A Multivariate Interpolation and Regression Enhanced Kriging Surrogate Model

Komahan Boopathy *
University of Dayton, Ohio, 45469, USA

Markus P. Rumpfkeil †
University of Dayton, Ohio, 45469, USA

We present a Kriging surrogate model that is enhanced with a Multivariate Interpolation and Regression (MIR) through a dynamic training point selection. We propose an adaptive training point selection strategy where MIR is used as a local surrogate model that guides the construction of the global Kriging surrogate model. The quality of the resulting MIR enhanced Kriging surrogate model is demonstrated for two-, five- and nine-dimensional analytic test functions. The results indicate that the model performs better than currently available Kriging surrogates as well as a previously enhanced Kriging surrogate that uses Dutch Intrapolation as a local surrogate model. Preliminary results of using variable-fidelity data in the construction of the MIR enhanced Kriging surrogate model are also presented and show promise.

Nomenclature

a_{vi}, a_{gi}	Basis functions for MIR	N_{loc}	Number of training points for local model
β	Wave number	N_t	Number of nodes in M-dimensional Cartesian mesh
f	Objective function	n	Taylor order (order of accuracy)
$\tilde{f}$	Approximated function value	n_e	Order of extrapolation
∇f	Gradient of objective function	n_i	Order of interpolation
γ	Magnitude of weights	P	Polynomial exactness parameter
M	Number of design variables	R_{dcok}	Correlation matrix
N	Number of training points		
x	Design variable		

I. Introduction and Motivation

Non-availability of analytic solutions to many engineering problems and the potential difficulty as well as high expenses in performing laboratory experiments have forced many researchers to increasingly rely upon computational simulations. Consequently, computational methods are extensively used in engineering research, also owing to the increasing computational power and sophistication of numerical algorithms. Despite the advancements made in computer hardware and the deployment of high performance computing (HPC), there exists an acute imbalance between the requirements and availability of computational power, especially when dealing with high-fidelity physics-based simulations. For example, a relatively straightforward gradient-based airfoil shape optimization requires many optimization iterations and hence flow solutions: the entire flow field needs to be solved for at each design iteration and the gradient also needs to be computed either through finite-difference or adjoint techniques,[1–3] which in total can require hundreds of flow solutions, potentially demanding enormous computational time and storage. With the relatively meager computational power at hand, the research community has to trade-off accuracy versus computational time or limit their design spaces in scope and freedom, which either way may lead to inefficient designs. In order to curb the predicaments involved in high-fidelity simulations through their computational burden, the idea of a surrogate model was introduced, that is to replace expensive function evaluations with an

*Graduate Student, Dept. of Mechanical and Aerospace Eng., boopathyk1@udayton.edu, Student Member AIAA
†Assistant Professor, Dept. of Mechanical and Aerospace Eng., Markus.Rumpfkeil@udayton.edu, Senior Member AIAA

approximate but inexpensive functional representation. Thus, when a surrogate model for an optimization problem is constructed with given training data, the most promising locations in the model can be explored with lesser computational cost. The accuracy of surrogate models can be increased efficiently by adding the exact function information in the most promising locations. This approach can save a lot of computational cost and enables the exploration of wider design spaces efficiently. A surrogate model being an approximate representation of the exact function space, has some error associated with it. A lot of today's surrogate model research is directed on improving the accuracy of existing models as well as developing versatile and robust new surrogates.[4–8]

A surrogate model and its corresponding functional approximation can be enhanced using various approaches such as,

1. determining training point locations adaptively rather than randomly,
2. finding multiple approximations for same quantity of interest and judicially deciding the best approximation (global multiple surrogate),
3. incorporating higher-order information in the construction of the surrogate such as gradients and Hessians,
4. using variable-fidelity data.

In this paper, we retain the usage of higher order information (gradients) and compare the effect of the distribution of the training points with that of Latin hypercube sampling.[9] We also demonstrate the usage of variable fidelity data on the accuracy of the surrogate.

I.A. Overview of Surrogate Models

In Peter and Marcelet[10] the performances of major surrogate models such as least square polynomials, multi-layer perceptron, radial basis function (RBF), and Kriging are compared for a two-dimensional turbomachinery problem, where the Kriging and RBF models show the best performance. Yet another comparative study[11] of Kriging, Datascape, and Second Order Regression (SOR) show that Kriging performs the best in terms of accuracy. Kriging yields accurate results especially when the training data used to build the models is sparse; when larger training sets are used Datascape produces more accurate models, but the latter is computationally more expensive than the other two methods. Besides the Kriging being better than many other current models available, the model predicts the function value by using stochastic processes, and has the flexibility to represent multi-modal functions. Thus, the Kriging model, originally developed in the field of geological statistics, has gained a lot of popularity.[6,7,12–19] In Zhao and Xui[20] a comparison of four surrogate modeling techniques (multivariate polynomial method, RBF, Kriging, and Bayesian neural networks) is carried out. The models are ranked based on accuracy, confidence, robustness, and efficiency. The results indicate that multivariate polynomial methods perform best among the four models tested. In Wang *et al.*[21,22] a Multivariate Interpolation and Regression (MIR) scheme is developed. It features remarkably high order of convergence, and is robust for a variety of node distributions. But in terms of computational time MIR is very expensive.[21,22] The existence of a polynomial based method (MIR) and a stochastic method (Kriging) which both show great promises (accuracy and versatility, respectively), and their respective weaknesses (computational cost and accuracy, respectively) have influenced us to try to exploit their respective advantages and mitigate their disadvantages by combining the two methods to build a better surrogate model.

Two popular strategies that one can adopted in the context of multiple surrogate methods are as follows:

- Calculate a mean function prediction using the predictions of two or more surrogates.
- Build local surrogate models and exploit the available information to guide the construction of a global surrogate.

Studies in Zhao *et al.*[20,23] show that when compared with the traditional approach of global meta-modeling using only a single metamodel, employing a local surrogate can improve the modeling accuracy considerably. Since there are several surrogate modeling techniques available, each having their own merits and demerits, the selection of the "best surrogate" to perform each of the two tasks, can be a difficult one.

In this paper, we do not calculate multiple "global" approximations since we are aiming to keep the computational cost in building the surrogate low, however, we do calculate multiple "local" approximations and use the gained information to aid in the selection of additional surrogate training points.

I.B. Higher-Order Information

An efficient gradient evaluation method based on adjoint formulations[1,3] is available, and has been adopted by the computational community for data-assimilation and design optimization problems over the last several decades. Thus, the introduction of gradient information within surrogate models as additional training data has attracted a lot of attention as well. For example, gradient enhanced Kriging models have been developed and have shown very beneficial results.[8,14,17,18] While adjoint methods provide an effective approach for computing first-order sensitivity derivatives, the ability to compute second-order sensitivity derivatives is also highly desirable for many science and engineering simulation problems.[24–28] For example, the availability of Hessian information allows the use of much stronger Newton optimization strategies, which holds the potential for greatly reducing the expense of solving difficult optimization problems. Since an efficient Hessian evaluation method has been developed by one of the authors,[29,30] it is also very promising to utilize the Hessian information within surrogate models in addition to the gradient information.[8,31] The observation is that for high-fidelity applications targeting a single output objective, the effort for computing the full gradient is comparable to the effort of computing the objective function itself; thanks to the adjoint formulation. Therefore, as the number of inputs, M, increases, using the output function and its derivative information is appealing, because it provides $M + 1$ pieces of information for roughly the cost of two function evaluations. Similarly, the Hessian provides $M \cdot (M + 1)/2$ pieces of information for roughly the cost of M function evaluations, since, in general, the most efficient full Hessian constructions require the solution of M forward linear problems (one corresponding to each input parameter).[24,27] Thus, one can reasonably expect to have to compute the output function overall far fewer times to obtain a good surrogate model when using gradient and Hessian information and this should also scale more reasonably to higher dimensions. However, we restrict ourselves to gradient enhancement only in this paper for simplicity.

Outline of the paper

The remainder of this paper is organized as follows. In Section II, we review some important concepts related to Kriging, Dutch Intrapolation and Multivariate Interpolation and Regression, and also some popular training point selection methods available in the literature. Section III, discusses our adaptive training point selection strategy that we propose in this work. Section IV demonstrates the quality of our improved surrogate model for two-, five- and nine-dimensional analytic test functions by comparing it to the performance of a standard Kriging model as well as a previously enhanced version[32] of Kriging that uses Dutch Intrapolation as local surrogate. We also extend our discussion to the advantage of variable-fidelity data in the building of the MIR enhanced Kriging surrogate model. Section V concludes this paper and outlines future research directions.

II. Literature Review

In this section, we provide a short overview of training point selection and surrogate models - Kriging, Dutch Intrapolation as well as MIR.

II.A. Training Point Selection

The location and number of training points used to construct any surrogate model has a significant effect on its accuracy. Thus, we briefly review some training point selection strategies and outline the merits and demerits associated with them. Training point selection approaches can be broadly classified into domain-based and response-based approaches.[33] In domain-based approaches, training points are chosen based on the information available from the design space (e.g. distance between two training points), whereas in response-based approaches, the training points are chosen based on the information provided by the surrogate model (e.g. mean squared error approach). The latter was developed to enhance the efficiency of the sampling process by using information from the existing metamodel. For example, in the response-based approach the

user could monitor the progress of the model and choose to stop or extend the sampling process. Domain-based sampling is based on space-filling concepts that try to fill the design space evenly with training points. It is, in general, not possible for the user to select the number of training points apriori to ensure a given accuracy, due to the non-linearity of most functions of interest. We now review some important domain- and response-based approaches from the literature.

Domain-based Approaches

MONTE CARLO: Monte Carlo (MC) techniques[34] are the simplest of all sampling methods. Here, a random number generator is used to select training point locations in the design space. A major drawback of MC is the fact that for a small amount of training points large areas of the design space may be left unexplored while others may be sampled densely.[35–37]

LATIN HYPERCUBE: Latin hypercube sampling (LHS) was proposed by McKay *et al.*[9] for designing computer experiments as an alternative to MC sampling techniques. The basic idea is to divide the range of each design variable into N bins of equal probability, which yields N^M bins in the design space, where M is the dimension of the problem. Subsequently, N training points are generated for each design variable such that no two values lie in the same bin (as shown in the left of Figure 1). The LHS algorithm generates training points in a box-shaped domain as follows,[36]

$$x_j^{(i)} = \frac{\pi_j^{(i)} + U_j^{(i)}}{N}, \quad \forall \quad 1 \leq j \leq M, \quad 1 \leq i \leq N \tag{1}$$

where $x_j^{(i)}$ is the jth-component of the ith-training point, $U \in [0, 1]$ is an uniform random number, and π is an independent random permutation of the sequence of integers $0, 1, \ldots, N-1$.

In the right of Figure 1, one can notice the random fluctuations in root mean square error ($RMSE$) of a generic surrogate model based on LHS. In spite of increasing the number of training points, the $RMSE$ does not necessarily decrease, because all these points are picked at random. Thus, a superior strategy for training point selection is required to ensure that the $RMSE$ will reduce when one increases the number of training points.

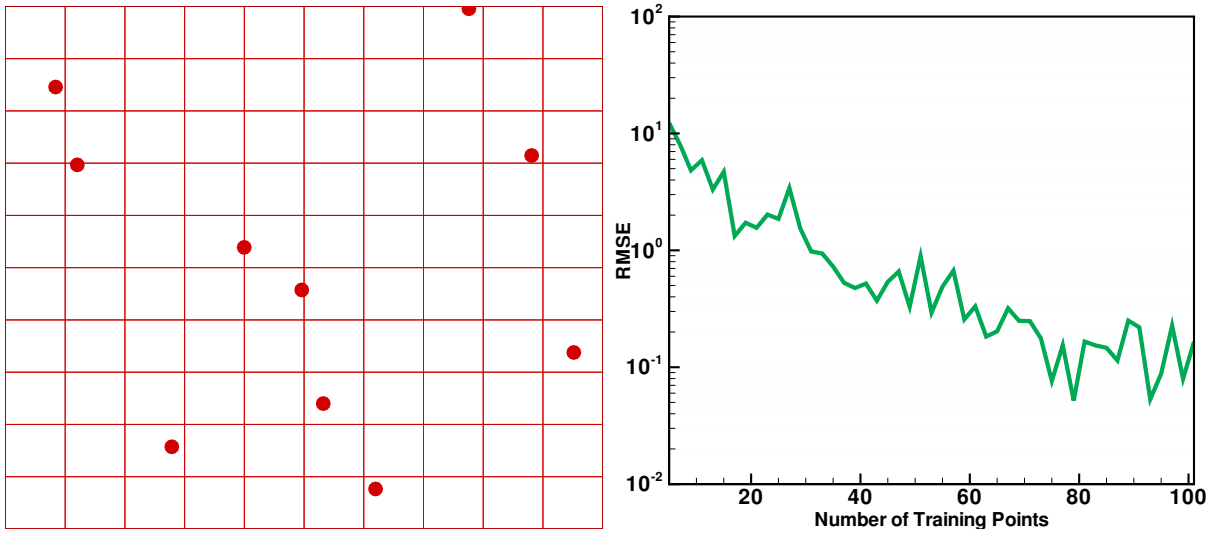


Figure 1: Left: An example of LHS for a two-dimensional design space. Right: A typical convergence history of a generic surrogate model using LHS or MC.

DELAUNAY TRIANGULATION: Delaunay triangulation is a geometrical method of training point selection, where the design space is divided into hyper-triangles and the training points are chosen at some geometrical significant location such as the centers of the hyper-triangles and midpoints of the edges as shown in Figure 2. A major drawback of the Delaunay triangulation is that it does not scale well to higher dimensions.[32]

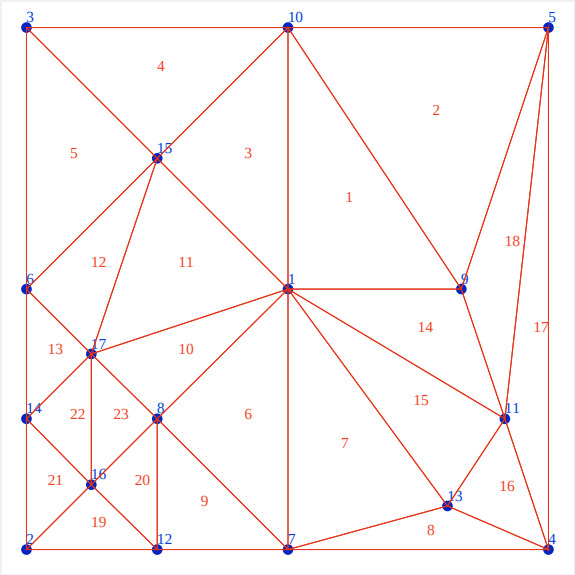


Figure 2: A Delaunay triangulation schematic is shown here. Points numbered 1 to 5 are the initial training points and 12 points have been added subsequently by splitting the design space into triangles.

Response-based Approaches

A brief overview of some response-based approaches is given in the following.

- Jones and Schonlau[12] proposed a sequential response surface methodology which starts with a smaller number of training points and adds additional training points at locations where the standard error is high. During this process, the training point set is updated, the metamodel is reconstructed, and the process of choosing new additional training points continued until the expected improvement from new training points has become sufficiently small.

- Alexandrov[38] proposed a metamodel management framework using a trust-region method for updating metamodels according to the improvement of the objective function during an optimization procedure.

- Messac *et al.*[39] developed a new methodology to quantify the surrogate error in different regions of the design space, which is called the Regional Error Estimation of Surrogate (REES) method. The REES method provides a model independent error measure that does not require any additional function evaluations. It works roughly as follows: after segregating the design space into sub-spaces (or regions) Variation of the Error with Sample Points (VESP) regression models are constructed to predict the accuracy of the surrogate in each subspace. These regression models are trained by the errors (the mean and the maximum error) evaluated for the intermediate surrogates in an iterative process. At each iteration, the intermediate surrogate is constructed using different subsets of training points and tested over the remaining points. Their results indicate that the REES measure is capable of evaluating the regional performance of a surrogate with reasonable accuracy and that one could use this error estimate to guide the surrogate-building process.

- Rosenbaum *et al.*[40] applied an adaptive sampling strategy where the samples are generated sequentially as well. At every stage of the adaptation process, a surrogate model is generated and assessed in order to find a new training point location at which the objective function is evaluated.

- Recently, a dynamic training point selection approach developed by one of the authors[32] has proven to yield better results than other domain-based strategies, where a local surrogate (Dutch Intrapolation) has been employed to guide the training point selection process. First, an initial number of training points are chosen via LHS and more training points are added in locations with the largest discrepancy between the function predictions of the Kriging and the local surrogate model. Thus, instead of just specifying the number of training points in the beginning and picking the points randomly, the model

is built by adding more training points in promising regions of the design space. This is similar to the concept of expected improvement[12,41] (EI) when optimizing with a Kriging model where a potential for improvement is used, which considers both estimated function values and uncertainties in the surrogate model, thereby keeping the balance between global and local search performance.

For a comprehensive review on training point selection strategies the reader is referred to Keane and Nair,[36] Arora,[33] and Forrester *et al.*[37]

II.B. Surrogate Model Review

In the following paragraphs, short accounts on Kriging, Dutch Intrapolation and Multivariate Interpolation and Regression are provided.

II.B.1. Kriging

The Kriging model was originally developed in the field of geological statistics by the South-African mining engineer *Danie G. Krige*.[6] Kriging was introduced in engineering design following the work of Sacks *et al*[42] and has been increasingly used in aerospace engineering and design.[14–17] Kriging predicts the function value by using stochastic processes and has the flexibility to represent multi-modal and non-linear functions. The basic formulation of Kriging is given as,[33]

$$\tilde{f}(x) = \mu + Z(x) \tag{2}$$

where $\tilde{f}$ is the approximated function value, μ models the mean behavior using a regression model, and $Z(x)$ models the local variation from the mean behavior using a Gaussian process with zero mean $E[Z(x)] = 0$.

A gradient enhanced direct as well as indirect Kriging model has previously been developed and has shown very beneficial results.[8,19,31] In the direct co-Kriging approach, the covariances between function values, function values and gradients, as well as gradients have to be considered within the correlation matrix as opposed to the original Kriging formulation where only the covariances between function values have to be considered. Thus, the correlation matrix R_{dcok} becomes asymmetric and its size increases to $N \cdot (M+1)$ (from N), where N is the number of training points, and M represents the number of input parameters. On the other hand, the formulation of the indirect co-Kriging model is exactly the same as that of the original Kriging model. In this approach, additional training points are constructed around a real training point by using a Taylor series extrapolation. Direct and indirect co-Kriging models produce identical results in the limit of small real to extrapolated point step sizes, although the direct approach is preferable due to its robustness and lack of tunable parameters. For more mathematical background on Kriging the reader is referred to previously published papers.[8,19,31]

II.B.2. Dutch Intrapolation

One can construct a local response surface using a hybrid of extrapolation and interpolation involving a few, already existing, training points x_i, $i = 1 \ldots, N_{loc}$, where N_{loc} is the number of training points used to construct the local response surface. The function values and available derivatives at each training point are used to construct extrapolated function values of order n_e for a test candidate location, x. The extrapolations from the training points are then weighted with a low-order interpolant of order n_i to find a unique function value $\tilde{f}(x)$. This approach has been coined Dutch Intrapolation[43,44] (DI) and it has been shown that the order of accuracy of the intrapolant is equal to $n_i + n_e$, that is, using function, gradient, and Hessian information for the extrapolations and second-order interpolation leads to a fourth-order accurate intrapolant. This is demonstrated in Figure 3 where the two-dimensional Rosenbrock function (a fourth-order polynomial) is represented exactly using function, gradient, and Hessian information in six training points chosen randomly using LHS. The Dutch extrapolation functions are normal multivariate Taylor expansions of order n_e with a correction term given in multi-index notation by[44]

$$\mathcal{T}^{n_e n_i}(x, x_i) = \sum_{|k|\geq 0}^{|k|\leq n_e} \frac{a_k^{n_e n_i}}{k!}(x - x_i)^k \nabla^k f(x_i) \quad \text{for } i = 1, \ldots, N_{loc} \quad \text{with } a_k^{n_e n_i} = \binom{n_e + n_i}{n_e}^{-1} \binom{n_e - k + n_i}{n_e - k} \tag{3}$$

where $f(x_i)$ is the objective function value at training point x_i. It is important to note that although the Dutch Taylor expansions are discussed here for general order n_e, practical applications are usually restricted

to low values of n_e. The range of practical applicability is similar to that of "normal" Taylor expansions. High-order Taylor expansions are often used in theoretical formulations, however, in practical applications their use is limited because the convergence with increasing order is typically very slow, and the region of convergence is very small. Thus, the Dutch Taylor expansions are to be used in small regions where the function to be approximated is well represented by a low-order polynomial, that is, where the Taylor expansion coefficients decrease quickly for increasing order. In addition, it becomes impractical to calculate higher-order derivatives of the objective function for high-fidelity physics-based simulations.

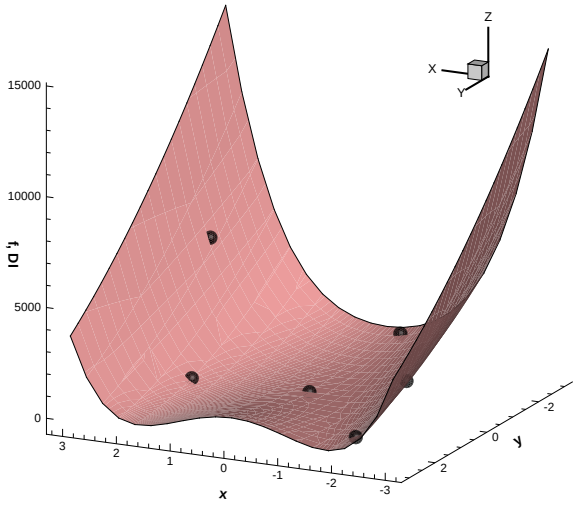


Figure 3: Comparison between two-dimensional Rosenbrock function and Dutch Intrapolation. The six spheres are the training point locations with function, gradient and Hessian information.

II.B.3. Multivariate Interpolation and Regression

Multivariate Interpolation and Regression (MIR) is a surrogate model where each data point is represented as a Taylor series expansion, and the higher order derivatives in the Taylor series are treated as random variables. The approximation coefficients are then chosen such that they minimize the objective function in each point by solving an equality constrained least squares problem. The approximation is an interpolation when the data points are given as exact, or a non-linear regression function if non-zero measurement errors are associated with the data points. Mathematically, the objective function, f, in an M-dimensional design space is approximated as[21,22]

$$\tilde{f} = \sum_{i=1}^{N_v} a_{vi}(x)\hat{f}(x_{vi}) + \sum_{i=1}^{N_g} a_{gi}(x)\nabla\hat{f}(x_{gi}) \tag{4}$$

where N_v is the number of training data points at which the function is evaluated and N_g is the number of gradient data points (if used), a_{vi} and a_{gi} are the basis functions of the resulting approximate polynomial, and $\hat{f}$ and $\nabla\hat{f}$ are the function f and gradient values ∇f added with their corresponding measurement errors σ_{vi} and σ_{gi}. The tunable parameters in MIR are the Taylor order n, the wave number β, magnitude of the weights γ and polynomial exactness parameter P. We would expect to improve the accuracy of the model by using a higher n, but round-off errors originating in the solution of the least squares problem propagate to the approximate function via equation (4).

Thus, higher n do not always guarantee an improved approximation. This is further explained in Wang *et al.*[21,22] The polynomial order finds the highest order polynomial for which the approximation $\tilde{f}$ is exact. It can be set by the user when higher order approximations are necessary. Other parameter of the scheme namely the wave number β and magnitude γ of the weights are not set by the user explicitly, but are calculated automatically and hence the reader is referred to Wang *et al.*[21,22] for details of the algorithm. Although other parameters of the approximation scheme can be computed automatically from the data points, the choice of an optimum Taylor order n is a tedious task and it varies from function to function and with the dimensionality of the problem as well.

III. Adaptive Training Point Selection

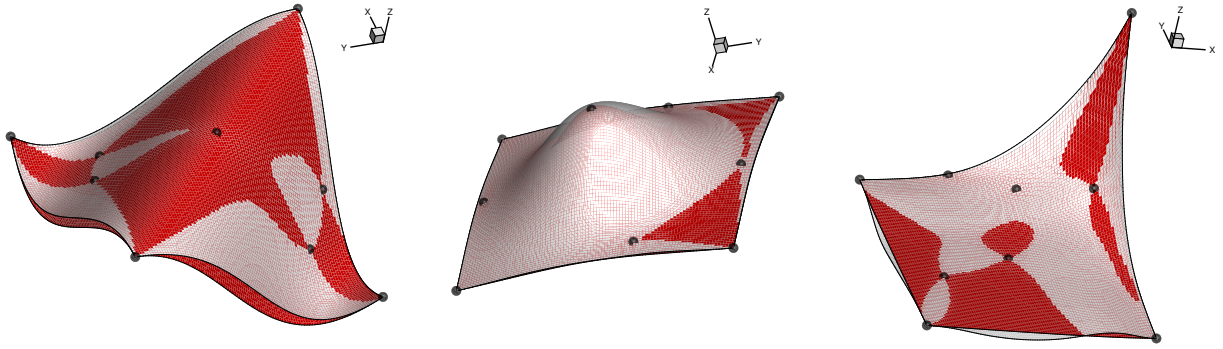
This section compares the performance of MIR to that of the original Kriging for some analytical test functions, followed by a discussion of our idea to use MIR as local guidance for choosing Kriging training points.

III.A. Analytic Test Functions

The following are the analytic test functions used in this paper on a hypercube $[-2, 2]^M$ for evaluation purposes:

1. A multi-dimensional Cosine function: $f_1(x_1, \ldots, x_M) = \cos(x_1 + \ldots + x_M)$
2. The multi-dimensional Runge function: $f_2(x_1, \ldots, x_M) = \frac{1}{1+x_1^2+\ldots+x_M^2}$
3. A multi-dimensional Exponential function: $f_3(x_1, \ldots, x_M) = e^{(x_1+\ldots+x_M)}$

A two-dimensional plot for each of the above test function is shown in Figure 4.



Cosine Runge Exponential

Figure 4: Analytic test functions in two-dimensions (red) together with a Kriging surrogate (white) constructed from nine training points.

The root mean square error ($RMSE$) between the actual f and approximated function values $\tilde{f}$ calculated on an M-dimensional Cartesian mesh with N_t total nodes is given by,

$$RMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(f_i - \tilde{f}_i)^2}. \tag{5}$$

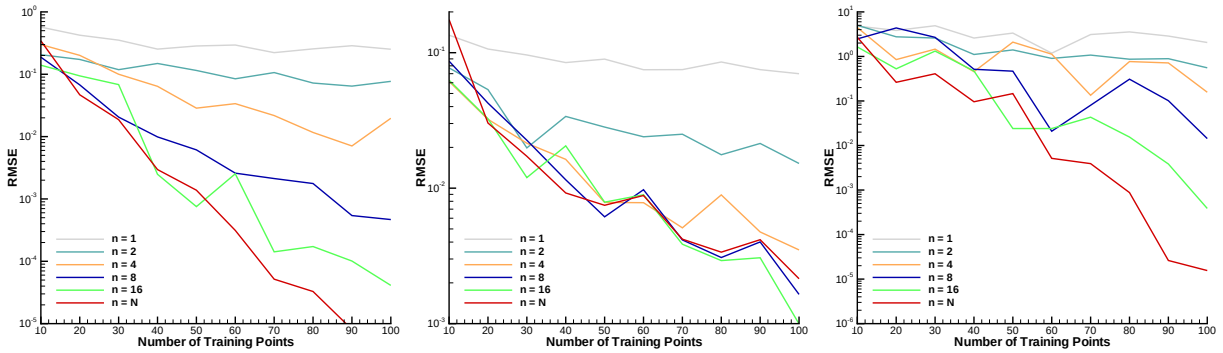
where f_i and $\tilde{f}_i$ are the actual and approximated function values in the sample point locations x_i, $i = 1 \ldots, N_t$, respectively. $RMSE$ is used throughout this paper as a measure of goodness.

III.B. Effect of Taylor Order on MIR

Figure 5 shows the effect of different Taylor orders on the accuracy of the MIR function value $\tilde{f}$. $RMSE$ is calculated on a Cartesian mesh with 101×101 nodes. It can be seen that lower n such as 1 or 2 yield a less accurate approximation whereas higher n usually produce a more accurate approximation, but they also demand much more computational time.[21,22]

III.C. RMSE Comparison between Kriging and MIR

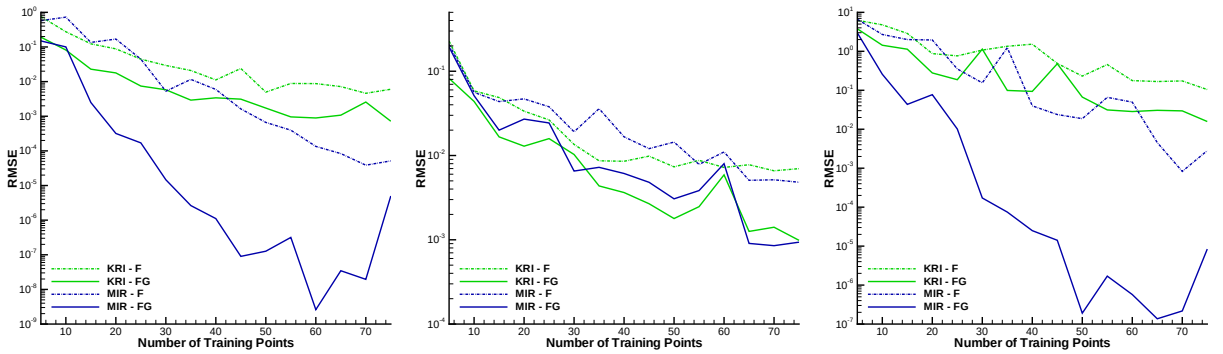
Figure 6 compares original Kriging (green lines) and MIR (blue lines) for the two-dimensional cosine, Runge, and exponential test functions. $RMSE$ is again calculated using 101×101 nodes. One can see that MIR approximates the two-dimensional cosine and exponential test functions better than the original Kriging surrogate model, however, the Kriging is performing slightly better for the Runge function. It can also be noted that the addition of gradient information (labeled FG, continuous lines) shows better results than the



Cosine Runge Exponential

Figure 5: $RMSE$ between the two-dimensional test functions and the MIR approximation vs. number of training points plotted for different Taylor orders, n. All training points are selected through LHS.

approximation with function values alone (labeled F, dotted lines). For this experiment, the Taylor order is set equal to the number of training points (N) when only function values are used and it is set to $(3 \cdot N)$ when both function and gradient information is used. A detailed comparison of MIR with other surrogate modeling methods and higher-dimensional test functions are reported in Wang *et al.*[21,22]



Cosine Runge Exponential

Figure 6: $RMSE$ between the two-dimensional test functions and the surrogate models vs. number of training points which are selected through LHS.

III.D. Summary of Comparisons

In spite of nice convergence properties as shown in the previous subsection, the main drawback of MIR is that it is computationally expensive and it scales poorly to higher dimensions and to more training points. Additionally, there is no good way to estimate the optimal Taylor order that is shown to have a significant impact on the results. A higher Taylor order can corrupt the solution and it is also computationally expensive. For example, from Figure 5 one can observe that for the Runge function a Taylor order of four exhibits almost the same performance as higher Taylor orders. However, MIR has a great potential to serve as a local surrogate as it is much cheaper as well as still accurate for a small number of training points. We exploit the information available from local MIR surrogates to guide the global training point selection process and we use a lower Taylor order in this work to maintain a balance between computation and accuracy.

In summary, the reasons for using MIR as a local rather than a global surrogate model are as follows:

1. The associated computational burden with MIR is very high[21,22]

2. The choice of an optimum Taylor order, n, is a tedious task and it varies from function to function as well as with the dimensionality of the problem

3. Kriging supports the usage of both high- and low-fidelity training points[45–48]
4. Kriging has the capability to represent multi-modal functions and can effectively capture non-polynomial functions[8]
5. The currently available version of MIR has severe memory restrictions for higher dimensions

A detailed account on our proposed adaptive training point selection method is provided next.

III.E. Our Adaptive Training Point Selection Framework

Figure 7 shows a schematic diagram of our MIR enhanced Kriging surrogate construction approach.

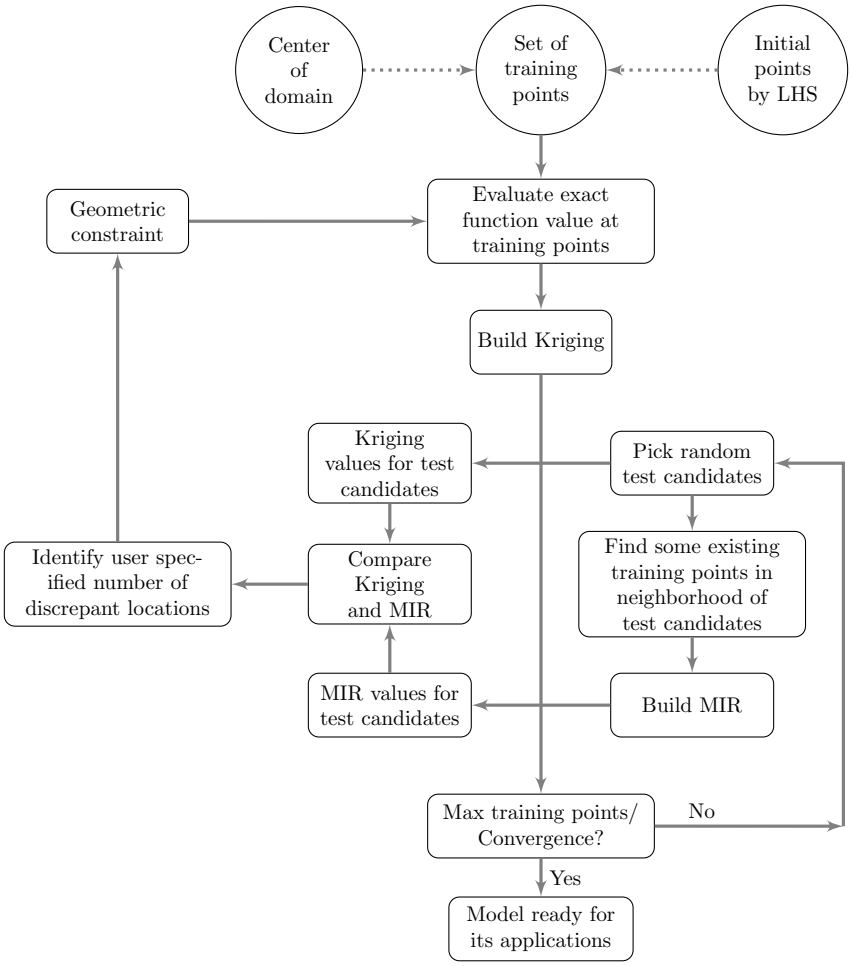


Figure 7: Our Kriging surrogate construction algorithm using MIR as local surrogate

The steps involved in the process are explained below.

- Start by evaluating the function value (also gradients, if desired) at the center of the domain.
- Then pick a user specified additional number of training points via LHS and evaluate their function (and gradient) value.
- Then repeat the following steps until convergence or until a maximum amount of function (and gradient) evaluations has been reached. (We define convergence as having the worst discrepancy below a certain threshold.)
    1. Specify a set of test candidates via LHS.

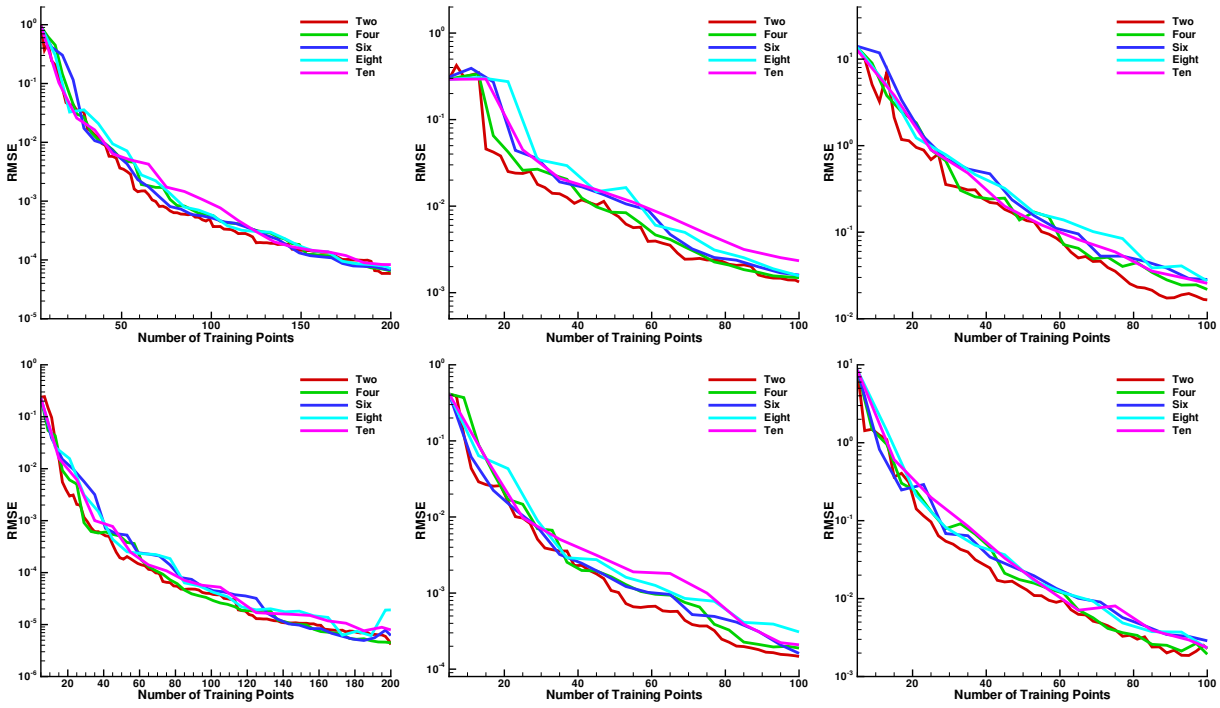
2. Construct a local function value for each test candidate using MIR involving an appropriate number of closest neighbors with function (and gradient) information.
3. Compare the global Kriging surrogate model function value predictions for the test candidates with the local MIR predictions.
4. Add a user-specified number of test candidates with the worst discrepancy between the two predicted values to the set of training points and evaluate the real function (and gradient) at these discrepant points and rebuild the Kriging surrogate.

- We also augment the selection process by geometric criteria, for example, we make sure that the distance of a test candidate to the nearest existing training point is above the average distance between all test candidates to their respective closest training point. This ensures that the training points are not clustered in one particular region and are sparse in other regions of the design space.

IV. Discussion of Results

IV.A. Avoidance of Training Point Clustering

In our approach, a user specified number of training points are added to the set of training points and the exact function (and gradient) is evaluated at those locations at each iteration. In Figure 8, we show the effect of number of training points chosen per cycle on the accuracy of the surrogate for all three analytic test functions in two-dimensions, with and without gradient information. Overall it can be observed that the surrogate is more accurate when only two points are added per cycle. However, adding only two points per cycle implies more computational burden since the Kriging has to be constructed more often to reach a fixed number of training points compared to adding more than two training points per cycle.

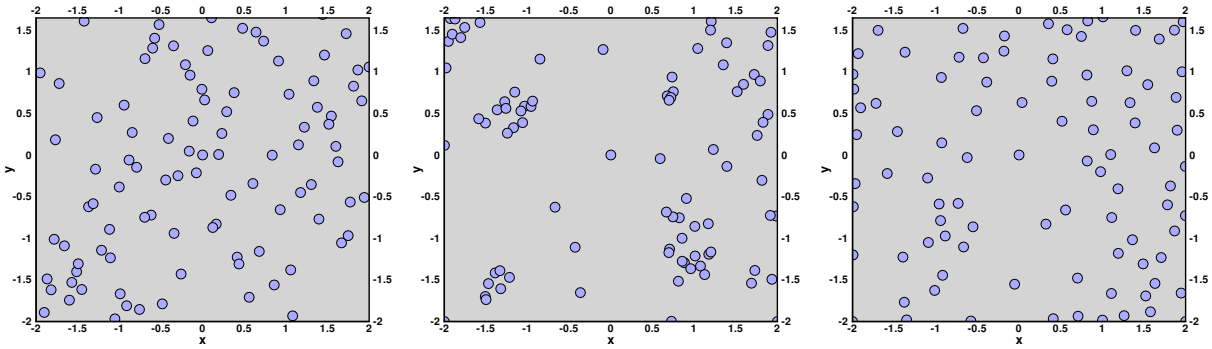


Cosine Runge Exponential

Figure 8: A comparison of number of training points added per cycle for two-dimensional test cases. Top : function values only, Bottom: function and gradient values.

Since the general behavior is preserved for all the tested cases (two, four, six, eight and ten), one may accelerate the convergence by adding more training points per cycle. But care must be taken to incorporate appropriate geometric constraints to prevent clustering of training points (see Figure 9) that can harm

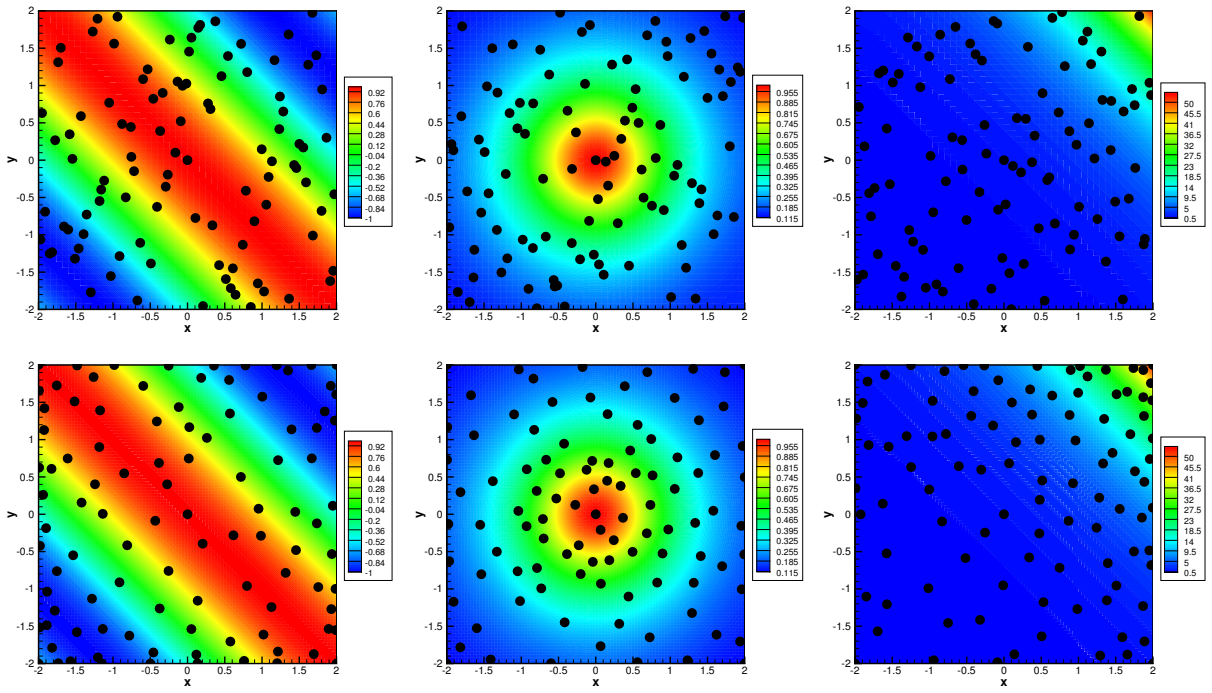
the convergence by ill-conditioning the correlation matrix. In our studies we observed that, whenever two or more training points lie very close to each other, fluctuations tend to occur in an otherwise monotone convergence. Attributing lack of enough training points as a possible cause for fluctuation, we repeated the experiments with a densely sampled domain and carefully observed when fluctuations occurred and identified "proximity" as the main cause. Hence it is very important to prevent proximity of training points by enforcing a geometrical constraint.



Latin hypercube | Adaptive-no constraint | Adaptive-with constraint

Figure 9: Training point distributions to demonstrate the need for a distance-based geometrical constraint. Total number of points is 105 and twenty points were added per cycle using the exponential test function, f_3, as an example.

In Figure 10, red and green regions characterize a larger change in the function value compared to other regions in the domain and the blue regions have a near zero gradient. One can clearly observe that our adaptive training point selection strategy chooses training points more efficiently than LHS by concentrating them in regions where the gradient is high. This behavior can save a lot of computational time when applied



Cosine | Runge | Exponential

Figure 10: Training point distributions ($N = 101$) for two-dimensional test functions shaded by their contours. Top: Latin hypercube sampling. Bottom: Our strategy.

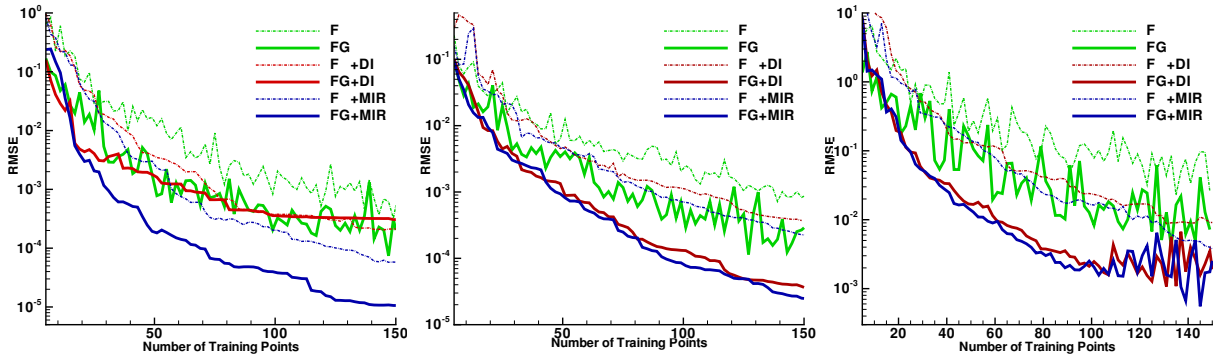
to high fidelity simulations by reducing the number of required function evaluations for an accurate surrogate model. Moreover, in terms of "space filling efficiency", our method is better than LHS as well. Evidently, in the top row of Figure 10 (which corresponds to LHS results) there are larger unsampled regions in the domain than in the bottom row, where the training points are more spread throughout the domain. Hence our surrogate model can capture the overall functional behavior better and locate local and/or global extrema more accurately.

IV.B. RMSE Comparisons

In this subsection we show plots of the $RMSE$ between the Kriging surrogate and the exact function versus the number of training points (with or without gradient information) used to construct the surrogate models for all three analytic test functions given in Subsection III.A. The $RMSE$ between the actual and approximate function are calculated on a Cartesian mesh with 10201, $1,889,568$ and $1,953,125$ total number of nodes for two-, five- and nine-dimensional test cases, respectively. The training points always include the center of the domain and the others are either all selected through LHS or we start with a user specified number of LHS training points and add additional points through dynamic training point selection as described in the Subsection III.E.

IV.B.1. Two-dimensional Results

From Figure 11 one can see that our enhanced Kriging (shown as blue lines) shows better results than other surrogate approaches namely, the original Kriging (shown as green lines) and the Kriging enhanced with DI (shown as red lines). In case of the two-dimensional cosine function one can see that with about thirty



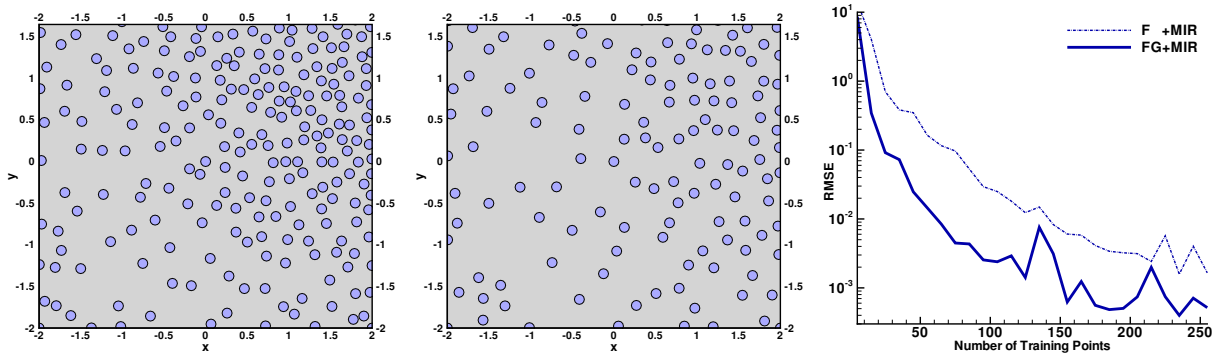
Cosine Runge Exponential

Figure 11: $RMSE$ versus the number of training points in two dimensions. Green, red, and blue lines show the performance of original Kriging, dynamically sampled Kriging with DI, and our new Kriging model, respectively.

training points with function and gradient information (FG), our model is able to outperform the previously enhanced Kriging with DI which requires about 150 function and gradient evaluations, to obtain the same level of accuracy. Furthermore, the addition of gradient information did not always produce a more accurate model in the case of Kriging with DI, whereas in our new model gradient information always improves the accuracy. For the Runge function, we see a significant improvement compared to the original Kriging but only a slightly better approximation when compared to the previously enhanced Kriging with DI. A similar observation applies to the exponential test function, f_3 as well.

Unfortunately, both DI and MIR based adaptive models incur fluctuations for the gradient cases (FG) of f_3, when the number of training points increase over 100. But the function only cases (F) of f_3 which have the same training point distribution (shown in Figure 12) are still showing monotone behavior. The gradient cases see an earlier onset of fluctuations because now the Kriging has to accommodate two potentially very different gradient values from two relatively close training points which actually causes numerical instabilities. Now it becomes an open question of when one should use higher-order information in the construction of the surrogate. Generally, based on our observations, when training points are concentrated in some parts of the domain, it is advisable to refrain from using higher-order information in those regions, whereas it is

advantageous to use higher-order information in sparse regions of the domain. Figures 12a and 12b show typical training point distributions when fluctuations are occurring in the convergence of f_3, for function (F) and gradient (FG) cases respectively.



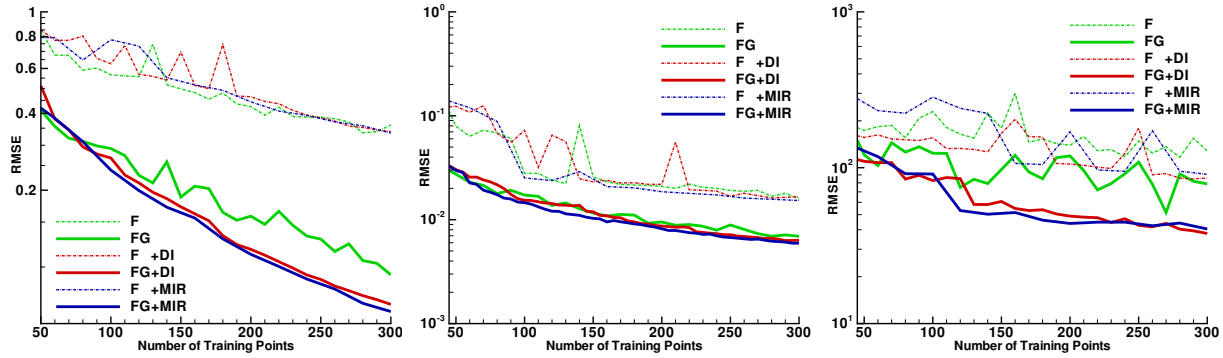
(a) $N = 255$ (b) $N = 155$ (c) $RMSE$ vs. N

Figure 12: Study of onset of fluctuations in convergence for f_3

It can also be observed that the Runge and exponential test cases have a single mode in the domain, whereas the cosine test function has multiple modes in the domain, as explained by the contours of Figure 10. Due to the presence of multiple potential good locations to add the training points (cosine test function), the Kriging surrogate features a remarkably good convergence behavior. Hence our enhanced approach can be a good choice for approximating multimodal functions and their corresponding surrogate based optimization and/or uncertainty quantification will be more accurate.

IV.B.2. Five-dimensional Results

Figure 13 shows the performance of our developed method for five-dimensional test functions. An improved convergence behavior can be noticed for all three test cases with fewer fluctuations compared to other approaches.

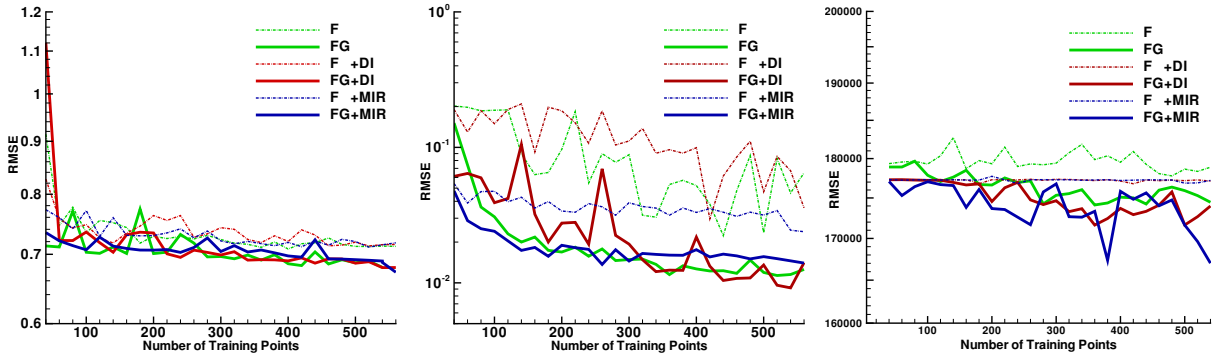


Cosine Runge Exponential

Figure 13: $RMSE$ versus the number of training points in five dimensions. Green, red, and blue lines show the performance of original Kriging, dynamically sampled Kriging with DI, and our new Kriging model, respectively.

IV.B.3. Nine-dimensional Results

Results for nine-dimensional test cases are shown in Figure 14. No consistent behavior can be seen as the surrogate suffers from the "curse of dimensionality". But it can be noticed that the addition of gradient information improves the approximation greatly. Since efficient Hessian calculation methods are available,[29,30] their use can hopefully improve the convergence even more.



Cosine Runge Exponential

Figure 14: $RMSE$ versus the number of training points in nine dimensions. Green, red, and blue lines show the performance of original Kriging, dynamically sampled Kriging with DI, and our new Kriging model, respectively.

IV.B.4. Summary of Observations

Overall, one can clearly see that all gradient (FG) enhanced Kriging models (shown as continuous lines) perform much better than the models that are only based on function evaluations (F) (shown as dotted lines), for all functions, approaches (original Kriging, Kriging with DI, and Kriging with MIR), and dimensions tested in this paper. It can also be inferred that the dynamic training point selection performs better than just selecting all training points through LHS. The dynamic training point selection also helps to reduce the effect due to randomness that more training points do not necessarily lead to a more accurate surrogate model. Another important feature that can be generally observed is that the Kriging with MIR tends to have a better convergence rate than the other two approaches. Thus, our new method provides improved approximations compared to other approaches, irrespective of whether the training points are with or without gradient information.

IV.C. Variable-fidelity Results

As can be inferred from Figures 11, 13 and 14, the number of training points needed by the surrogate increases exponentially with the dimension of the problem. As each training point is, in general, computationally expensive to obtain in physics-based simulations, we seek to show the potential advantage of multi-fidelity modeling (variable-fidelity) in this subsection since lower-fidelity data can usually be obtained computationally cheaper (e.g. coarser mesh data, Euler solution instead of Navier-Stokes solution, etc.). For our test cases, we artificially perturb the high-fidelity test functions f_1, f_2, and f_3 given in Subsection III.A as f_{L1}, f_{L2}, and f_{L3}, where $f_{Li} = 0.1 \cdot f_i$, $i = 1, \ldots, 3$. Since the variable-fidelity model involves some high-fidelity data as well, for ease of comparison, we adopt the following method. If N_h and N_l are the number of high- and low-fidelity training points in the variable-fidelity surrogate, respectively, the $RMSE$ is compared with that of the regular Kriging model constructed with N_H training points

$$N_H = N_h + \frac{N_l}{K} \tag{6}$$

where K is a constant representing savings in computational time. Though there is hardly any difference in computational time for analytic function evaluations (high- or low-fidelity), we assume $K = 5$, i.e. the low-fidelity model is five times cheaper to evaluate than the high-fidelity one. In practical applications, K can be estimated and could easily be larger.

The Kriging incorporates the correlations in the low-fidelity data-values rather than the actual function-values to build the surrogate. We use a large number of low-fidelity data-values and fewer high-fidelity data-values (Kriging is forced to pass through the high-fidelity data points) to build variable-fidelity MIR enhanced Kriging surrogates. We compare their result with MIR enhanced Kriging surrogates built only with high-fidelity training points in Table 1. It can be inferred that the use of variable fidelity data can increase the accuracy of the surrogate model significantly. Depending on the computational budget, one can choose the number of low- and high-fidelity training points appropriately.

Table 1: $RMSE$ comparisons for high- and variable-fidelity Kriging models

Dimensions	Test case	N_h	N_l	$RMSE$ on variable-fidelity model	N_H	$RMSE$ on high-fidelity model
2	f_{L1} & f_1	25	125	0.64×10^{-3}	50	0.27×10^{-2}
2	f_{L2} & f_2	25	125	0.11×10^{-2}	50	0.80×10^{-2}
2	f_{L3} & f_3	25	125	0.13×10^{0}	50	0.14×10^{0}
5	f_{L1} & f_1	40	400	0.71×10^{0}	120	0.73×10^{0}
5	f_{L2} & f_2	40	400	0.31×10^{-1}	120	0.23×10^{-1}
5	f_{L3} & f_3	40	400	0.22×10^{3}	120	0.24×10^{3}
9	f_{L1} & f_1	40	800	0.72×10^{0}	200	0.73×10^{0}
9	f_{L2} & f_2	40	800	0.36×10^{-1}	200	0.33×10^{-1}
9	f_{L3} & f_3	40	800	0.17×10^{6}	200	0.17×10^{6}

V. Conclusions

We described an enhanced Kriging surrogate model approach which uses MIR as a local surrogate model to guide a dynamic training point selection process. We demonstrated the quality of the surrogate by comparison with two-, five- and nine-dimensional analytic test functions and showed it to be a good candidate for improving local deviation and space-filling designs. We also obtain a more monotone behavior in that additional training points always lead to a more accurate surrogate model. The potential of our method to approximate multi-modal functions better than other approaches has been illustrated. Though our model maintains its performance for moderately higher dimensions, it still suffers from the "curse of dimensionality". Finally, we showed promising results incorporating variable-fidelity data. Our future plans entail the application of our improved surrogate model to uncertainty quantification (UQ) and robust optimizations (RO) applied to computational fluid dynamics (CFD) problems.

Acknowledgments

We would like to thank Wataru Yamazaki for his Kriging model and Qiqi Wang for his MIR model.

References

[1]Pironneau, O., "On Optimum Design in Fluid Mechanics," *Journal of Fluid Mechanics*, Vol. 64, No. 1, 1974, pp. 97–110.

[2]Jameson, A., "Optimum Aerodynamic Design Using Control Theory," Computational Fluid Dynamics Review, Hafez, M., Oshima, K. (eds), Wiley: New York, pp. 495-528, 1995.

[3]Errico, R. M., "What is an adjoint model?" *Bulletin of the American Meteorological Society*, Vol. 8(11), 1997, pp. 2577–2591.

[4]Wiener, N., "The Homogeneous Chaos," *American Journal of Mathematics*, Vol. 60(4), 1938, pp. 897–936.

[5]Xiu, D. and Karniadakis, G. E., "Modeling Uncertainty in Flow Simulations via Generalized Polynomial Chaos," *Journal of Computational Physics*, Vol. 187(1), 2003, pp. 137–167.

[6]Cressie, N., "The Origins of Kriging," *Mathematical Geology*, Vol. 22, No. 3, 1990, pp. 239–252.

[7]Koehler, J. R. and Owen, A. B., "Computer Experiments," *Handbook of Statistics, Ghosh, S., Rao, C.R., (Eds.)*, pp. 261-308, 1996.

[8]Yamazaki, W., Rumpfkeil, M. P., and Mavriplis, D. J., "Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model," AIAA Paper, 2010-4363, 2010.

[9]McKay, M. D., Conover, W. J., and Beckman, R. J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239–245.

[10]Peter, J. and Marcelet, M., "Comparison of Surrogate Models for Turbomachinery Design," *WSEAS Transactions on Fluid Mechanics*, Vol. 3(1), 2008.

[11]Gano, S., Kim, H., and Brown, D., "Comparison of Three Surrogate Modeling Techniques: Datascape, Kriging, and Second Order Regression," AIAA Paper, 2006-7048, 2006.

[12]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, pp. 455–492.

[13]Simpson, T. W., Korte, J. J., Mauery, T. M., and Mistree, F., "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," AIAA Paper, 98-4755, 1998.

[14]Chung, H. S. and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," AIAA Paper, 2002-0317, 2002.

[15]Martin, J. D. and Simpson, T. W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol. 43, No.4, 2005, pp. 853–863,.

[16]Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413–420.

[17]Laurenceau, J. and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498–507.

[18]Laurenceau, J. and Meaux, M., "Comparison of Gradient and Response Surface Based Optimization Frameworks Using Adjoint Method," AIAA Paper, 2008-1889, 2008.

[19]Yamazaki, W., Mouton, S., and Carrier, G., "Efficient Design Optimization by Physics-Based Direct Manipulation Free-Form Deformation," AIAA Paper, 2008-5953, 2008.

[20]Zhao, D. and Xue, D., "A comparative study of metamodeling methods considering sample quality merits," *Structural and Multidisciplinary Optimization*, Vol. 42 (6), 2010, pp. 923–938.

[21]Wang, Q., Moin, P., and Iaccarino, G., "A rational interpolation scheme with super-polynomial rate of convergence," *SIAM Journal of Numerical Analysis*, Vol. 47, No. 6, 2010, pp. 4073–4097.

[22]Wang, Q., Moin, P., and Iaccarino, G., "A High-Order Multi-Variate Approximation Scheme for Arbitrary Data Sets," *Journal of Computational Physics*, Vol. 229, No. 18, 2010, pp. 6343–6361.

[23]Zhao, D. and Xue, D., "A multi-surrogate approximation method for metamodeling," *Engineering with Computers*, Vol. 27, No. 2, 2011, pp. 139–153.

[24]Sherman, L. L., Taylor III, A. C., Green, L. L., and Newman, P. A., "First- and second-order aerodynamic sensitivity derivatives via automatic differentiation with incremental iterative methods," *Journal of Computational Physics*, Vol. 129, 1996, pp. 307 – 331.

[25]Dixon, L. C., "Automatic Differentiation: Calculation of the Hessian," *Encyclopedia of Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, pp. 82–86.

[26]Taylor III, A. C., Green, L. L., Newman, P. A., and Putko, M., "Some Advanced Concepts in Discrete Aerodynamic Sensitivity Analysis," *AIAA Journal*, Vol. 41 No. 7, 2003, pp. 1224–1229.

[27]Ghate, D. and Giles, M. B., "Inexpensive Monte Carlo uncertainty analysis," *Recent Trends in Aerospace Design and Optimization, Tata McGraw-Hill, New Delhi*, 2006, pp. 203–210.

[28]Chalot, F., Dinh, Q., Herbin, E., Martin, L., Ravachol, M., and Roge, G., "Estimation of the impact of geometrical uncertainties on aerodynamic coefficients using CFD," AIAA Paper, 2068-2008, April, 2008.

[29]Rumpfkeil, M. P. and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method," AIAA Paper, 2010-1268, January, 2010.

[30]Rumpfkeil, M. P. and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method with Applications," *AIAA Journal*, Vol. 48, No. 10, 2010, pp. 2406–2417.

[31]Rumpfkeil, M. P., Yamazaki, W., and Mavriplis, D. J., "Uncertainty Analysis Utilizing Gradient and Hessian Information," Sixth International Conference on Computational Fluid Dynamics, ICCFD6, St. Petersburg, Russia, July 12-16, 2010.

[32]Rumpfkeil, M. P., Yamazaki, W., and Mavriplis, D. J., "A Dynamic Sampling Method for Kriging and Cokriging Surrogate Models," AIAA Paper, 2011-883, 2011.

[33]Arora, J. S., *Optimization of Structural and Mechanical Systems*, World Scientific Publishing Co. Pte. Ltd., 2007.

[34]Metropolis, N. and Ulam, S., "The Monte Carlo method," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 335–341.

[35]Sobol, I., *A primer for the Monte Carlo Method*, CRC Press, 1994.

[36]Keane, A. and Nair, P., *Computational Approaches for Aerospace Design*, John Wiley & Sons, 2005.

[37]Forrester, A., Sobester, A., and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*, John Wiley & Sons, 2008.

[38]Alexandrov, N. M., Dennis, J. E., Lewis, R. M., and Torczon, V., "A Trust-Region Framework for Managing the Use of Approximation Models in Optimization," *Structural Optimization*, Vol. 15, 1998, pp. 16–23.

[39]Mehmani, A., Chowdhury, S., Zhang, J., and Messac, A., "Regional Error Estimation of Surrogates (REES)," AIAA Paper, 2012-5707, 2012.

[40]Rosenbaum, B. and Schulz, V., "Efficient response surface methods based on generic surrogate models," *SIAM Journal of Scientific Computing*, Vol. Submitted, 2012.

[41]Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, Vol. 34(3), 2006.

[42]Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. (4), 1989, pp. 409–423.

[43]Kraaijpoel, D. A., *Seismic ray fields and ray field maps: theory and algorithms*, Ph.D. thesis, Universiteit Utrecht, The Netherlands, 2003.

[44]Kraaijpoel, D. and van Leeuwen, T., "Raising the order of multivariate approximation schemes using supplementary derivative data," *Procedia Computer Science*, Vol. 1(1), 2010, pp. 307–316.

[45]Han, Z. H., Zimmermann, R., and Goertz, S., "On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context," CEAS 2009 European Air and Space Conference, 2009.

[46]Han, Z. H., Zimmermann, R., and Goertz, S., "A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data," AIAA Paper, 2010-1225, 2010.

[47]Yamazaki, W. and Mavriplis, D. J., "Derivative-Enhanced Variable Fidelity Surrogate Modeling for Aerodynamic Functions," AIAA Paper, 2011-1172, 2011.

[48]Han, Z. H., Goertz, S., and Zimmermann, R., "Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function," *Aerospace Science and Technology*, doi:10.1016/j.ast.2012.01.006, 2012.