G is a matrix containing the derivatives of the basis functions and δY is a vector of gradient observations. Using this new definition of regression parameters, the mean value of the model can be predicted as [9]:

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

where w is the vector of covariances between the test point function values and the gradient at each training point and $\underline{R}$ is a Schur complement involving the covariance matrix . The variance associated with this prediction can now be calculated as:

$$V[y_*] = cov(\vec{x}_*,\vec{x}_*) - [k_*^T w_*^T] \underline{K}^{-1} \begin{bmatrix} k_* \\ w_* \end{bmatrix} + \underline{R}(\vec{x}_*) \underline{A}^{-1} \underline{R}(\vec{x}_*)^T \quad (3)$$

where $\underline{A}$ and $\underline{R}$ are now defined in the block sense presented previously.

A. Hyperparameter Fitting

With the functional form of the covariance function specified, only the hyperparameters need to be determined in order to fully specify the covariance function. For this work, the hyperparameters are determined by maximizing the marginal likelihood function for the data given the functional form we have assumed. Its formula can be extended to include gradient observations by treating K as a block matrix and including derivative observations in the quadratic terms and the regression matrix: $log(p(y|X;\theta)) = -\frac{1}{2}Y^T K^{-1} Y + \frac{1}{2}Y^T C Y - \frac{1}{2}log|K| - \frac{1}{2}log|A| - \frac{n-s}{2}log2\pi$, where $C = K^{-1} H A^{-1} H^T K^{-1}$ [5]. Here, A is the regression matrix ($A = H^T K^{-1} H$) and s is the number of terms in the mean model. The model can be easily modified to account for the gradient information. The optimization is carried out using a combination of L-BFGS and an active set algorithm. The L-BFGS algorithm is used to get within the vicinity of an extrema. At this point, the active set algorithm is used to further hone in on the maximum and ensure the maximum is global and not just local. To further avoid local maxima, a fixed number of initial starting points (~ 5) are tested. Despite these mitigating practices, the selection of local minimum represents a problem for this method of determining hyperparameters. The local minimum problem creates problems with repeatability of results. The local minimum problem seems to be a bigger problem for models built with gradient observations (GEK model) then traditional Kriging methods.

III. Results for Simulation Data

In order to assess the utility of the Gradient Enhanced Universal Kriging (GEK) Model for representing the uncertainty space relating to engineering simulations and predicting the error associated with regression, data derived from thermo-hydraulic simulations of a sodium cooled fast reactor was used. The simulation problem is defined by twelve parameters relating to the thermodynamic and heat transfer relations present in the reactor. The main output of interest for this simulation is the peak fuel pin temperature. The model and previous work that approximated the uncertainty propagation via regression with Hermite polynomials is described in [7]. In order to improve upon this regression based representation, the gradient enhanced Kriging model was applied to the data. The performance of the Kriging model was compared to other approaches for uncertainty propagation, such as regression or ordinary Kriging. Additionally, the ability of the Kriging model to predict the propagation uncertainty by means of the variance in the Kriging model was tested.

For these tests, a set of 500 peak fuel temperature values were generated using the original code. For the tests, the polynomial order in regression and mean function as well as the number of training points were varied. Additionally, for one of the tests error was intentionally introduced into the model training by basing the model on partially converged function results. This test was performed to assess the ability of both models to capture the true uncertainty space with imprecise training data, as well as test the ability of the

TABLE I
COMPARISON OF ERROR FOR KRIGING AND REGRESSION MODELS

Data Set	GEK RMS	Regr. RMS	GEK Max	Regr. Max
1	0.11554	0.47118	0.70207	2.194
2	0.58351	0.76058	2.5731	3.2553
3	0.77163	1.1982	3.2202	4.8668
4	0.77163	1.289	3.2204	5.0067

TABLE II
STATISTICS FOR KRIGING PREDICTION

Data Set	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
1	0.690	0.882	0.950
2	0.378	0.568	0.668
3	0.362	0.626	0.720
4	0.362	0.626	0.720

TABLE III
COMPARISON OF ERROR BETWEEN COVARIANCE FUNCTIONS

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

TABLE IV
COMPARISON OF DATA DISTRIBUTION BETWEEN COVARIANCE FUNCTIONS

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

Kriging model to predict the higher error associated with this model. The accuracy of the models was determined based on comparing model predictions to the 500 function values using the RMS error and max difference as measures of the error.

In order to gauge the relative performance of the various uncertainty propagation aproaches, such as GEK and regression-based approach, the max error and RMS error of the the approaches models for several data sets was measured and compared. For these tests, GEK and regression used the same training data and basis functions (for the mean function in case of GEK). To assess the overall performance of the GEK versus the regression model, the performance was tested on one or more of four scenarios.These were 1) a maximum third order polynomial based on 8 training points, 2) a maximum second order polynomial based on 6 points, 3) a second order polynomial based on 4 points and 4) a second order polynomial based on 4 training points with partially converged function values.

A. GEK versus Regression

The maximum difference between model and actual function values as well as RMS error between model and actual values are found in table I for the four different scenarios. In addition to observing the effect on mean prediction accuracy, the ability of the variance to provide a confidence interval on the result was assessed. To assess this quality of the model, the fraction of validation points within 1,2 and 3 standard deviations of the mean prediction was examined. Ideally, these fractions should correspond to a normal distribution with 68% falling within the first standard deviation, 95% within the second and 99% within three standard deviations. The results for each scenario are given in Table II. For these tests, the square exponential covariance function was used.

In the context of simulation, where the cost is dominated by acquiring the training data, creation of the regression data and Kriging model have essentially the same cost. For all scenarios tested, the GEK approach outperformed the regression based model. For the third order regression case the normal approximation is an excellent one. For the other cases the normal distribution is not a very good one (the tails are fatter) but the 3σ intervals contain indeed a large proportion of the data. It should be pointed out, however, that the last case contains 2 and 4 sampling points, an exceedingly small number. Experimentation with other kernels is in progress the other 3 data sets, in order to determine the influence of the choice on the kernel on the results.

To assess the role played by choice of covariance function, a Kriging model for the third order regression test case was created with each of the five different covariance functions considered for this work. Table III shows the RMS error and Max error for each covariance function and Table IV contains the distribution of the data around the mean prediction.

As these results show, the choice of covariance function can dramatically change the accuracy of the model predictions. As expected from the tests using explicit functions, the squared exponential gives the most accurate mean predictions. Unexpectedly, the squared exponential also gives a distribution of the data which is closest to a normal distribution, although the Matern functions and the Cubic Spline 2 covariance also give reasonable data distributions. Again, these results imply that the distributions predicted from the assumption of the Gaussian Process become more valid as the accuracy of the model increases.

B. Error Prediction

Finally, the error associated with the original regression was predicted using the Kriging model. This was first performed using the 4 scenarios outlined previously. In order to assess the prediction of the error, the point-wise actual error was plotted along with the mean and worst case error predictions. In order to provide a quantitative measure of the prediction, the correlation between the mean error and worst case error predictions with the true error was calculated for each scenario. As a final measure, the fraction of validation points encompassed by the worst case error prediction was calculated in order to determine the validity of the error bound. Because the worst case error prediction is based on the 99% confidence interval of the Kriging prediction, this last measure is equivalent to the fraction of data points whose actual function value lie within

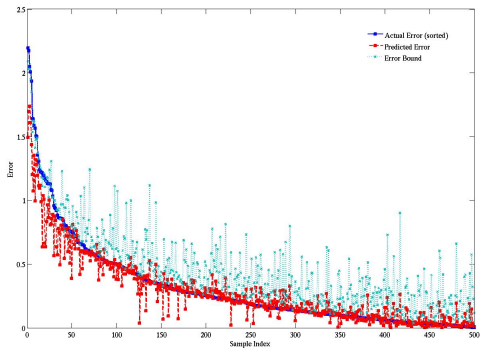

Fig. 1. Point-wise error prediction with actual error for third order regression based on 8 training points

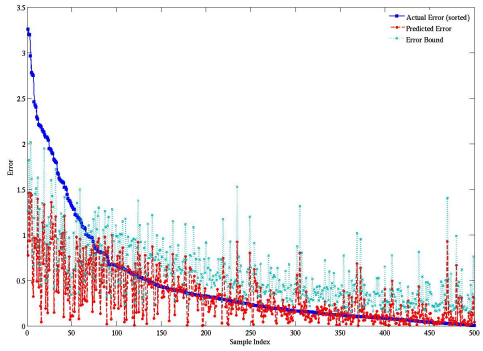

Fig. 2. Point-wise error prediction with actual error for second order regression based on 6 training points

three standard deviations of the Kriging prediction. For these tests, the square exponential covariance function was used.

For the third order regression case, the error prediction closely matches the actual error and the worst-case error estimate just bounds the actual error in the domain. This result is unsurprising as the error in the Kriging approximation of the function is low. For the other test cases, the error prediction is less accurate and worst-case error bounds only a portion of the actual error. For all the test cases, the error prediction is positively correlated with the actual error and a relatively large fraction of the data is encompassed by the error bound. Hence, even for cases in which the Kriging model still contains significant approximation error, the error prediction and bound still provides some information as to the distribution of the error throughout the domain and a rough estimate as to level of error present in the regression model. In cases in which

TABLE V
CORRELATION BETWEEN ACTUAL ERROR AND PREDICTED ERROR

Data Set	Corr. for Mean	Corr. for Worst-case	Fraction Covered
1	0.965	0.913	0.950
2	0.592	0.630	0.668
3	0.848	0.782	0.720
4	0.864	0.822	0.720

TABLE VI
CORRELATION BETWEEN ACTUAL ERROR AND PREDICTED ERROR FOR EACH COVARIANCE FUNCTION

Cov. Fcn.	Corr. Mean	Corr. Worst-case	Fraction Covered
Cubic Spline 1	-0.217	0.050	0.758
Cubic Spline 2	0.791	0.666	0.930
Squared Exponential	0.965	0.913	0.950
Matern-3/2	0.594	0.493	0.932
Matern-5/2	0.859	0.737	0.928

the number of training points is limited by computational budget, the Kriging model error predictor may represent the only possible test for the accuracy of the regression model.

Because the accuracy of the mean predictions and the magnitude of the variance are sensitive to the choice of covariance function, the error prediction for the third order regression was calculated using the five different covariance functions in this work. The correlation with the actual error and the fraction of validation points within the error bound can be found in Table III-B. We see that the squared exponential and the Matern functions correlate far better the predicted error with the actual error.

IV. CONCLUSIONS AND FURTHER WORK

In conclusion, the Universal Gradient Enhanced Kriging (GEK) model provides accurate representations of a function space with limited number of function samples. It combines the strenghts of both regression and Kriging models: it exhibits fast convergences as the polynomial order is increased and provides unbiased estimates of function behavior as the number of training points is increased. In addition to function predictions, the variance of the Kriging model provides a reasonable confidence bounds for the prediction. For all cases tested, the Kriging model provided an error prediction which was positively correlated with the actual error in the domain. The correlation was greatest for scenarios in which the Kriging model represented a significant accuracy improvement over the regression model. In addition, we find that, as in the regression case that has preceded this work [7], the gradient information greatly reduces the number of samples and computational effort needed at the same accuracy level. Future work will include issues in propagation and approximation of adjoint derivative information in novel architectures, multi-model/multi-resolution uncertainty estimates, and assessment of the accuracy of uncertainty representation of GEK approaches for a larger spectrum of test codes.

ACKNOWLEDGMENT

This work was supported by the US DOE CSGF Fellowship Program (Brian Lockwood) and the U.S. Department of Energy through contract DE- AC02-06CH11357.

REFERENCES

[1] J.P. Chiles and P. Delfiner. *Geostatistics: modeling spatial uncertainty*. Wiley, New York, 1999.

[2] B.P. Eynon. Statistical analysis of precipitation chemistry measurements over the eastern United States. II. Kriging analysis of regional patterns and trends. *J. Appl. Meteorol.;(United States)*, 27, 1988.

[3] L. Hascoët and B. Dauvergne. Adjoints of large simulation codes through automatic differentiation. *REMN Revue Européenne de Mécanique Numérique / European Journal of Computational Mechanics*, 17:63–86, 2008.

[4] G. Hudson and H. Wackernagel. Mapping temperature using kriging with external drift: theory and an example from Scotland. *International Journal of Climatology*, 14(1):77–91, 1994.

[5] C. E. Rasmussen and C. K. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, 2006.

[6] C.E. Rasmussen and C. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, Massachusets., 2006.

[7] Oleg Roderick, Mihai Anitescu, and Paul Fischer. Polynomial regression approaches using derivative information for uncertainty quantification. *Nuclear Science and Engineering*, 164(2):122–139, February 2010.

[8] S. Sakata, F. Ashida, and M. Zako. Kriging-based approximate stochastic homogenization analysis for composite materials. *Computer Methods in Applied Mechanics and Engineering*, 197(21-24):1953–1964, 2008.

[9] E. Solak, R. Murray-Smith, W.E. Leithead, D.J. Leith, and C.E. Rasmussen. Derivative observations in gaussian process models of dynamical systems. *Advances in Neural Information Processing Systems*, pages 1057–1064, 2003.

[10] M.L. Stein. *Interpolation of spatial data: Some theory for Kriging*. Springer-Verlag, Berlin, 1999.

[11] S. Vijayakumar, A. D'souza, T. Shibata, J. Conradt, and S. Schaal. Statistical learning for humanoid robots. *Autonomous Robots*, 12(1):55–69, 2002.