11th World Congress on Structural and Multidisciplinary Optimisation
07th -12th, June 2015, Sydney Australia

One-dimensional Function Extrapolation Using Surrogates

Yiming Zhang[1], Nam-Ho Kim[2], Chan-Young Park[3], Raphael T. Haftka[4]

[1] University of Florida, Gainesville, FL, USA, yimingzhang521@ufl.edu
[2] University of Florida, Gainesville, FL, USA, nkim@ufl.edu
[3] University of Florida, Gainesville, FL, USA, cy.park@ufl.edu
[4] University of Florida, Gainesville, FL, USA, haftka@ufl.edu

1. Abstract
Surrogate modeling is commonly used to estimate function values efficiently and accurately at unsampled points. The estimation procedure is called interpolation when target points are inside the convex hull of sampled points while extrapolation otherwise. This paper explores one-dimensional deterministic function extrapolation using surrogates. We first define a new error metric, relative average error, for quantifying overall performance of extrapolation technique. Ordinary Kriging and Linear Sheppard surrogates proved to be safer on several challenging functions than polynomial response surfaces, support vector regression or radial basis neural functions. This reflected that prediction of these surrogates converge to mean value of samples at points far from samples.

It's commonly recognized that long-range extrapolation is likely to be less accurate than short-range extrapolation. Two kinds of effective extrapolation distance are defined to indicate how far we can extrapolate test functions. We propose using the correlation between the nearest sample and the prediction point given by Ordinary Kriging as indicator of effective extrapolation distance. The relationship between effective extrapolation distance and corresponding correlation over the distance is examined by several test functions. A large value of correlation is associated with effective extrapolation distance.

2. Keywords: One-dimensional extrapolation, Extrapolation distance, Kriging, Error metric

3. Introduction
In surrogate modeling, it is common to sample a function f at several points and fit them with an explicit function in order to estimate the function at other points[1]. This is often required for optimization or reliability analysis in which thousands of function evaluations are common, and each sample often means expensive simulation or costly or time-consuming experiment.

Function estimation is defined as interpolation when target points are inside the convex hull of sampled data points and extrapolation otherwise. For one-dimensional samples, convex hull is the smallest interval containing the samples. Although many research results have been reported on the accuracy of surrogate modeling, most focused on the prediction accuracy in interpolation. Extrapolation is usually associated with large estimation errors [2] and commonly encountered in three situations:

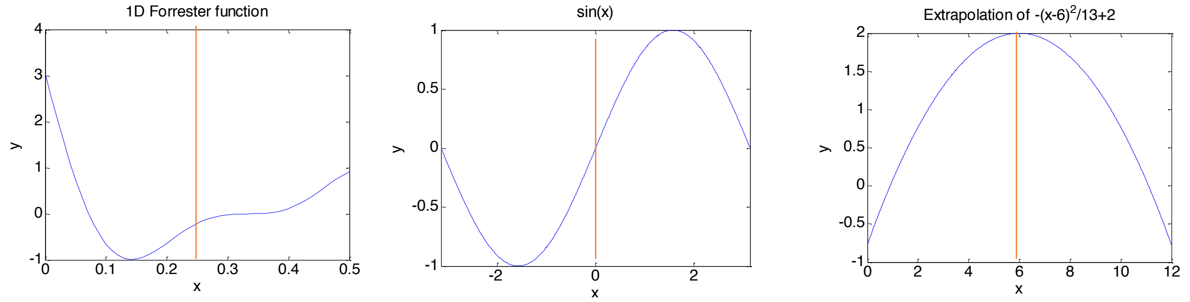
1) Sampling pattern such as Latin Hypercube sampling is adopted, which typically does not sample at or near the boundaries of sampling region.
2) For function estimation in high-dimensional space, we usually cannot afford enough points to avoid extrapolation. For example, in twenty-dimensional box, more than million points (2^{20}) are required.
3) Region of interest changes after samples are collected[3].

Besides the above conditions, extrapolation may be useful when the target points cannot be sampled via simulation or experiment due to the need to know future events, inadequacy of simulation software or high cost to perform experiments[4]. As a first step to explore effective extrapolation scheme in engineering problems, attention is limited here to one-dimensional (1D) function extrapolation.

This paper investigates general issues on extrapolation using surrogates. Section 4 illustrates possible behavior of surrogates for extrapolation and potentials of extrapolation using surrogates. Section 5 proposes an error metric designed for extrapolation. Extrapolation of a few examples using five surrogates are compared in Section 6. Section 7 discusses the possibility of estimating extrapolation distance using surrogates.

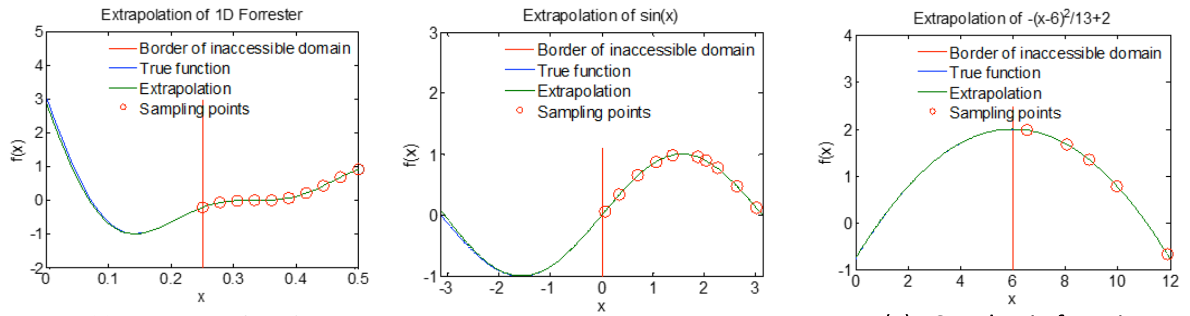
4. Possible behaviours of extrapolation using surrogates
Estimation of several analytical functions using ordinary kriging is presented to illustrate possible behavior in the extrapolation region, which is assumed here to be inaccessible. Figure1 presents three 1D functions which have different function behavior between the accessible and inaccessible domains. These three functions are estimated

using Ordinary Kriging in Fig.2. It is seen extrapolation results approximate the true function value surprisingly well. Since extrapolation in Kriging is based on correlation between function values based on distance, we evaluate first the correlation between inaccessible domain and accessible domain.



(a) Forrester function (b) sin(x) (c) Quadratic function

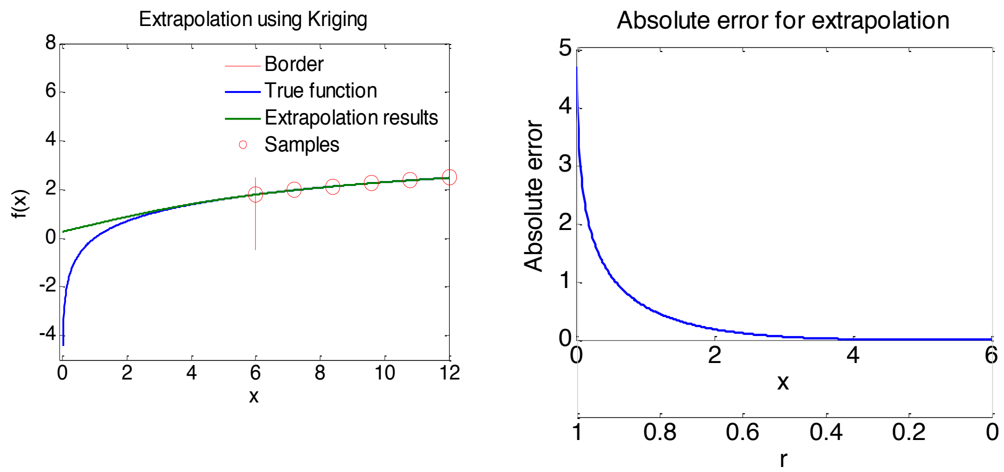
Figure 1: (a) Forrester function, (b) $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$ and (c) $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$. Vertical line denotes border of inaccessible domain which is on the left and accessible domain which is on the right



(a) Forrester function (b) sin(x) (c) Quadratic function

Figure 2: Extrapolation of (a) Forrester function, (b) $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$ and (c) $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$. Vertical line denotes border of inaccessible domain which is on the left and accessible domain which is on the right

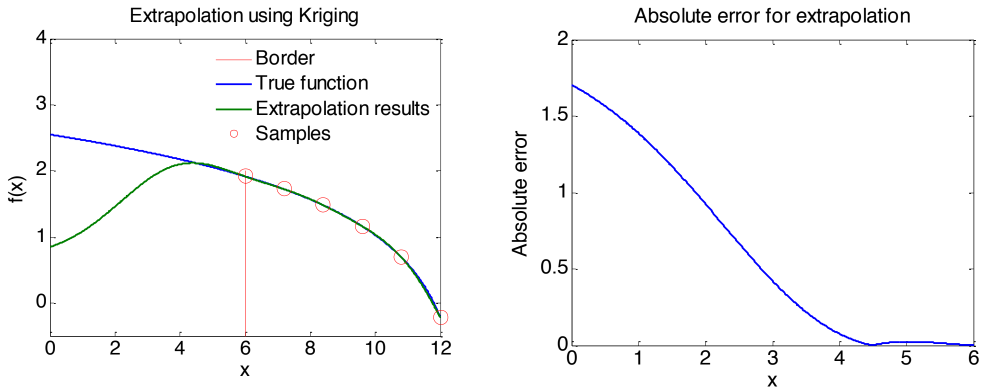
Denote by r the ratio of length of extrapolation distance to that of accessible domain. As expected, extrapolation error increases with r as shown in Fig. 3 for the log function. In this paper, the length of inaccessible domain and accessible domain are set to be equal, which would typically be considered as long range extrapolation.



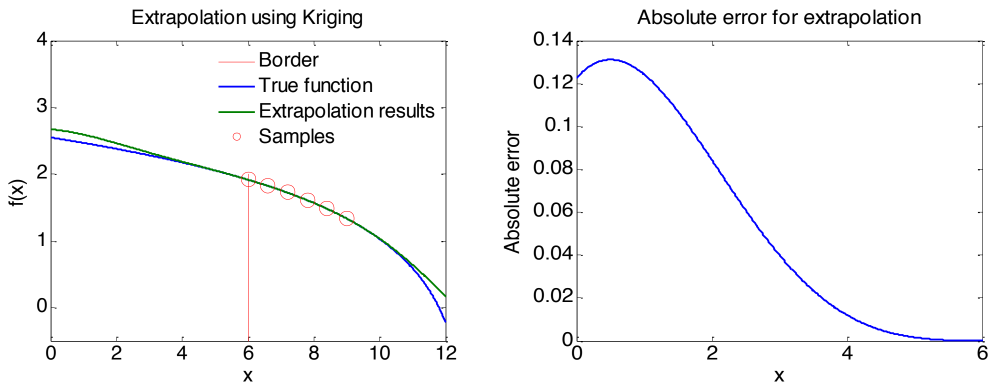
(a) Extrapolation results (b) Absolute error for extrapolation

Figure 3: Extrapolation of $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$

Another factor determining the extrapolation accuracy is how close the samples to the boundary are. Figure 4 illustrates that by using samples close to the boundary. Extrapolation accuracy improves obviously after shifting samples close to inaccessible domain.



(a) Full sampling domain



(b) Reduced sampling domain

Figure 4: Extrapolation of $H=2\left(F(X_1)-F(X_2)-\sum_{i=1}^{n}\frac{\partial F(X_2)}{\partial x_i}\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i}-x_{i,2}^{p_i})\right)$ **from (a) Full sampling domain and (b) Reduced sampling domain**

The examples and discussion above identified three issues, which are correlation between inaccessible domain and accessible domain, relative extrapolation distance and absolute extrapolation distance, which are important for extrapolation research.

5. An error metric for 1D extrapolation

We denote extrapolation result by $\hat{f}(x)$. The performance of one-dimensional extrapolation technique may be quantified by various error measures. Relative error $e_{cr}(x)$ is

$$e_{cr}(x)=\left|\frac{\hat{f}(x)-f(x)}{f(x)}\right| \tag{1}$$

$e_{cr}(x)$ may be misleading when the function changes sign. So one often uses the range of the function instead of the function value for normalization. In addition, for extrapolation, we are often interested in the error in predicting change from a boundary point x_b. This error, $e_{ec}(x)$ is:

$$e_{ec}(x)=\left|1-\frac{\hat{f}(x)-f(x_b)}{f(x)-f(x_b)}\right| \tag{2}$$

For example, if based on this year's record we predict that gas prices will rise from \$4/gallon today to \$5/gallon a year from now, and they rise only to \$4.50, we may consider $e_{ec}(x)$=100% this as rather than $e_{cr}(x)$=11% error. Of course, this error measure will fail if the change in the function is near zero, and for this case an alternate relative error is defined as $e_r(x)$. Denoting $range(f)$ as the range of true function in the extrapolation domain, $e_r(x)$ is:

$$e_r(x)=\left|\frac{\hat{f}(x)-f(x)}{range(f)}\right| \tag{3}$$

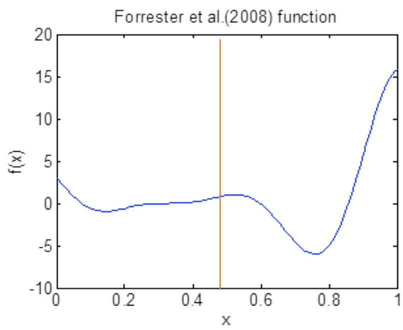
$e_r(x)$ is used in the following for extrapolation comparisons. We may also use $range(f(x)) = \max(f_1) - \min(f_1)$, where f_1 denotes function value in the range between extrapolation point x and the closest sample. In order to evaluate the overall performance of extrapolation, we use the average error AE in extrapolation domain:

$$AE = \frac{\int_{x_t}^{x_b} e_r(x)dx}{x_b - x_t} \tag{4}$$

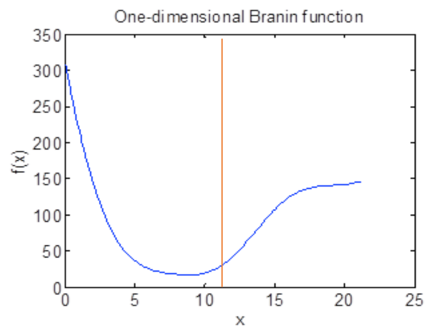
6. Surrogate comparison for extrapolation

Surrogates have different performance for interpolation and extrapolation. We test the performance of five popular surrogates: Ordinary Kriging, Polynomial response surface (PRS), Radial basis neural network (RBNN), Linear Shepard (moving least squares), and Support vector regression (SVR). Four test functions were extracted from well-known multidimensional functions taken from [5] and shown in Fig. 5.

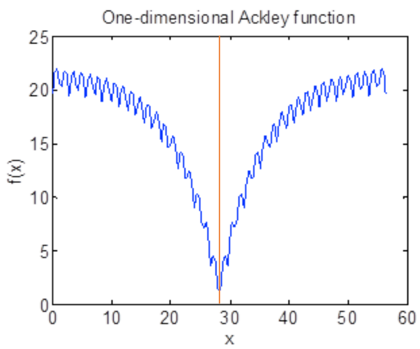
The number of sampling points along each line is 6. Sampling points are generated using Latin hypercube sampling with 5 iterations, which introduces randomness in the position of the samples. To average out the effect of the positions of the sampling points, 30 sets of samples are generated for each test function, and the mean value of AE for all the sample sets are computed. The extrapolation results of test functions are listed in Table 1. All the surrogates except Ordinary Kriging and Linear Shepard can generate huge errors. Kriging and Linear Shepard do not extrapolate well, but do not incur huge errors. This is because function estimation using Ordinary Kriging and Linear Shepard are weighted sum of samples and they eventually revert to the mean of the samples. Ordinary Kriging was then selected for further testing in the next section.



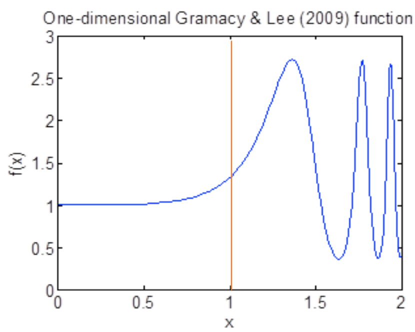
(a) Forrester function



(b) 1D Braining-Hoo function



(c) 1D Ackley function



(d) 1D Gramacy & Lee function

Figure 5: Four test functions for surrogate selection. Domains close to origin are inaccessible domain.

Table 1: Average AE for extrapolation of test functions using 30 sets of samples

Surrogate models	Test functions			
	Average AE of Forrester et al.(2008) function	Average AE of one-dimensional Branin-Hoo function	Average AE of one-dimensional Ackley function	Average AE of one-dimensional Gramacy & Lee (2009) function

Kriging	0.80	0.13	0.31	1.44
Quadratic PRS	13.36	0.97	2.01	6.13
Cubic PRS	7.54	0.53	4.02	60.22
Quartic PRS	93.62	3.18	8.67	442.14
RBNN	2322.5	0.2	0.54	25612
Linear Shepard	0.44	0.19	0.56	1.85
SVR	20.7	0.26	0.76	188.08

7. Estimating extrapolation distance

Kriging is based on a correlation structure between points based on their distance. Large correlation between extrapolation points to closest sample may indicate reliable extrapolation. We defined two types of effective extrapolation distance and tried to find the relation between extrapolation distances of test functions and corresponding correlation over that extrapolation distance.

7.1. Effective extrapolation distance

Ordinary Kriging assumes that the error at a point is normally distributed with a mean of zero and a given standard deviation. Error bounds here are set be 95% confidence interval of this normal distribution. The conservative extrapolation distance d is defined for measuring how far the error bounds of the surrogate bound the true function.

The second extrapolation distance is denoted as accuracy distance. Accurate distance is inside conservative distance and in which estimated error bounds of the points are less than 30% of $range(f)$. We make an exception to the requirement of being within the error bounds when they are very tight, allowing error bounds to be off by 1% of $range(f)$). Two types of effective extrapolation distance are illustrated in Fig. 6.

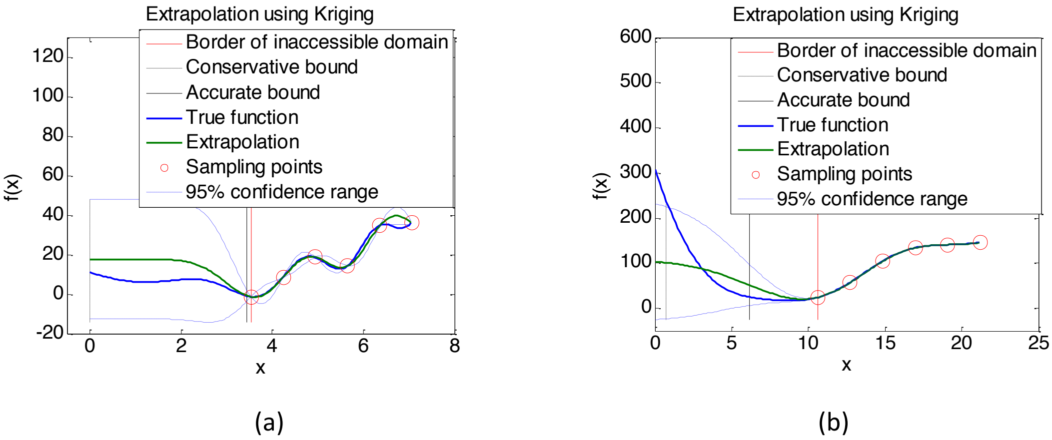


Figure 6. Conservative and accurate distances for extrapolation

7.2 Identification of effective extrapolation distance using correlation

The error estimates get less dependable as we go deeper into the extrapolation domain. We tried to find certain indicators of effective extrapolation distance based on Kriging. Prediction of Ordinary Kriging is based on the assumption that correlation between function values decay with distance at a rate controlled by θ, with the correlation r between two points at a distance l being equal to

$$r(l) = \exp\left(-\theta \times l^2\right) \tag{5}$$

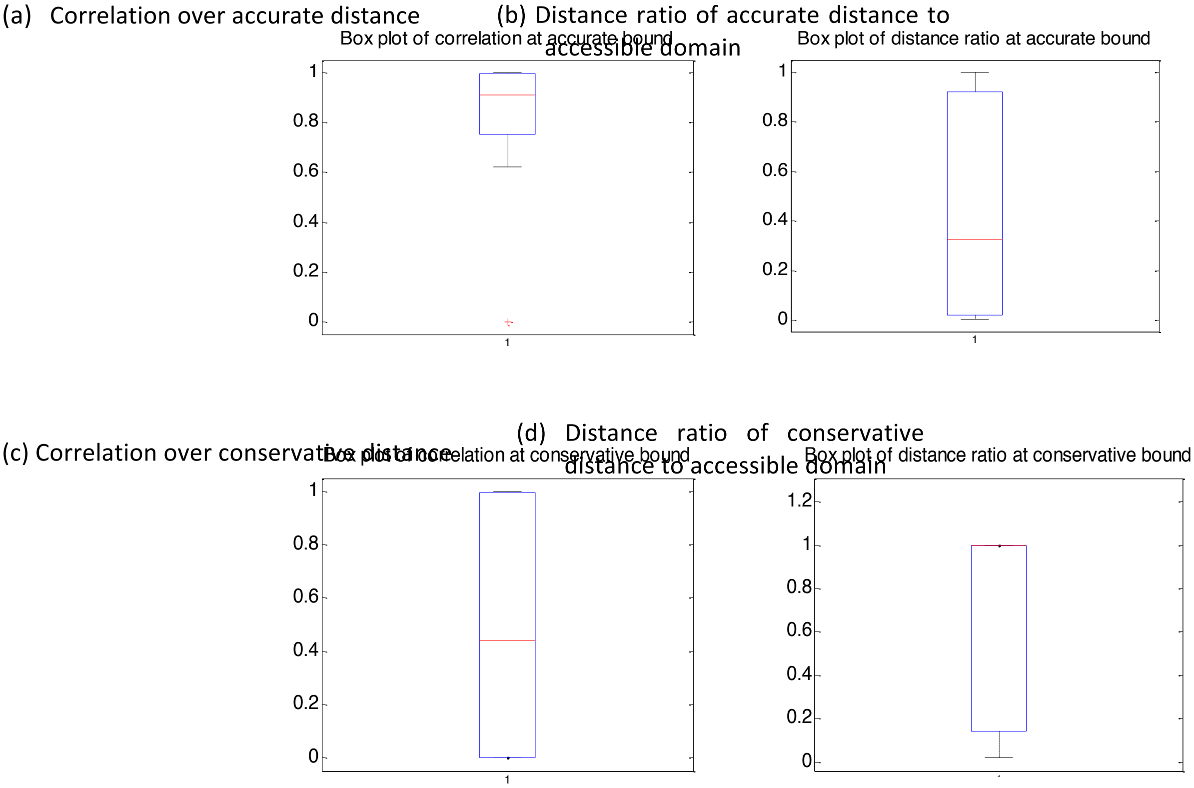
θ is usually found by maximizing likelihood of observing that the samples come from Gaussian process. Large θ means short wavelength, large curvature, fast changing function, and reverse for small θ. It is reasonable to expect that as the correlation between function values in the sampling domain and extrapolation domain diminishes, the reliability of the error estimates deteriorates. Distance corresponds to given small correlation as possible measure of how far we can go.

We have performed a test to figure out the relation between effective extrapolation distance and corresponding correlation value. 10 multi-dimensional functions: Branin-Hoo function, Ackley function, Gramacy & Lee (2009) function, Hartmann 3-D Function, Hartmann 6-D Function, Sasena Function, Friedman Function, Zhou (1998) Function, Franke's Function, Dette & Pepelyshev (2010) curved Function.

These functions are commonly used for testing algorithm performance and can be found from [5]. 3 lines are

Figure 7: Correlation and distance ratio over effective distance while extrapolating 30 1D functions

extracted towards one random vertex from each function. We use 6 uniformly spaced samples to train Kriging. In Fig.7, we present corresponding correlation value between effective extrapolation bounds and closest sample. Accurate bounds are associated with large correlation value. The third quartiles of correlation values corresponding to accurate bounds and conservative bounds are both 0.99. The box plots of distance ratios are dispersed and imply extrapolation accuracy vary with functions.



8. Summary

This paper first illustrated the possibility of extrapolating 1D function using surrogates and proposed an average error metric designed for quantifying the performance of extrapolation technique. Testing extrapolation on several challenging functions indicated that Kriging and Linear Shepard were safer than other surrogates. We defined two types of effective extrapolation distance and correlation of Ordinary Kriging has been demonstrated as a possible indicator for effective extrapolation distance based on the tests of 30 one-dimensional functions.

9. Acknowledgement

This work is supported by the U.S. Department of Energy, National Nuclear Security Administration, Advanced Simulation and Computing Program, as a Cooperative Agreement under the Predictive Science Academic Alliance Program, under Contract No. DE-NA0002378.

10. References

[1] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker, "Surrogate-based analysis and optimization," *Progress in aerospace sciences,* vol. 41, pp. 1-28, 2005.

[2] A. Chaudhuri and R. T. Haftka, "Optimistic Bias in Surrogate Prediction near Surrogate Optima."

[3] V. Balabanov, O. Weckner, and J. Wu, "Reducing Error of Polynomial Approximation Outside of Designated Design Space for Practical Problems," 2014.

[4] D. An, J.-H. Choi, N. H. Kim, and S. Pattabhiraman, "Fatigue life prediction based on Bayesian approach to incorporate field data into probability model," *Structural engineering & mechanics,* vol. 37, p. 427, 2011.

[5] “Virtual Library of Simulation Experiments: Test Functions and Datasets”, [Online] Available: http://www.sfu.ca/~ssurjano/optimization.html