

Yiming Zhang
Department of Mechanical
and Aerospace Engineering,
University of Florida,
Gainesville, FL 32611-6250
e-mail: yimingzhang521@ufl.edu

Chanyoung Park
Department of Mechanical
and Aerospace Engineering,
University of Florida,
Gainesville, FL 32611-6250
e-mail: cy.park@ufl.edu

Nam H. Kim
Department of Mechanical
and Aerospace Engineering,
University of Florida,
Gainesville, FL 32611-6250
e-mail: nkim@ufl.edu

Raphael T. Haftka
Department of Mechanical
and Aerospace Engineering,
University of Florida,
Gainesville, FL 32611-6250
e-mail: haftka@ufl.edu

Function Prediction at One Inaccessible Point Using Converging Lines

The focus of this paper is a strategy for making a prediction at a point where a function cannot be evaluated. The key idea is to take advantage of the fact that prediction is needed at one point and not in the entire domain. This paper explores the possibility of predicting a multidimensional function using multiple one-dimensional lines converging on the inaccessible point. The multidimensional approximation is thus transformed into several one-dimensional approximations, which provide multiple estimates at the inaccessible point. The Kriging model is adopted in this paper for the one-dimensional approximation, estimating not only the function value but also the uncertainty of the estimate at the inaccessible point. Bayesian inference is then used to combine multiple predictions along lines. We evaluated the numerical performance of the proposed approach using eight-dimensional and 100-dimensional functions in order to illustrate the usefulness of the method for mitigating the curse of dimensionality in surrogate-based predictions. Finally, we applied the method of converging lines to approximate a two-dimensional drag coefficient function. The method of converging lines proved to be more accurate, robust, and reliable than a multidimensional Kriging surrogate for single-point prediction. [DOI: 10.1115/1.4036130]

Keywords: extrapolation and interpolation, Bayesian inference, surrogate model, design of experiments, multiple models

1 Introduction

In surrogate modeling, it is common to sample a function at scattered points and fit the samples with an explicit function to approximate function values at unsampled points [1,2]. The estimation procedure is called interpolation when target points are inside the convex hull of sampled points, while it is termed extrapolation [3,4] otherwise, as shown in Fig. 1. For one-dimensional samples, for example, the convex hull is set by the lower and upper bounds of samples. Surrogate models are routinely used as interpolation schemes and may be inadequate for extrapolation, which is commonly encountered while approximating high-dimensional functions [5]. A surrogate is especially useful to estimate function value when the point of interest cannot be sampled via simulation or experiment due to extreme conditions, lack of data, limitations of simulation software, or an experiment that is too dangerous to perform. For example, 16 out of the 500 simulations were reported to have failed for an aeroelastic helicopter blade simulation in Glaz et al. [6] during design optimization.

Several valuable efforts have been initiated toward reliable estimation at inaccessible points. Neural networks, an advanced framework, have been reported to be misleading for extrapolation beyond the region of samples [4] and are therefore mostly used for interpolation. Balabanov et al. [7] and Hooker and Rosset [8] demonstrated that regularization methods can improve extrapolation accuracy beyond sample range even if they compromise interpolation accuracy. Wilson and Adams [9] proposed a Gaussian process kernel function for better uncovering data patterns and leading to improved prediction accuracy of forecasting.

Richardson extrapolation has been commonly adopted to predict the response to finite element analysis toward extremely refined mesh that is unaffordable [10,11]. A specific polynomial form was used for Richardson extrapolation to incorporate physical knowledge about convergence order. Hooker [12] proposed an empirical trend function when predicting toward inaccessible point. In prognosis for remaining life of mechanical systems, application of Gaussian process surrogate and neural network model has been discussed based on a heuristic study [13]. Besides, numerous papers can be found on forecasting one-dimensional time series data at inaccessible points (i.e., future) in financial fields and weather forecasting, e.g., see Refs. [14,15].

Based on this literature, it is challenging to estimate prediction accuracy at inaccessible points. Surrogate predictions, in particular, risk increasing uncertainty when prediction points are far from samples [7,16,17]. This motivated us to estimate a function value by creating multiple independent predictions from different data sets. Introducing multiple predictions may reduce the uncertainty

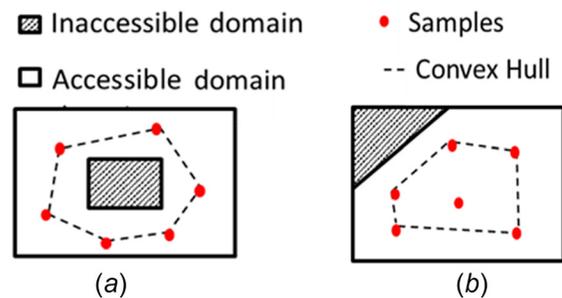


Fig. 1 Illustration of interpolating and extrapolating regions in design of experiments: (a) interpolating inaccessible region and (b) extrapolating inaccessible region

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received August 9, 2016; final manuscript received February 6, 2017; published online March 21, 2017. Assoc. Editor: Gary Wang.

of blind predictions. A new sampling pattern, called the method of converging lines, is proposed in this work. As a first step toward improving the prediction capability for long-distance prediction using surrogate model, we explore the possibility of predicting a multidimensional function using multiple one-dimensional lines. The main idea is to select samples along lines toward the inaccessible point. One-dimensional surrogates are then constructed based on the samples along each line in order to predict the function value at the inaccessible point. The method of converging lines transforms multidimensional function prediction into a series of one-dimensional predictions, which provide multiple estimates at the inaccessible point. Combining multiple predictions based on Bayesian inference is proposed to enhance the prediction accuracy.

The paper is organized as follows: In Sec. 2, we present the concept and technical details for the method of converging lines. In Sec. 3, we examine numerical properties of proposed approach using two algebraic functions. In Sec. 4, we apply the method of converging lines to approximate a two-dimensional drag coefficient function to demonstrate potential application. One-dimensional surrogate prediction is performed using ordinary Kriging due to its excellence for interpolating noise-free function.

2 Method of Converging Lines

2.1 Principal Idea. The procedure to select locations where the function is to be sampled is called the design of experiments (DOE). For fitting a surrogate, it is common to select space-filling DOEs (e.g., Latin hypercube sampling (LHS) [18]), which are geared to provide a good representation of the design domain. The method of converging lines, in contrast, serves the prediction at a single target point, $\mathbf{x}_{\text{target}}$. This method locates sampling points along several lines all intersecting at the target point.

The difference between LHS and the sampling scheme of the method of converging lines in two- and three-dimensional space is illustrated in Fig. 2. In the figure, the target point, $\mathbf{x}_{\text{target}}$, is at

the origin, and the domain within a normalized distance of 0.5 from $\mathbf{x}_{\text{target}}$ is set to be inaccessible. Three converging lines are shown in Fig. 2(b). The orientations of the line l_i may vary with applications in order to generate a good estimate. Along a given line, local coordinate $x_{\text{local}}^{(j)}$ of j th sample point $\mathbf{x}^{(j)}$ is the Euclidean distance between $\mathbf{x}^{(j)}$ and $\mathbf{x}_{\text{target}}$ as

$$x_{\text{local}}^{(j)} = \|\mathbf{x}^{(j)} - \mathbf{x}_{\text{target}}\| \quad (1)$$

where $x_{\text{local}}^{(j)}$ is a non-negative scalar and zero when $\mathbf{x}^{(j)} = \mathbf{x}_{\text{target}}$. The local coordinate of the sample, $\mathbf{x}^{(j)}$, is first normalized to be within [0,1] along each dimension to eliminate the effect of different scales.

The method of converging lines is expected to be ideal in certain conditions, where sampling domain is relatively convenient to access in terms of cost and feasibility, and sampling the target point is impossible or exponentially more challenging. In this paper, we focused on long-distance prediction, namely, the length over inaccessible domain equals to that from the accessible domain. Under these two expected scenarios, it might be desirable to perform pointwise prediction using multiple lines.

2.2 One-Dimensional Function Estimation Using Ordinary Kriging.

The underlying assumption for prediction is that the function behaves similarly in the accessible domain and inaccessible domain. Therefore, surrogate models providing acceptable accuracy in sampling domain would be appropriate to estimate the function at inaccessible points within a certain distance from samples. When this assumption fails, the surrogate prediction is likely to fail, whether we use a multidimensional surrogate or the method of converging lines.

Ordinary Kriging based on correlation kernel is adopted for one-dimensional modeling due to its superior performance for approximating noise-free data. Surrogate prediction based on ordinary Kriging is reported to avoid large errors at inaccessible points [19]. Kriging assumes that the distance between sample points reflects a spatial correlation. The value of the function at a point is correlated to the values at neighboring points based on their separation. The correlation is strong to nearby points and weak with far away points, but strength does not change based on location. This unique assumption provides a measure of uncertainty of the prediction. We adopt the Gaussian model to approximate the correlation between point \mathbf{x} and point \mathbf{s} as

$$C(f(\mathbf{x}), f(\mathbf{s}), \theta) = \prod_{i=1}^n \exp(-\theta_i(x_i - s_i)^2) \quad (2)$$

where n is the dimensionality of variable \mathbf{x} and \mathbf{s} , θ_i is the hyperparameter along i th direction [20]. n is 1 for applying one-dimensional prediction using the method of converging lines.

Prediction at an untested point \mathbf{x} is formulated as

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\beta}^T \mathbf{r} \quad (3)$$

$$r_i = \exp\left(-\sum_{k=1}^n \theta_k (s_k^{(i)} - x_k)^2\right)$$

where $s^{(i)}$ denotes i th sample point, and $\hat{\mu}$ is the mean of samples. The hyperparameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are obtained by maximizing the likelihood that the data come from a Gaussian process. Ordinary Kriging is implemented using the surrogate toolbox of Viana and Goel [21].

The prediction of ordinary Kriging and the corresponding uncertainty are expressed as a conditional probability density function (PDF) for a given data set \mathbf{y} as $p(y(\mathbf{x})|\mathbf{y})$, which follows a student's t-distribution [22]. A student's t-distribution is well approximated by a normal distribution and converges to a normal distribution as the number of samples increases. In this paper, the

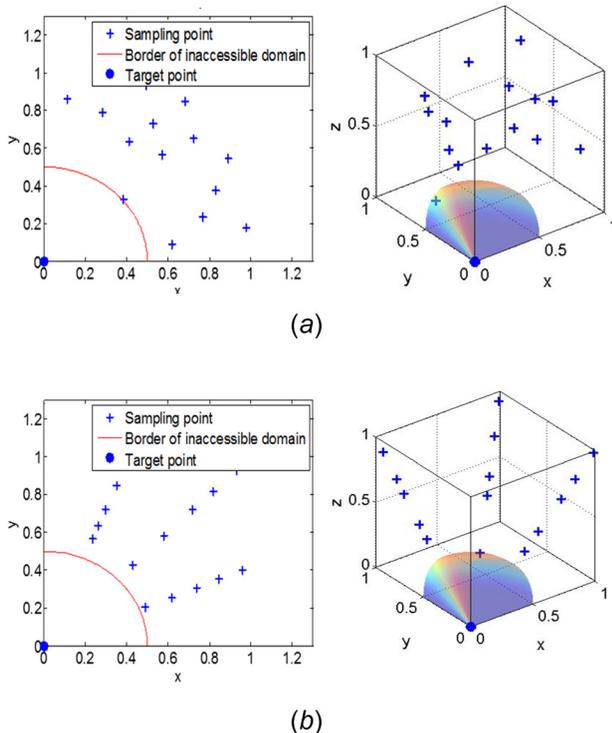


Fig. 2 Illustration of (a) space-filling sampling (LHS) and (b) method of converging lines using 15 samples when the target point is at the origin and the domain at distance of less than 0.5 from the origin is inaccessible

surrogate prediction with uncertainty is expressed using a normal distribution for computational efficiency. The conditional PDF of a prediction is defined as a normal distribution $N(\hat{y}(\mathbf{x}), \sigma(\mathbf{x})^2)$, where $\hat{y}(\mathbf{x})$ is the Kriging predictor and $\sigma(\mathbf{x})$ is the standard deviation at \mathbf{x} .

2.3 Combination of Prediction From Multiple Lines. For making a prediction in n -dimensional space, one-dimensional surrogate allows higher accuracy than using an n -dimensional surrogate with the same number of samples. The method of converging lines proposes to combine multiple predictions at the inaccessible point with independently built multiple one-dimensional surrogates toward the point to reduce the chance of large error. The true response $y(\mathbf{x}_{\text{target}})$ is predicted with multiple one-dimensional Kriging surrogates, and the prediction of the i th surrogate and the corresponding prediction uncertainty are defined through a conditional PDF for the given i th data set as $p(y_i(\mathbf{x}_{\text{target}})|\mathbf{y}_i)$, which is a normal distribution having a mean of the Kriging predictor $\hat{y}_i(\mathbf{x}_{\text{target}})$ and a variance of the prediction variance at the target point $\mathbf{x}_{\text{target}}$.

The multiple predictions can be combined to obtain a single most probable estimate. Various combination schemes are available to combine multiple predictions [23,24]. Here, we adopt Bayesian inference [25,26] to combine Kriging predictions. In the Bayesian inference, the conditional PDF of the true response from the Kriging based on i th data set can be interpreted as likelihood functions as $p(y(\mathbf{x}_{\text{target}})|\mathbf{y}_i) \propto l(y(\mathbf{x}_{\text{target}})|\mathbf{y}_i)$. By assuming independence of prediction uncertainties between lines, the posterior distribution of the true response is obtained by multiplying the likelihood functions. With m one-dimensional predictions, the posterior distribution is expressed as

$$p\left(y(\mathbf{x}_{\text{target}})\left|\bigcup_{i=1}^m \mathbf{y}_i\right.\right) \propto \prod_{i=1}^m l(y(\mathbf{x}_{\text{target}})|\mathbf{y}_i) \quad (4)$$

The posterior distribution in Eq. (4) can be further derived based on the approximation using normal distributions as

$$\begin{aligned} p\left(y(\mathbf{x}_{\text{target}})\left|\bigcup_{i=1}^m \mathbf{y}_i\right.\right) &\propto \prod_{i=1}^m p\left(y(\mathbf{x}_{\text{target}})\left|\hat{y}_i(\mathbf{x}_{\text{target}}), \sigma_i(\mathbf{x}_{\text{target}})^2\right.\right) \\ &\propto \prod_{i=1}^m \exp\left[-\frac{(y(\mathbf{x}_{\text{target}}) - \hat{y}_i(\mathbf{x}_{\text{target}}))^2}{2\sigma_i(\mathbf{x}_{\text{target}})^2}\right] \\ &= \exp\left\{-\frac{(y_{\text{target}} - \hat{y}(\mathbf{x}_{\text{target}}))^2}{2\sigma(\mathbf{x}_{\text{target}})^2}\right\} \end{aligned} \quad (5)$$

where $p(y(\mathbf{x}_{\text{target}})|\hat{y}_i(\mathbf{x}_{\text{target}}), \sigma_i(\mathbf{x}_{\text{target}})^2)$ is a PDF of a normal distribution at $y(\mathbf{x}_{\text{target}})$ for the given mean of $\hat{y}_i(\mathbf{x}_{\text{target}})$ and standard deviation of $\sigma_i(\mathbf{x}_{\text{target}})$. $\hat{y}(\mathbf{x}_{\text{target}})$ and $\sigma(\mathbf{x}_{\text{target}})$ are the combined prediction and the standard deviation representing the prediction uncertainty as

$$\begin{cases} \hat{y}(\mathbf{x}_{\text{target}}) = \frac{\sum_{i=1}^m \hat{y}_i(\mathbf{x}_{\text{target}})}{\sum_{i=1}^m \frac{1}{\sigma_i(\mathbf{x}_{\text{target}})^2}} \\ \sigma(\mathbf{x}_{\text{target}}) = \sqrt{\frac{1}{\sum_{i=1}^m \frac{1}{\sigma_i(\mathbf{x}_{\text{target}})^2}}} \end{cases} \quad (6)$$

Based on Eqs. (4) and (5), the posterior distribution $p(y(\mathbf{x}_{\text{target}})|\bigcup_{i=1}^m \mathbf{y}_i)$ follows a normal distribution, $N(\hat{y}(\mathbf{x}_{\text{target}}), \sigma(\mathbf{x}_{\text{target}})^2)$. From the expression of $\sigma(\mathbf{x}_{\text{target}})$ in Eq. (6), it can be

expected that surrogate prediction from the line with minimum prediction variance will be most influential on the posterior distribution because the combined prediction is a weighted average of the predictions based on the prediction variances.

Compared with multidimensional approximation, the method of converging lines is essentially a set of predictions from one-dimensional (1D) surrogates. The 1D surrogate is likely to be more robust than multidimensional approximation for two reasons: (a) The number of hyperparameters to be estimated for building a one-dimensional surrogate is one, while n hyperparameters have to be estimated for an n -dimensional surrogate. Therefore, the cost of evaluating samples to make a reasonable prediction is affordable regardless dimensionality while that to build n -dimensional surrogate increases excessively (the curse of dimensionality). (b) The 1D surrogate can be easily visualized reducing the chance that failure of the surrogate fitting will go unnoticed.

One important feature of the method of converging lines is that lines may have diverse performance. Some lines may be friendly for mathematical modeling, namely, enabling small and reliable prediction variance. These good lines will dominate combined prediction based on Bayesian method. By potentially improving the quality of surrogate models and combining multiple lines, the method of converging lines is expected to be more reliable regarding pointwise estimation. Another bonus feature is that the number of sampling points required along lines for accurate prediction may not depend on the dimensionality of the function. For example, if we were in 20-dimensional space, prediction with a standard LHS may require hundreds of points, while it is possible that three lines with 20 points may still suffice for prediction with the method of converging lines. The effect of dimensionality will be discussed in Sec. 3.

Several conditions may lead to large prediction errors even under the valid assumption that the inaccessible domain had a similar trend as that of the accessible domain:

- (1) Long-distance prediction usually risks large errors. Lacking samples near inaccessible domain would make it challenging to determine surrogate parameters and interpret goodness-of-fit.
- (2) Different surrogate models or parameter settings is a major source of uncertainty. It is more challenging to select appropriate models for predictions toward inaccessible points especially when no prior information for noise is available.
- (3) Some lines may not be trusted due to physical complexity or inappropriate modeling. Outlier prediction could be identified and excluded when most lines have similar predictions, but one line is significantly different. Outlier prediction comes from the surrogate model with misleading estimation and prediction variance.

Note, however, that these difficulties are likely to be even more acute if the same prediction is attempted with a multidimensional surrogate, especially that only a single estimate is obtained.

3 Numerical Properties of Method of Converging Lines

Even though it is tough to determine what function could be approximated and how to predict toward inaccessible points, we tried to identify and discuss several important factors based on numerical test functions. We examined the effect of lines regarding prediction accuracy, compared prediction from multiple lines with that from multidimensional approximation, and effect of high-dimensionality.

Extrapolation has been reported to be more challenging than interpolation in general [19]. Therefore, the following analysis distinguishes between these two scenarios. We adopt ordinary Kriging for prediction considering it could avoid huge error regarding prediction toward inaccessible points [19].

Table 1 Target point, variable range, and inaccessible domain of 8D Dette function

| | Target point x_{target} | Variable range | Inaccessible domain |
|---------------|----------------------------------|------------------|------------------------|
| Extrapolation | $x_i = 1$ | $x_i \in [0, 1]$ | $x_i \in [0.5, 1]$ |
| Interpolation | $x_i = 0.5$ | $x_i \in [0, 1]$ | $x_i \in [0.25, 0.75]$ |

Table 2 Target point, variable range, and inaccessible domain of 100D Styblinski–Tang function

| | Target point x_{target} | Variable range | Inaccessible domain |
|---------------|----------------------------------|-------------------|---------------------|
| Extrapolation | $x_i = 3$ | $x_i \in [-1, 3]$ | $x_i \in [1, 3]$ |
| Interpolation | $x_i = 1$ | $x_i \in [-1, 3]$ | $x_i \in [0, 2]$ |

3.1 Algebraic Illustration Functions. This section examines properties of the method of converging lines based on two algebraic functions, Dette function and Styblinski–Tang function, from a virtual library of simulation experiments [27]. Dette function given in Eq. (7) is an 8D algebraic function proposed by Dette and Pepelyshev [28]. This function is highly curved in some variables and less in others. The function is evaluated on the hypercube $x_i \in [0, 1]$ for all $i = 1, \dots, 8$

$$f(x) = 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 + 16\sqrt{x_3 + 1}(2x_3 - 1)^2 + \sum_{i=4}^8 \ln\left(1 + \sum_{j=3}^i x_j\right) \quad (7)$$

For extrapolation, the target point for prediction was selected to be the vertex where all the eight variables were at their upper value, $x_i = 1, i = 1, 2, \dots, 8$, as provided in Table 1. The inaccessible domain was set to have the same length as the accessible domain along all the variable directions. For interpolation, the target point was set to be $x_i = 0.5$ for $i = 1, \dots, 8$, which is the center of variable space. The inaccessible domain was set to be the hypercube centered at the target point with a width of 0.5. Therefore, the inaccessible domain had the same length as the accessible domain.

The second test function is the Styblinski–Tang function, which is a popular multimodal function for testing optimization algorithms with user-defined dimensionality as shown in Eq. (8). The dimensionality d is set to be 100 to study the effect of high-dimensionality on prediction. The function is evaluated on the hypercube $x_i \in [-1, 3]$ for all $i = 1, \dots, d$. Approximating this test

function directly is challenging for classical surrogate models considering its dimensionality

$$f(x) = \frac{1}{2} \sum_{i=1}^d [x_i^4 - 16x_i^2 + 5x_i] \quad (8)$$

For extrapolation, the target point for prediction was selected again to be the vertex, where all the variables were at their upper bounds, $x_i = 3$ as provided in Table 2. The inaccessible domain was set to be a hypercube cornered at the target point and had same length as the accessible domain along all the variable directions. For interpolation, the target point was set to be $x_i = 1$, which was the center of variable space. The inaccessible domain was set to be the hypercube centered at the target point with a width of 1. Therefore, the inaccessible domain had the same length as the accessible domain.

3.2 Method of Converging Lines Versus Multidimensional Approximation. We first compared the proposed approach with multidimensional approximation based on Dette function for extrapolation and interpolation, respectively.

3.2.1 Extrapolation. For the method of converging lines, three lines were selected toward the target point. The other ends of the lines were randomly selected from the rest of the 2^8 vertices of the domain. We selected 100 sets of lines randomly to compensate for the effect of line selection. Six samples were evenly spaced in the accessible domain. A typical set of extrapolation results based on ordinary Kriging was shown in Fig. 3. The function along line 1 in Fig. 3(a) is unimodal like a bowl. Figure 3(b) showed line 2 which is wavy. Line 3 in Fig. 3(c) was close-to-linear and most friendly for approximation. In the case in Fig. 3, combined prediction was close to line 3, which had a much smaller prediction variance. Note that even without knowledge of the true function, the ability to visualize the data and predictions shown in Fig. 3 is useful for spotting which surrogate is likely to be trusted.

For a multidimensional approximation, 18 samples were selected using LHS with 5000 iterations with the MATLAB function lhsdesign. Sampling domain was restricted to be the same as that to generate multiple lines. The inaccessible domain was a hypercube cornered at the target point.

One hundred sets of predictions were generated using multiple lines and multiple LHS designs. Prediction accuracy was quantified using absolute percent error and is summarized using boxplot in Fig. 4. Prediction from multiple lines had the error median to be 0.018%, and 75% predictions were less than 0.1%. In contrast, the minimum error of multidimensional approximation was 25%.

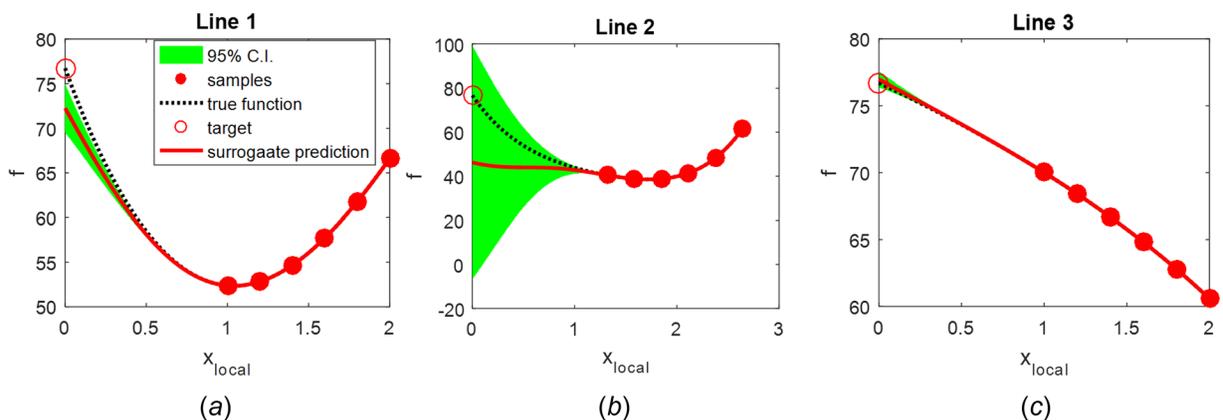


Fig. 3 A typical set of extrapolation results from 8D Dette function. The three lines are from randomly selected vertices to the target vertex.

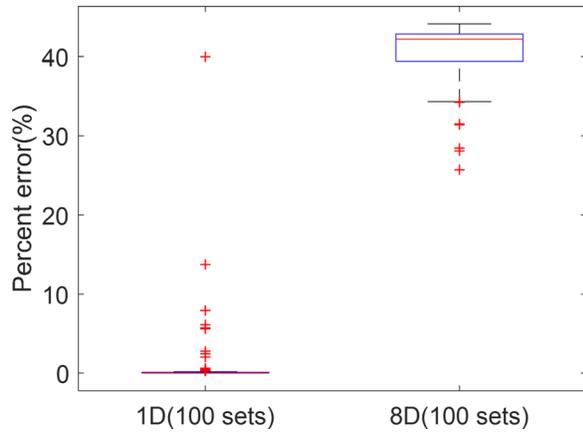


Fig. 4 Summary of extrapolation results from multiple lines and 8D Kriging based on Dette function

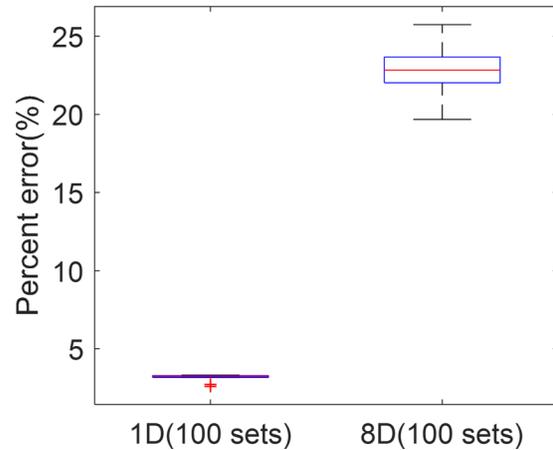


Fig. 6 Summary of interpolation results from multiple lines and 8D Kriging based on Dette function

3.2.2 Interpolation. For the method of converging lines, three lines were selected to intersect at the target point. The ends of lines were randomly selected from all the vertices. A typical set of interpolation results using Kriging is shown in Fig. 5. All the lines had a bowl shape with different minima. All the Kriging models had a similar level of prediction variance and overestimated function value at the target point slightly. Again, we note that even without knowing the exact function, figures such as Fig. 5 provide some protection against the failure of the fitting process.

For a multidimensional approximation, 18 samples were generated using LHS in the accessible domain. One hundred sets of prediction were generated to estimate randomness and summarized in Fig. 6. Prediction using multiple lines had the error median to be 3.24% and maximum error to be 3.31%. In contrast, prediction using 8D Kriging had the error median to be 22.8% and minimum error to be 19.7%. Multiple lines generated more accurate predictions compared to 8D Kriging for this test function.

3.3 Applicability to Very High-Dimensionality. We adopted Styblinski–Tang function with 100 dimensions to illustrate applicability to high-dimensionality. The dimension is too high to be approximated by any regular surrogate models with a reasonable number of samples. Therefore, the method of converging lines might be the only option for a function with such a high dimension.

For extrapolation, three lines were randomly selected from the rest of the 2^{100} vertices. Six sampling points were evenly spaced

in the accessible domain along each line. A typical set of extrapolation results is given in Fig. 7. The test function along all the three lines was multimodal. Prediction error was small at domain close to samples and increased with prediction distance in the extrapolation domain. This example illustrates the difficulty in extrapolating long distance. Not only the prediction can deteriorate with distance, but for a function that changes its behavior, even the uncertainty estimate may be flawed.

One hundred sets of random lines were generated toward same target point to compensate for variability. The error of 100 predictions is summarized in Fig. 8. The prediction error ranged around 28%. In this case, 100-dimensional surrogates are not possible, the performance of the proposed method of converging lines was not compared with multidimensional surrogates, but the prediction error would be huge if 18 samples were able to fit a 100-dimensional surrogate.

For interpolation, three lines are selected toward the target point. The other end of lines is randomly selected from all the vertices. A typical set of interpolation results is shown in Fig. 9. All the lines have a bowl shape, and the Kriging models overestimated function value at the target point. One hundred sets of prediction are generated to compensate for randomness and summarized in Fig. 10. Prediction using multiple lines had the error between 7.4% and 9.3%. Based on the tests using the 100D Styblinski–Tang function, the method of lines enables estimation of the high-dimensional function at one point using multiple lines. Such estimation is practically impossible with a small number of samples using a multidimensional surrogate. In particular, it is

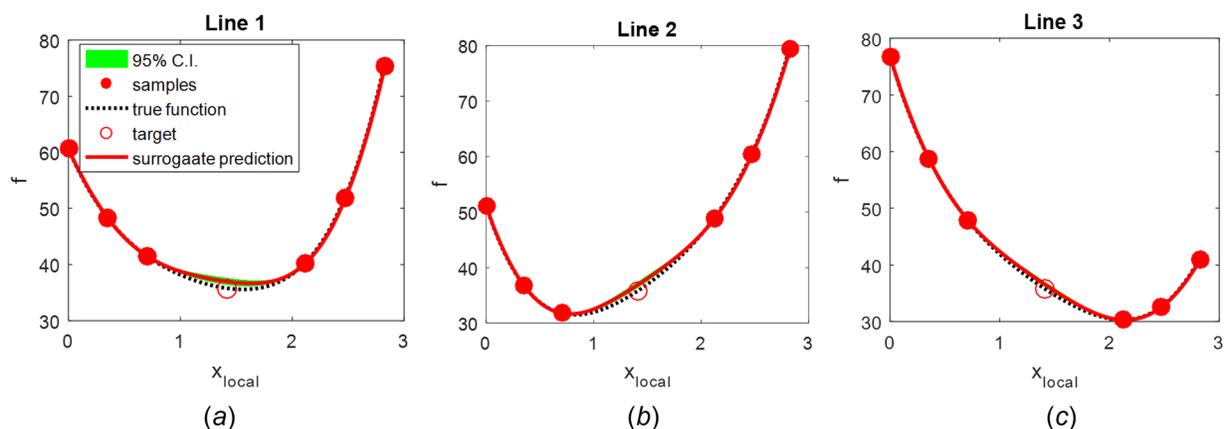


Fig. 5 A typical set of interpolation results from 8D Dette function

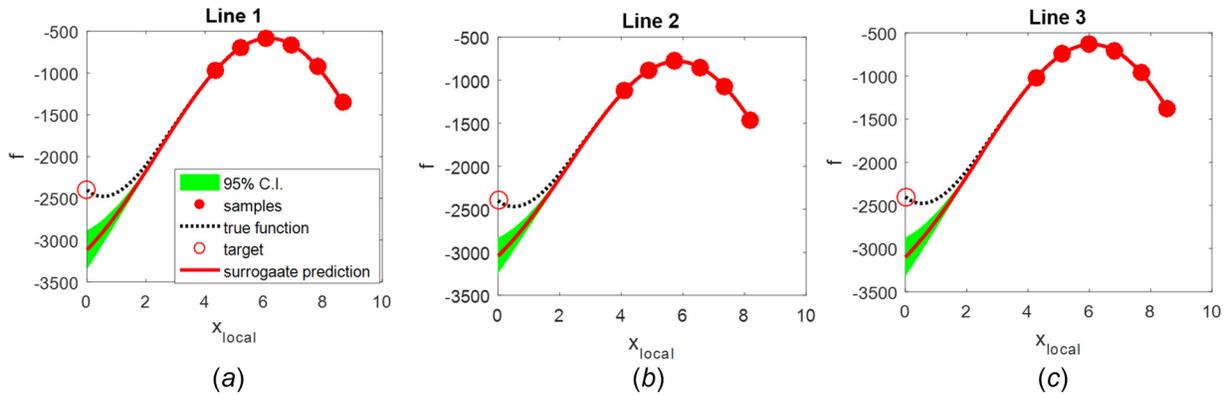


Fig. 7 A typical set of extrapolation results from 100D Styblinski–Tang function

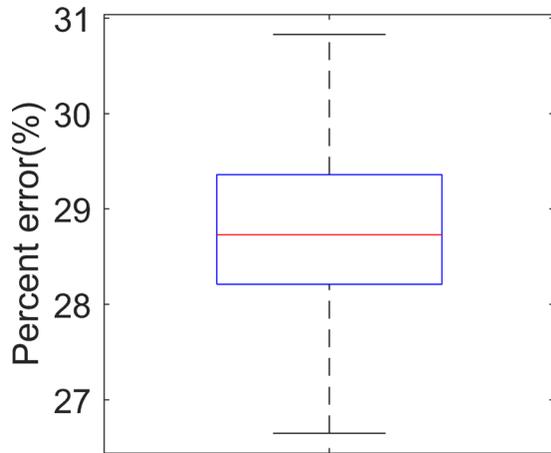


Fig. 8 One hundred sets of extrapolation results from multiple lines based on 100D Styblinski–Tang function. The target point was at $x_t = 3$ with the function value to be -2400 . One hundred sets of converging lines were randomly selected among the vertices excluding the target point.

impossible to fit a Kriging surrogate with 18 points in 100-dimensional space.

4 Long-Distance Extrapolation and Interpolation of a 2D Drag Coefficient

A drag coefficient model of a spherical particle [29] is presented in this section to demonstrate long-distance function

estimation at an inaccessible point using the method of converging lines.

4.1 Introduction to the Drag Coefficient Function. The drag coefficient is a dimensionless quantity that measures the quasi-steady drag coefficient of a spherical particle in compressible flow. For a given particle, the drag coefficient, C_D , is a function of Mach number, M , and Reynolds number, Re . The range of applicability for C_D in this paper is limited to the supersonic ($1 < M \leq 1.75$) domain to avoid the physical discontinuity at $M = 1$. The Reynolds number is limited to $Re \leq 2 \times 10^5$ to avoid turbulence of the attached boundary layer. It is considered that the governing physics is consistent within these limits. The analytical expression for C_D is given in $C_D = f(Re, M)$ in the Appendix. The dependence of the drag coefficient on Reynolds number and Mach number is shown in Fig. 11, which shows that the behavior can be made smoother by transforming Re to logarithmic coordinates. The target point, variable range for analysis, and inaccessible domain of the drag coefficient function are summarized in Table 3 and plotted in Fig. 12. For extrapolation, prediction domain is a square located on the upper right corner as shown in Fig. 12(a). For interpolation, prediction domain is set to be a square in the center of variable range as shown in Fig. 12(b). The length of sampling domain and prediction domain is the same along each line for both cases.

4.2 Extrapolation Results

4.2.1 Extrapolation of Drag Coefficient Function Using the Method of Converging Lines. Instead of sampling entire range of sampling domain, six uniform samples were selected from half of the sampling domain as shown in Fig. 12(a). This was because the prediction variance decreased noticeably when samples were

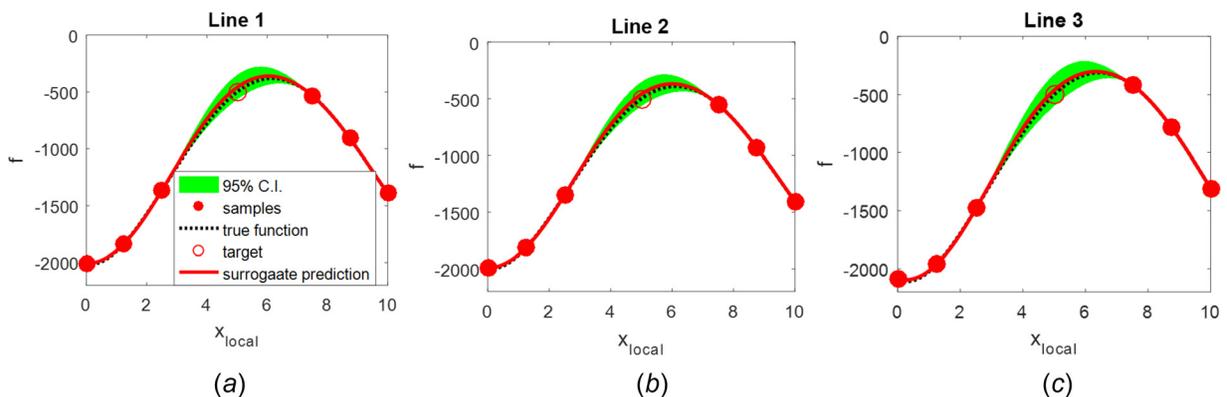


Fig. 9 A typical set of interpolation results from 100D Styblinski–Tang function

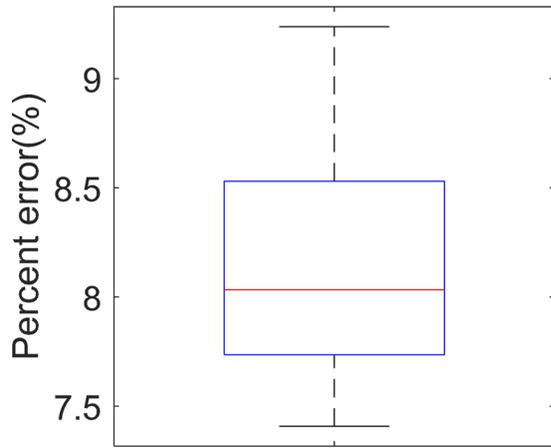


Fig. 10 One hundred sets of interpolation results from multiple lines based on 100D Styblinski–Tang function. The target point was at $x_i=1$ with the function value to be -500 . One hundred sets of converging lines were randomly selected among the vertices excluding the target point.

closer to the target point. When the inaccessible domain is in the interpolation region, the sampling scheme in Fig. 12(b) can be used, which will be considered in Sec. 4.3. Table 4 summarizes statistics of the surrogate predictions. Predictions of function value along variable M with constant Re (line 3) had the smallest prediction variance as shown in Table 4 and Fig. 13(c), indicating better accuracy along this line.

By substituting extrapolation results of each line from Table 4 into Eq. (7), most probable estimation can be obtained as $p(y_{\text{target}}|y_1(\mathbf{x}_{\text{target}}), y_2(\mathbf{x}_{\text{target}}), y_3(\mathbf{x}_{\text{target}})) \sim N(0.9755, 0.0003^2)$. Note that the posterior distribution was dominated by the prediction along line 3 because the prediction variance (σ^2) was much smaller than for the other lines. The small prediction variance denotes high reliability of extrapolation. Comparing with true function value 0.9766, absolute error is 0.001 and relative error is 0.1%. Note also that the less accurate lines (lines 1 and 2) provide a reasonable warning of their lack of accuracy in their prediction uncertainties, σ_i .

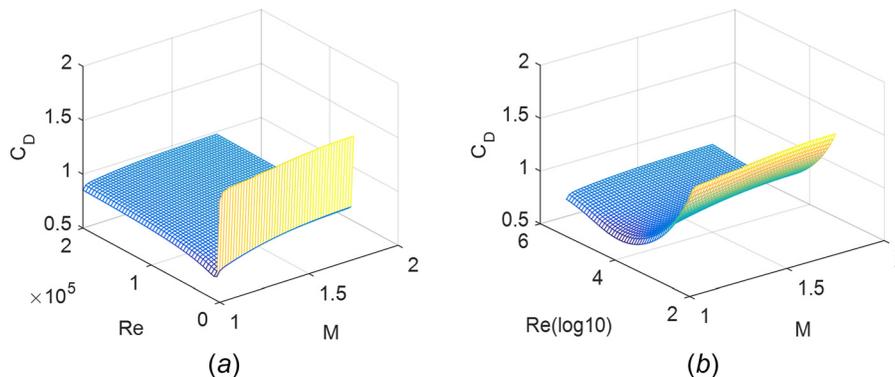


Fig. 11 Dependence of drag coefficient on Re and M : (a) natural coordinate and (b) \log_{10} coordinate

Table 3 Target point, variable range, and inaccessible domain of drag coefficient function

| | Target point $\mathbf{x}_{\text{target}}$ | Variable range | Inaccessible domain |
|---------------|---|--|---|
| Extrapolation | $\log_{10}(Re) = 5.301$ $M = 1.75$ | $\log_{10}(Re) \in [2, 5.301]$ $M \in [1, 1.75]$ | $\log_{10}(Re) \in [3.65, 5.301]$ $M \in [1.375, 1.75]$ |
| Interpolation | $\log_{10}(Re) = 3.6505$ $M = 1.375$ | $\log_{10}(Re) \in [2, 5.301]$ $M \in [1, 1.75]$ | $\log_{10}(Re) \in [2.8253, 4.4757]$ $M \in [1.1875, 1.5625]$ |

4.2.2 Comparison of 1D and 2D Surrogates. One interesting question is how sensitive the prediction is to the location of lines. The sensitivity to the selection of lines was tested by varying the angle of lines. In Fig. 12, the angle between lines and the horizontal axis with constant $M=1.75$ was set to be α_1, α_2 , and α_3 , respectively, where $\alpha_1 = 0$ deg, $\alpha_2 = 45$ deg, and $\alpha_3 = 90$ deg in the figure. In order to test the effect of lines, α_1 varied from 0 deg to 5 deg, α_2 from 40 deg to 50 deg, and α_3 from 85 deg to 90 deg in the increment of 1 deg. Three hundred and ninety six sets of three lines were generated based on different combinations of α_1, α_2 , and α_3 . Extrapolation results based on each set of lines were plotted using boxplot in Fig. 14(b). All the extrapolation results concentrated near the true function value, 0.9766, varying from 0.9751 to 0.9756.

In order to examine the performance of multidimensional extrapolation, two-dimensional (2D) Kriging is adopted. Eighteen samples are selected in the reduced sampling domain using LHS (5000 iterations maximizing minimum distance) as shown in Fig. 14(a). As converging lines benefitted from the reduced domain, 2D Kriging also benefited from it based on our tests. Two-dimensional extrapolation is repeated 20 times to compensate for the randomness of sampling pattern. For a typical 2D extrapolation out of the 20 cases, i.e., corresponding to the case with median absolute error, the prediction was $N(0.9869, 0.0313^2)$. In Fig. 14(b), we can see that boxplot of 2D Kriging prediction has a median equal to 0.9799 and standard deviation to be 0.0261. Two-dimensional Kriging generated larger variation and bias than converging lines even with 20 sets of data. The method of converging lines proved to be more reliable and robust for one-point extrapolation of drag coefficient function.

4.3 Interpolation Results

4.3.1 Interpolation of Drag Coefficient Function Using the Method of Converging Lines. Interpolation using method of converging lines was performed similarly as shown in Fig. 12(b). We selected samples in the half of sampling domain to reduce prediction variance. Interpolation results are plotted in Fig. 15 from three lines. Predictions at target point are provided in Table 5.

As in the case of extrapolation, the prediction variance on line 3 is by far the smallest. By substituting prediction results of each line from Table 5 into Eq. (7), most probable estimation can be

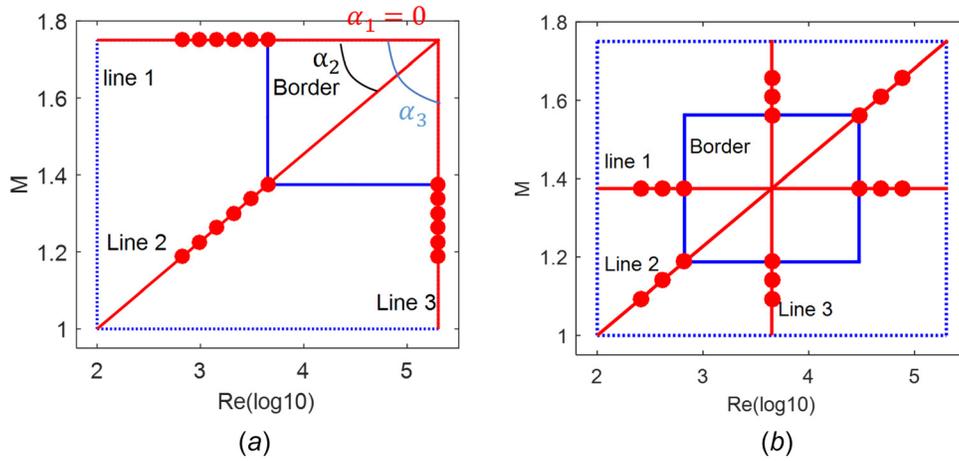


Fig. 12 Lines and samples selection for extrapolating and interpolation of drag coefficient function. The solid dots denote the location of samples. (a) Extrapolation and (b) interpolation.

Table 4 Mean value of prediction and its uncertainty given by standard deviation for extrapolation along three lines shown in Fig. 12(a)

| | Line 1 | Line 2 | Line 3 |
|------------|--------|--------|--------|
| μ_i | 1.021 | 0.9352 | 0.9755 |
| σ_i | 0.0786 | 0.0725 | 0.0003 |

obtained as: $p(y_{\text{target}} | y_1(\mathbf{x}_{\text{target}}), y_2(\mathbf{x}_{\text{target}}), y_3(\mathbf{x}_{\text{target}})) \sim N(0.9223, 0.000005^2)$, essentially identical to the result of line 3. Comparing with true function value, 0.9224, absolute error is 0.0001 and relative error is 0.01%.

4.3.2 Comparison of 1D and 2D Surrogates. We have more freedom to select lines for interpolation while fixing intersection

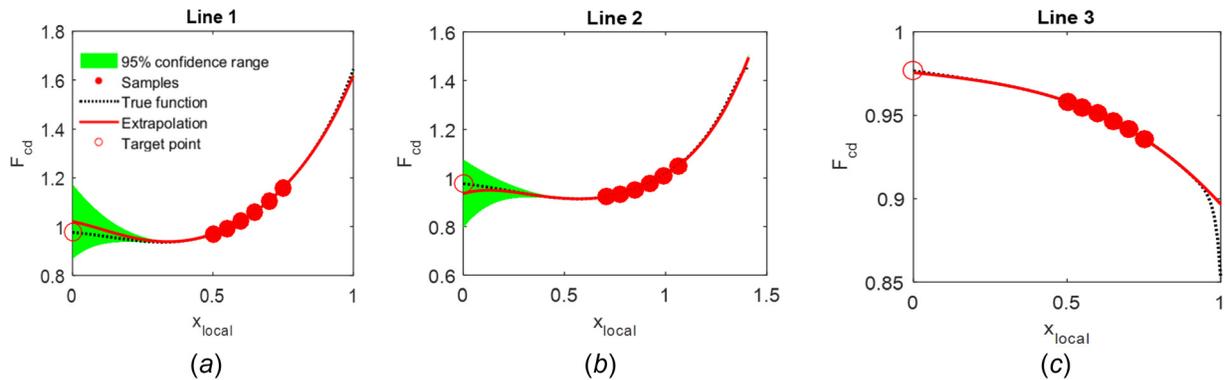


Fig. 13 Extrapolation results of drag coefficient function using samples from reduced sampling domain, as shown in Fig. 14: (a) line 1, (b) line 2, and (c) line 3

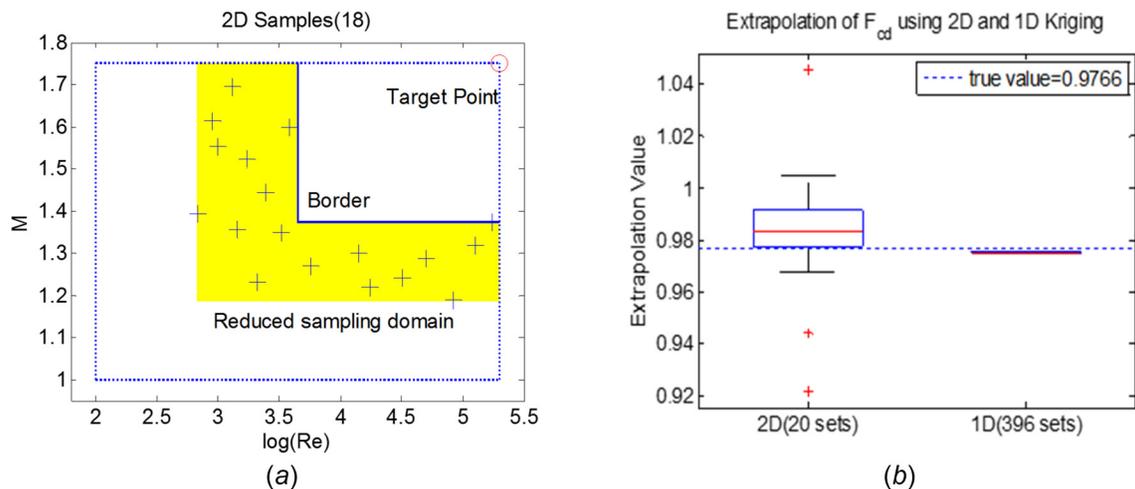


Fig. 14 Extrapolation results of drag coefficient function using samples from reduced sampling domain, as shown in Fig. 12: (a) line 1, (b) line 2, and (c) line 3

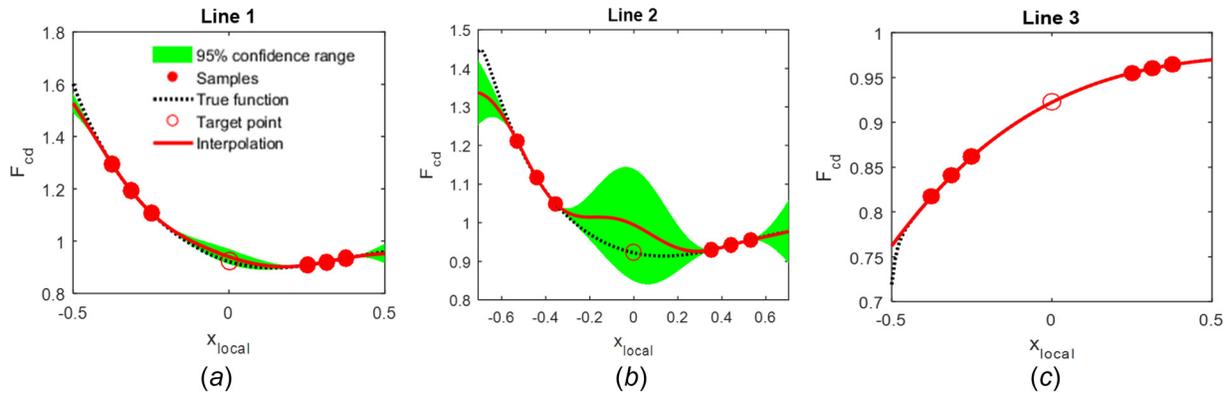


Fig. 15 Interpolation results of drag coefficient function: (a) line 1, (b) line 2, and (c) line 3

Table 5 Interpolation results along three lines

| | Line 1 | Line 2 | Line 3 |
|------------|--------|--------|--------------------|
| μ_i | 0.9408 | 0.9943 | 0.9223 |
| σ_i | 0.0146 | 0.0742 | 5×10^{-6} |

angle between lines. Therefore, the intersection angle between lines 1, 2, and 3 were fixed to be 45 deg successively. Three lines rotated counterclockwise gradually from 0 deg to 45 deg in 20 steps. Twenty sets of 1D interpolation results were then obtained accordingly. For 2D Kriging, 18 samples are selected in the sample domain using LHS (5000 iterations maximizing minimum distance for the following analysis as shown in Fig. 16(a)). For a typical 2D interpolation out of the 20 cases, i.e., corresponding to the case with median absolute error, the prediction was $N(0.9317, 0.0012^2)$. It is not only less accurate than the method of converging lines but also has a substantial underestimate of the uncertainty in the prediction in the prediction variance.

Predictions of 1D and 2D Kriging were summarized in Fig. 16(b). One-dimensional Kriging predictions varied from 0.9223 to 0.9232. Two-dimensional Kriging prediction generated a median to be 0.9255 and standard deviation equal to 0.0195. The box plot shows that for some DOEs, the 2D prediction was grossly inaccurate.

5 Conclusion

This paper presents a method for improving the accuracy of function estimation at an inaccessible point by taking advantage of the fact that prediction is needed at only one point instead of everywhere. The proposed method of converging lines transforms a multidimensional prediction into a series of 1D predictions. Most probable estimation is obtained by combining consistent predictions from different lines using Bayesian inference while assuming all the lines are independent.

The one-dimensional prediction was performed using Kriging surrogate model. Function prediction is only feasible under the assumption that the function behaves similarly in the sampling domain and prediction domain. We illustrated its applicability to high-dimensional functions with two algebraic examples, 8D Dette function and 100D Styblinski–Tang function. The proposed approach proved to be more accurate and reliable than multidimensional approximation for 8D Dette function. The proposed approach enabled moderately accurate estimation of 100D Styblinski–Tang function with 18 samples. The two examples also illustrated another advantage of the method, which is an easy visualization of the fitted surrogate. Finally, the examples illustrated that combining multiple predictions using Bayesian inference could increase the chance of obtaining valid estimation.

Next, the method was illustrated for a physical example, extrapolation and interpolation of a two-dimensional drag coefficient. For this example, as well, the method of converging lines proved to be much more reliable and accurate than 2D Kriging.

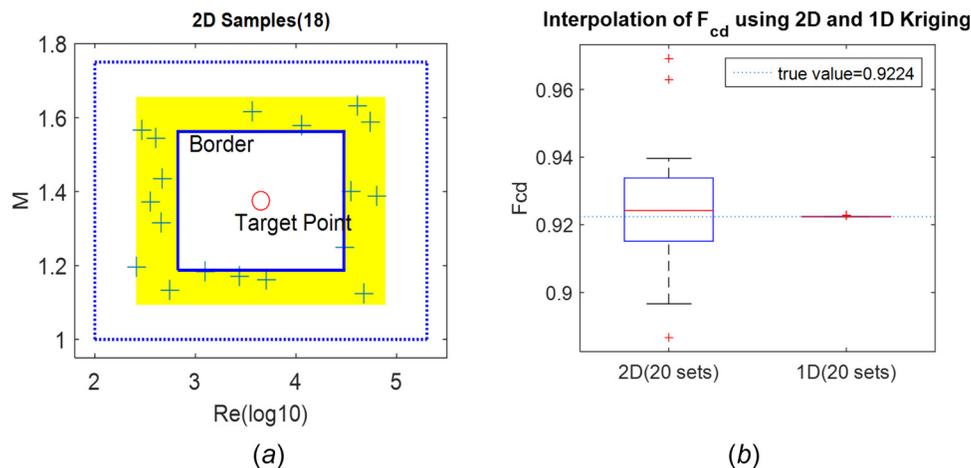


Fig. 16 Interpolation results using 2D and 1D Kriging: (a) 2D samples and (b) comparison of 2D and 1D interpolation

In this paper, we assume that predictions from multiple lines are independent to apply Bayesian methods. The correlation among lines exists to some extent especially at the domain close to the converged target point. Introducing correlation among lines will be an interesting topic for future work.

Acknowledgment

This work was 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.

Appendix: Analytical Expressions for Drag Coefficient

The analytic expression for drag coefficient function was cited from Ref. [29]. The range of applicability for f_c is $\text{Re} \leq 2 \times 10^5$, $M \leq 1.75$, and $\text{Kn} \leq 0.01$, where Kn denotes Knudsen number. $\text{Re} = 2 \times 10^5$ is the subcritical Reynolds number above which the attached boundary layer becomes turbulent. Attention is limited to continuum flows, namely, $\text{Kn} < 0.01$. C_D is given in Eq. (A1) for subcritical ($M \leq M_{cr} \approx 0.6$), supersonic ($1 < M \leq 1.75$), and intermediate ($M_{cr} < M \leq 1$) Mach number regimes

$$C_D(\text{Re}, M) = \begin{cases} C_{D,\text{std}}(\text{Re}) + [C_{D,M_{cr}}(\text{Re}) - C_{D,\text{std}}(\text{Re})] \frac{M}{M_{cr}} & \text{if } M \leq M_{cr} \\ C_{D,\text{sub}}(\text{Re}, M) & \text{if } M_{cr} < M \leq 1.0 \\ C_{D,\text{sup}}(\text{Re}, M) & \text{if } 1.0 < M \leq 1.75 \end{cases} \quad (\text{A1})$$

where

$$C_{D,\text{std}}(\text{Re}) = \frac{24}{\text{Re}} (1 + 0.15\text{Re}^{0.687}) + 0.42 \left(1 + \frac{42500}{\text{Re}^{1.16}} \right)^{-1} \quad (\text{A2})$$

$$C_{D,M_{cr}}(\text{Re}) = \frac{24}{\text{Re}} (1 + 0.15\text{Re}^{0.684}) + 0.513 \left(1 + \frac{483}{\text{Re}^{0.669}} \right)^{-1} \quad (\text{A3})$$

$$C_{D,\text{sup}}(\text{Re}, M) = C_{D,M=1}(\text{Re}) + [C_{D,M=1.75}(\text{Re}) - C_{D,M=1}(\text{Re})] \zeta_{\text{sup}}(M, \text{Re}) \quad (\text{A4})$$

$$C_{D,M=1}(\text{Re}) = \frac{24}{\text{Re}} (1 + 0.118\text{Re}^{0.813}) + 0.69 \left(1 + \frac{3550}{\text{Re}^{0.793}} \right)^{-1} \quad (\text{A5})$$

$$C_{D,M=1.75}(\text{Re}) = \frac{24}{\text{Re}} (1 + 0.107\text{Re}^{0.867}) + 0.646 \left(1 + \frac{861}{\text{Re}^{0.634}} \right)^{-1} \quad (\text{A6})$$

$$\zeta_{\text{sup}}(M, \text{Re}) = \sum_{i=1}^3 f_{i,\text{sup}}(M) \prod_{\substack{j \neq i \\ j=1}}^3 \frac{\log \text{Re} - C_{j,\text{sub}}}{C_{i,\text{sub}} - C_{j,\text{sub}}} \quad (\text{A7})$$

$$f_{1,\text{sup}}(M) = -2.963 + 4.392M - 1.169M^2 - 0.027M^3 - 0.233 \exp[(1 - M)/0.011] \quad (\text{A8})$$

$$f_{2,\text{sup}}(M) = -6.617 + 12.11M - 6.501M^2 + 1.182M^3 - 0.174 \exp[(1 - M)/0.01] \quad (\text{A9})$$

$$f_{3,\text{sup}}(M) = -5.866 + 11.57M - 6.665M^2 + 1.312M^3 - 0.350 \exp[(1 - M)/0.012] \quad (\text{A10})$$

$$C_{1,\text{sup}} = 6.48; \quad C_{2,\text{sup}} = 8.93; \quad C_{3,\text{sup}} = 12.21 \quad (\text{A11})$$

$$C_{D,\text{sub}}(\text{Re}, M) = C_{D,M_{cr}}(\text{Re}) + [C_{D,M=1}(\text{Re}) - C_{D,M_{cr}}(\text{Re})] \zeta_{\text{sub}}(M, \text{Re}) \quad (\text{A12})$$

$$\zeta_{\text{sub}}(M, \text{Re}) = \sum_{i=1}^3 f_{i,\text{sub}}(M) \prod_{\substack{j \neq i \\ j=1}}^3 \frac{\log \text{Re} - C_{j,\text{sub}}}{C_{i,\text{sub}} - C_{j,\text{sub}}} \quad (\text{A13})$$

$$f_{1,\text{sub}}(M) = -1.884 + 8.422M - 13.70M^2 + 8.162M^3 \quad (\text{A14})$$

$$f_{2,\text{sub}}(M) = -2.228 + 10.35M - 16.96M^2 + 9.840M^3 \quad (\text{A15})$$

$$f_{3,\text{sub}}(M) = 4.362 - 16.91M + 19.84M^2 - 6.296M^3 \quad (\text{A16})$$

$$C_{1,\text{sub}} = 6.48; \quad C_{2,\text{sub}} = 9.28; \quad C_{3,\text{sub}} = 12.21 \quad (\text{A17})$$

References

- [1] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., 2005, "Surrogate-Based Analysis and Optimization," *Prog. Aerosp. Sci.*, **41**(1), pp. 1–28.
- [2] Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., and Wu, C. J., 2006, "Building Surrogate Models Based on Detailed and Approximate Simulations," *ASME J. Mech. Des.*, **128**(4), pp. 668–677.
- [3] Brooks, D. G., Carroll, S. S., and Verdini, W. A., 1988, "Characterizing the Domain of a Regression Model," *Am. Stat.*, **42**(3), pp. 187–190.
- [4] Barnard, E., and Wessels, L., 1992, "Extrapolation and Interpolation in Neural Network Classifiers," *IEEE Control Syst.*, **12**(5), pp. 50–53.
- [5] Shan, S., and Wang, G. G., 2010, "Metamodeling for High Dimensional Simulation-Based Design Problems," *ASME J. Mech. Des.*, **132**(5), p. 051009.
- [6] Glaz, B., Goel, T., Liu, L., Friedmann, P. P., and Haftka, R. T., 2007, "Application of a Weighted Average Surrogate Approach to Helicopter Rotor Blade Vibration Reduction," *AIAA Paper No. 2007-1898*.
- [7] Balabanov, V., Weckner, O., and Wu, J., 2014, "Reducing Error of Polynomial Approximation Outside of Designated Design Space for Practical Problems," *AIAA Paper No. 2014-2303*.
- [8] Hooker, G., and Rosset, S., 2012, "Prediction-Based Regularization Using Data Augmented Regression," *Stat. Comput.*, **22**(1), pp. 237–249.
- [9] Wilson, A. G., and Adams, R. P., 2013, "Gaussian Process Kernels for Pattern Discovery and Extrapolation," 30th International Conference on Machine Learning (ICML), Atlanta, GA, June 16–21, pp. 1067–1075.
- [10] Roache, P. J., 1998, *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, NM.
- [11] Sidi, A., 2003, *Practical Extrapolation Methods: Theory and Applications*, Vol. 10, Cambridge University Press, Cambridge, UK.
- [12] Hooker, G., 2004, "Diagnostics and Extrapolation in Machine Learning," *Ph.D. thesis*, Stanford University, Stanford, CA.
- [13] An, D., Kim, N. H., and Choi, J.-H., 2015, "Practical Options for Selecting Data-Driven or Physics-Based Prognostics Algorithms With Reviews," *Reliab. Eng. Syst. Saf.*, **133**, pp. 223–236.
- [14] George, J. J., 2014, *Weather Forecasting for Aeronautics*, Academic Press, New York.
- [15] Rapach, D. E., and Zhou, G., 2013, "Forecasting Stock Returns," *Handbook of Economic Forecasting*, Vol. 2(Pt. A), Elsevier, North Holland, The Netherlands, pp. 328–383.
- [16] Chaudhuri, A., and Haftka, R. T., 2014, "Optimistic Bias in Surrogate Prediction Near Surrogate Optima," *AIAA paper No. 2014-2438*.
- [17] Durelli, A., and Rajaiiah, K., 1981, "Quasi-Square Hole With Optimum Shape in an Infinite Plate Subjected to In-Plane Loading," *ASME J. Mech. Des.*, **103**(4), pp. 866–870.
- [18] Stein, M., 1987, "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, **29**(2), pp. 143–151.
- [19] Zhang, Y., Kim, N.-H., Park, C.-Y., and Haftka, R. T., 2015, "One-Dimensional Function Extrapolation Using Surrogates," 11th World Congress on Structural and Multidisciplinary Optimisation (WCSMO), Sydney, Australia, June 7–12, pp. 1–6.
- [20] Rasmussen, C. E., 2004, "Gaussian Processes for Machine Learning," *Advanced Lectures on Machine Learning*, Springer, Berlin, pp. 63–71.
- [21] Viana, F. A. C., 2011, "SURROGATES Toolbox User's Guide," Gainesville, FL, Version 3.0 <https://sites.google.com/site/srgtstoolbox/>

- [22] Bernardo, J., Berger, J., Dawid, A., and Smith, A., 1992, "Some Bayesian Numerical Analysis," *Bayesian Statistics*, Vol. 4, Oxford University Press, Oxford, UK, pp. 345–363.
- [23] Goel, T., Haftka, R. T., Shyy, W., and Queipo, N. V., 2007, "Ensemble of Surrogates," *Struct. Multidiscip. Optim.*, **33**(3), pp. 199–216.
- [24] Viana, F. A., Haftka, R. T., and Steffen, V., Jr., 2009, "Multiple Surrogates: How Cross-Validation Errors Can Help Us to Obtain the Best Predictor," *Struct. Multidiscip. Optim.*, **39**(4), pp. 439–457.
- [25] Murphy, K. P., 2007, "Conjugate Bayesian Analysis of the Gaussian Distribution," [Technical Report](#).
- [26] Box, G. E., and Tiao, G. C., 2011, *Bayesian Inference in Statistical Analysis*, Vol. 10, Wiley, New York.
- [27] Surjanovic, S., and Bingham, D., 2013, "Virtual Library of Simulation Experiments: Test Functions and Datasets," Simon Fraser University, Burnaby, BC, Canada, accessed May 13, 2015, from <http://www.sfu.ca/~ssurjano>
- [28] Dette, H., and Pepelyshev, A., 2012, "Generalized Latin Hypercube Design for Computer Experiments," *Technometrics*, **50**(4), pp. 421–429.
- [29] Parmar, M., Haselbacher, A., and Balachandar, S., 2010, "Improved Drag Correlation for Spheres and Application to Shock-Tube Experiments," *AIAA J.*, **48**(6), pp. 1273–1276.