Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Stochastic Polynomial Interpolation for Uncertainty Quantification With Computer Experiments

Matthias Hwai Yong TAN
Department of Systems Engineering and Engineering Management, 
City University of Hong Kong, Hong Kong
(matthtan@cityu.edu.hk)

Multivariate polynomials are increasingly being used to construct emulators of computer models for uncertainty quantification. For deterministic computer codes, interpolating polynomial metamodels should be used instead of noninterpolating ones for logical consistency and prediction accuracy. However, available methods for constructing interpolating polynomials only provide point predictions. There is no known method that can provide probabilistic statements about the interpolation error. Furthermore, there are few alternatives to grid designs and sparse grids for constructing multivariate interpolating polynomials. A significant disadvantage of these designs is the large gaps between allowable design sizes. This article proposes a stochastic interpolating polynomial (SIP) that seeks to overcome the problems discussed above. A Bayesian approach in which interpolation uncertainty is quantified probabilistically through the posterior distribution of the output is employed. This allows assessment of the effect of interpolation uncertainty on estimation of quantities of interest based on the metamodel. A class of transformed space-filling design and a sequential design approach are proposed to efficiently construct the SIP with any desired number of runs. Simulations demonstrate that the SIP can outperform Gaussian process (GP) emulators. This article has supplementary material online.

KEY WORDS: Bayesian approach; Design and analysis of experiments; Deterministic simulations; Interpolation uncertainty; Polynomial metamodels.

1. INTRODUCTION

Due to advances in computing, there is widespread interest in the development and use of engineering and physical models. Many of the models are formulated as differential equations, and solved numerically using methods such as the finite element method. One of the main purposes of these models is to predict one or more outputs as a function of a set of inputs. However, because solving the models is time consuming, it is difficult to use the models to predict the outputs at a large number of input settings, which is essential for purposes such as uncertainty propagation, sensitivity analysis, robust design, and optimization. Thus, metamodels are constructed using data from a computer experiment to approximate the computer output.

Polynomial metamodels are used in almost every field of engineering (mechanical design, manufacturing, aeronautics, electronics, bioengineering, nuclear engineering, etc.) as surrogates of complex computer models. They are advantageous for some applications due to their simplicity. Response surface methodology (Box and Draper 2007) has been used in computer experimentation to build low-order polynomials metamodels for prediction and optimization (Simpson et al. 1998; Wang and Shan 2007). Nonintrusive uncertainty quantification methods, which are methods that treat the computer model as a black box, are currently receiving a lot of attention from researchers. Many of the nonintrusive methods use polynomial metamodels. Since the seminal work of Ghanem and Spanos (1991), there is increasing interest in the use of polynomial chaos methods for approximating solutions of stochastic differential equations. Polynomial chaos methods approximate the computer output using polynomials that are orthogonal with respect to the input distribution (Xiu and Karniadakis 2002). Consider the case with input vector $u$: if $\{\psi_0, \psi_1, \ldots\}$ is a sequence of polynomials that are orthogonal with respect to the density $w$ of $u$, then $Y(u) = \sum_{\alpha=0}^{\infty} \beta_{\alpha} \psi_{\alpha}(u)$ is a polynomial chaos expansion of $Y$. In practice, truncated polynomial chaos expansions with estimated coefficients are employed. This gives simple approximations of the output with very useful properties that can significantly simplify the tasks of global sensitivity analysis (Sudret 2008; Crestaux, Le Maître, and Martinez 2009) and robust parameter design (Welch et al. 1990; Chen, Jin, and Sudjianto 2006). The use of truncated polynomial chaos expansions can in these applications allow straightforward computation of the functional analysis of variance (ANOVA) decomposition, sensitivity indices, and mean and variance models.

There are two types of approaches for constructing polynomial chaos expansions: the stochastic Galerkin and collocation approaches (Le Maître and Knio 2010; Xiu 2010). The construction of polynomial chaos expansions or interpolating polynomial metamodels via computer experiments for approximating the solution of stochastic differential equations is called stochastic collocation. This approach has three variants: least squares, interpolation, and projection. Some authors (e.g., Le Maître...
Polynomial interpolation has traditionally been a research area of applied mathematicians. Univariate polynomial interpolation (Phillips 2003; Trefethen 2013) is a well-understood subject. However, many issues in multivariate polynomial interpolation are still not well understood. Research on this subject has focused on three areas: characterization and construction of point sets so that a unique polynomial interpolates the data (Chung and Yao 1977; Narayan and Xiu 2013), selection of an interpolating polynomial space given a set of points (De Boor and Ron 1990; Narayan and Xiu 2012), and the use of sparse grids (Barthelmann, Novak, and Ritter 2000). Interpolation error bounds have also been derived (Sauer and Xu 1995; Barthelmann, Novak, and Ritter 2000). However, these bounds are not useful in practice because they depend on the function being interpolated. A review of the recent literature on multivariate polynomial interpolation is given in Gasca and Sauer (2000).

In the statistical literature, the stationary Gaussian process (GP) model with continuous or differentiable sample paths (Sacks et al. 1989; Currin et al. 1991) is widely used as a metamodel. It captures interpolation uncertainty as an infinite-dimensional posterior process and is one of a handful of models that capture the uncertainty probabilistically. Articles that address the problem of constructing interpolating or nearly interpolating polynomial metamodels include Pistone and Wynn (1996), Holliday et al. (1999), Bates, Giglio, and Wynn (2003), and Bates, Maruri-Aguilar, and Wynn (2014). However, these polynomial interpolators do not allow quantification of interpolation uncertainty.

Despite the popularity and usefulness of interpolating polynomials for approximating deterministic computer models, statistical methods for quantifying polynomial interpolation uncertainty and designing experiments to reduce this uncertainty are lacking. This research develops a stochastic interpolating polynomial (SIP) model, which is an orthonormal polynomial-based model with the capability to capture interpolation uncertainty through the Bayesian posterior of the output. It enables probabilistic statements to be made about interpolation errors based on the data. The SIP can be used to provide point estimates and credible intervals for the output and other quantities such as sensitivity indices that depend on the output. This allows assessment of the effect of interpolation uncertainty, which is extremely important in practice. It is seldom possible to reduce interpolation uncertainty to a negligible level because the computer runs are expensive (which is the reason for constructing the metamodel in the first place). Practitioners must take into account interpolation uncertainty when basing decisions on metamodel predictions or risk being misled by the metamodel. For example, it is demonstrated in Apley, Liu, and Chen (2006) and Tan and Wu (2012) that ignoring interpolation uncertainty can give misleading results in robust parameter design.

A class of transformed space-filling design is proposed for constructing the SIP. These designs can be of any size, which gives the experimenter flexibility in choice of design size. A sequential design approach that reduces interpolation uncertainty adaptively as more data becomes available is also developed. The approach chooses follow-up runs one at a time to eliminate output uncertainty at points with maximum prediction variance. In contrast, the sparse grid method, which is widely used to construct interpolating polynomials, requires large design sizes for high degree polynomial interpolation and contain large gaps between allowable sizes. Consequently, the SIP can require less computation time and cost compared to sparse grids to achieve the same level of accuracy.

This article is organized as follows. Section 2 introduces the SIP model. Section 3 discusses specification of prior distributions for the model parameters. Section 4 proposes initial designs for uni-dimensional and multi-dimensional input spaces. In Section 4.1, the performance of the SIP is also compared with Lagrange interpolation for univariate input spaces. A sequential design approach is introduced in Section S2 of the online supplement. Section 5 compares the performance of the SIP with the GP model in multi-dimensional input spaces using test functions that include real physical models. Concluding remarks are given in Section 6.

2. STOCHASTIC INTERPOLATING POLYNOMIAL

The SIP is a polynomial that interpolates the data. Despite the lack of experimental error, interpolation uncertainty can be quantified through a Bayesian approach that introduces a prior for the functional relationship between outputs and inputs, as in GP modeling. This approach is taken in the proposed research. While it is possible to work with almost any polynomial basis functions, we shall work with polynomials that are orthonormal with respect to probability distributions as this is convenient and useful. Let $w_1, \ldots, w_d$ be probability density functions and $\{\psi_j : j \in N_0 = \{0, 1, \ldots\}\}$, where $\psi_j$ is a polynomial of degree $j$, be the sequence of polynomial basis functions in $R$ that are orthonormal with respect to the probability density $w_i$. These functions can be easily generated with a three-term recurrence relation (Gautschi 2004). Then, $\{\psi_{\alpha} = \prod_{j=1}^d \psi_{\alpha_j} : \alpha = (\alpha_1, \ldots, \alpha_d) \in N_0^d\}$ is a set of polynomial basis functions in $R^d$ that are orthonormal with respect to $w = \prod_{j=1}^d w_j$. Let $\chi \subset R^d$ be the design region, $\mathbf{u} = (u_1, \ldots, u_d) \in \chi$ be the possibly centered and scaled computer input, and $N_+(d, P) = \{(\alpha_1, \ldots, \alpha_d) \in N_0^d : 0 < \sum_{j=1}^d \alpha_j \leq P\}$. The symbol $N_+(d, P)$ refers to a set of $d$-tuples of nonnegative integers whose total sum is smaller than or equal to $P$. This corresponds to exponents of monomials whose degree does not exceed $P$. We model the computer output as

$$ Y(u) = \mu + \beta_0 + \sum_{\alpha \in N_+(d, P)} \beta_{\alpha} \psi_{\alpha}(u),$$

which is a polynomial of degree smaller than or equal to $P$. The parameter $\mu$ is introduced for convenience in writing the prior for the $\beta_{\alpha}$’s later (so that the prior mean of all $\beta_{\alpha}$’s can be set to
zero). This is equivalent to omitting $\mu$ in (1) and setting the prior mean of $\beta_0$ to $\mu$. There are a total of $q = (d + P)!/(d!P!)$ polynomial coefficients (i.e., the $\beta_i$’s including $\beta_0$).

It is often convenient and advantageous to use $\psi_\alpha$’s such that $w = \prod_{i=1}^d w_i$ is the probability density of the input $u$ (assumed to have independent components) and to take $\chi = \prod_{i=1}^d \chi_i$, where $\chi_i$ is the support of $w_i$, for the purposes of uncertainty propagation and sensitivity analysis. In this case, a useful property of the model (1) is that if $Y = Y_0 + \sum_{i=1}^d Y_i + \sum_{j=1}^{d-1} \sum_{i=1}^d Y_{i,j} + \cdots + Y_{1,\ldots,d}$ denotes the functional ANOVA decomposition of $Y$ (where $Y_0(\theta)$ is the mean, $Y_{ij}(\theta)$ is the main effect of input $i$, $Y_{ij}(\theta)$ is the interaction between inputs $i$ and $j$, etc.), then

$$Y_{i_1,\ldots,i_d} = \sum_{\alpha \in F(i_1,\ldots,i_d)} \beta_\alpha \psi_\alpha,$$  

(2)

where $F(i_1,\ldots,i_d) = \{(\alpha_1,\ldots,\alpha_d) \in \mathbb{N}_+(d, P) : \alpha_i > 0 \text{ if and only if } i \in \{i_1,\ldots,i_d\}\}$ (Sudret 2008). It follows that the variance components of $Y$ are given by

$$V_{i_1,\ldots,i_d} = \sum_{\alpha \in F(i_1,\ldots,i_d)} \beta_\alpha^2.$$  

(3)

The simple formulas (2) and (3) for the functional ANOVA and variance decompositions are key advantages of the model (1) over the GP model because for the GP model, computation of these decompositions requires computation of iterated multi-dimensional integrals in general (Oakley and O’Hagan 2004). Note that if the purpose of the computer experiment is simply to build a metamodel of the form (1) to predict the computer model on $\chi = [-1,1]^d$ (which is a common objective), we suggest either choosing the $\psi_\alpha$’s to be tensor products of orthonormal Legendre polynomials or choosing the $\psi_\alpha$’s to be tensor products of orthonormal Chebyshev polynomials. These choices give good results (see Section 5). For example, consider the case where $d = 1$ and $w_1$ is the uniform distribution over $[-1,1]$, then $\psi_0(u) = 1$, $\psi_1 = \sqrt{3}u_1$, $\psi_2 = \sqrt{3}(\frac{1}{2}(3u_1^2 - 1))$, $\psi_3 = \sqrt{3}\frac{1}{2}(5u_1^2 - 3u_1)$, $\ldots$ are orthonormal Legendre polynomials. If $d = 2$, both $w_1$ and $w_2$ are uniform distribution’s over $[-1,1]$ and $P = 2$, then the model (1) is

$$Y(u) = \mu + \beta_{0,0}\sqrt{3}u_1 + \beta_{0,1}\sqrt{3}u_2 + \beta_{2,0}\sqrt{3}\frac{1}{2}(3u_1^2 - 1) + \beta_{0,2}\sqrt{3}\frac{1}{2}(3u_2^2 - 1) + \beta_{1,1}\sqrt{3}u_1\sqrt{3}u_2.$$  

(4)

By (1), we have the main effect $Y_{(1)} = \beta_{1,0}\sqrt{3}u_1 + \beta_{2,0}(\sqrt{3}/2)(3u_1^2 - 1)$ and interaction $Y_{(1,2)} = \beta_{1,1}\sqrt{3}u_1\sqrt{3}u_2$.

Let $D = \{u_1,\ldots,u_n\} \subset \chi$ be a design and $Y = (Y_1,\ldots,Y_m)^T$ be the vector of computer outputs at the design points. By (1), $Y = \mu \mathbf{1}_n + X\beta$, where $\mathbf{1}_n$ is an $n \times 1$ vector of 1’s, $X$ is the model matrix with elements $\psi_\alpha(u_i)$ in row $i$, and $\beta$ is the $q \times 1$ vector of the $\beta_\alpha$’s. We assume that $P$ is chosen large enough so that $q > n$ and $D$ is chosen in such a way that $X$ has full row rank. Let $\beta | \theta \sim N(0, \sigma^2 S)$ (a normal distribution with mean 0 and nonsingular covariance matrix $\sigma^2 S$), where $S$ depends on the parameter $r$ and, $\theta = (\mu, \sigma^2, r)^T$ is a vector of hyperparameters to be estimated. We introduce the parameter $r$ to allow realistic modeling of $\beta$, as will be discussed in Section 3 (see also Joseph 2006). This is to ensure that the induced prior for $Y$ generates well-behaved surfaces with high probability and can be tuned (with an estimate of $r$) to produce accurate posterior predictions. Let $u^1,\ldots,u^m$ be arbitrary points in $\chi$, and $Y_0 = (Y(u^1),\ldots,Y(u^m))^T$. By (1), $Y_0 = \mu \mathbf{1}_m + X_0\beta$, where $X_0$ has elements $\psi_\alpha(u^i)$ in row $i$. The posterior predictive distribution $Y_0 | (Y, \theta)$ can be derived from the joint distribution of $Y_0$ and $Y$ given $\theta$, which is

$$Y_0 | (Y, \theta) \sim N \left[ \mu \mathbf{1}_m + \left( X_0 \mathbf{1}_n \right) \beta, \sigma^2 \left( X_0 S X_0^T \right) \right].$$  

(5)

Note that (5) is a degenerate normal distribution (which has a singular covariance matrix) for some choices of $u^1,\ldots,u^m$ and $m$ large enough. However, due to the assumption that $X$ has full row rank and $S$ is nonsingular, $Y_0 | \theta$ is a nondegenerate normal distribution, which is necessary to avoid existence of linear combinations of the $Y_i$’s with zero variance that may be inconsistent with the data. By Theorem 1.2.11 (p. 12) of Muirhead (2005), which holds for the degenerate normal case, $Y_0 | (Y, \theta)$ is a normal distribution with mean

$$E(Y_0 | Y, \theta) = \mu \mathbf{1}_m + (X_0 S X_0^T)^{-1}(X_0 Y - \mu \mathbf{1}_m),$$  

(6)

and covariance matrix

$$\text{cov}(Y_0 | Y, \theta) = X_0 [S - (X_0 S X_0^T)^{-1}(X_0 S)] X_0^T \sigma^2.$$  

(7)

It can be seen that if $m = n$ and $u^i = u_i, i = 1,\ldots,n$, then $Y_0 | (Y, \theta) = Y$ with probability 1 because setting $X_0 = X$ in (6) and (7) gives $E(Y_0 | Y, \theta) = Y$ and $\text{cov}(Y_0 | Y, \theta) = 0$. Thus, the SIP possesses the interpolation property. Choose $m$ and $u^1,\ldots,u^m$ so that $X_0$ has full column rank. Then, since $Y_0 = \mu \mathbf{1}_m + X_0\beta$, we obtain $\beta = (X_0^T X_0)^{-1}X_0^T (Y_0 - \mu \mathbf{1}_m)$, which gives

$$\beta | (Y, \theta) \sim N((X_0 S X_0^T)^{-1}(Y - \mu \mathbf{1}_m), \sigma^2 [S - (X_0 S X_0^T)^{-1}(X_0 S)]).$$  

(8)

The degenerate normal posterior distribution of $\beta$ given by (8) is needed for quantifying uncertainty in estimating the functional ANOVA and variance decompositions via (2) and (3). Consider the case of interpolating a function with two inputs at the four points $D = \{0.24, 0.82, -0.94, -0.63, -0.11, -0.87, 0.99, 0.13\}$ with the SIP model (4). We find that $X$ is a matrix with six columns and full row rank (rank four). Let $S$ be the $6 \times 6$ identity matrix, $m = 1$, and $u^i = (u_1^i, u_2^i)$. Then, $X_0 = (1,\sqrt{3}u_1^1,\sqrt{3}u_2^1,\sqrt{3}\frac{1}{2}(3u_1^1)^2 - 1,\sqrt{3}\frac{1}{2}(3u_2^1)^2 - 1,\sqrt{3}\frac{1}{2}(3u_1^2)^2)$ and (6) gives $E(Y_0 | Y, \theta) = \mu + b_0 + b_{1,0}\sqrt{3}u_1^2 + b_{0,1}\sqrt{3}u_2^2 + \sqrt{3}\frac{1}{2}(3u_1^2)^2 - 1 + b_{0,2}\sqrt{3}\frac{1}{2}(3u_2^2)^2 - 1 + b_{1,1}\sqrt{3}u_1^2\sqrt{3}u_2^2,$ (where $(b_{0,0}, b_{0,1}, b_{1,0}, b_{2,0}, b_{0,2}, b_{1,1}) = X^T (X_0 S X_0^T)^{-1}(Y - \mu \mathbf{1}_m)$, if $\mu = 0$ and the true function is $0.577\sqrt{3}u_1 + 1.155\sqrt{3}u_2$, then $(b_{0,0}, b_{0,1}, b_{1,0}, b_{2,0}, b_{0,2}, b_{1,1}) = (0.074, 0.543, 1.156,$
−0.051, −0.092, −0.008). Thus, the SIP accurately captures the linear trend.

One possible choice of the parameter \( \theta \) is its maximum likelihood estimate (MLE). Alternatively, we can use the maximum a posteriori estimate (MAPE) (see Robert 2007, p. 166). Let \( p(\theta) \) be the prior density of \( \theta \). The MAPE \( \hat{\theta} \) of \( \theta \) is obtained by maximizing

\[
p(\theta|Y) \propto p(Y|\theta)p(\theta) = p_N(Y; \mu_1, \sigma^2 X S X^T)p(\theta),
\]

where \( p_N(Y; \mu, \Sigma) \) denotes the pdf of the normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). If \( p(\theta) \propto p(r) \), that is, \( p(\mu, \sigma^2|r) \propto 1 \), it can be shown that for fixed \( r \), setting \( \mu \) and \( \sigma^2 \) to

\[
\hat{\mu} = \left[ X S X^T \right]^{-1} 1_n^T Y, \quad \hat{\sigma}^2 = \frac{1}{n} \left( Y - \hat{\mu} 1_n^T \right)^T \left( Y - \hat{\mu} 1_n \right),
\]

maximize \( p(\theta|Y) \). Thus, the MAPE \( \hat{\theta} \) of \( r \) can be found by minimizing the profile function \( -2 \ln p(Y|\theta) = (\mu, \sigma^2)^T p(r) \). Apart from constants that do not depend on \( r \), the profile function is

\[
\Lambda(r) = n \ln \hat{\sigma}^2 + 2 \ln |X S X^T| - 2 \ln p(r).
\]

This section discusses the specification of the prior \( p(r) \) of \( r \) and the covariance matrix \( S \) of the prior \( \beta|\theta \sim N(0, \sigma^2 S) \). A procedure for selecting \( P \) is also described.

A key to making the SIP work in practice is to use a sparse parameterization of \( S \) that can be tuned to produce well-behaved low-curvature functions (which are frequently observed in practice) with high probability. A sparse parameterization of \( S \) is crucial to allow tractable and reliable estimation of the parameters of \( S \). For simplicity, we assume that the \( \beta_i \)’s are a priori independent. Since high-degree polynomial basis functions tend to contribute more to the curvature of the metamodel and higher order terms in Taylor expansions of functions tend to have decreasing importance, it makes sense to use a prior for \( \beta \) such that \( \beta_\alpha \) has larger variance than \( \beta_{\alpha'} \) when \( \alpha_i \leq \alpha'_i \) (\( \alpha_i \) is the \( i \)th component of \( \alpha \) and \( \alpha'_i \) is the \( i \)th component of \( \alpha' \)) for all \( i \) and \( \alpha_k < \alpha'_k \) for some \( k \). Moreover, since different inputs tend to have different degrees of influence on the output, it is sensible to use a prior for \( \beta_\alpha \) such that the decrease of the variance of \( \beta_\alpha \) with an increase in \( \alpha_i \) depends on \( i \). Thus, a suitable prior for \( \beta_\alpha \) is

\[
\beta_\alpha \sim N \left( 0, \sigma^2 \prod_{i=1}^d r_i^{\alpha_i} \right),
\]

where \( r_i \in (0, 1) \). Note that if \( \alpha_i < \alpha'_i \) for all \( i \) and \( \alpha_k < \alpha'_k \) for some \( k \), then clearly \( \prod_{i=1}^d r_i^{\alpha_i} < \prod_{i=1}^d r_i^{\alpha'_i} \) and that if \( r_i \) is close to zero, then the \( i \)th input is inert. A similar prior for the regression coefficients for two-level factorial designs is derived by Joseph (2006). Other priors such as mixture priors may be used but they will not be considered in this article. Because orthonormal polynomials are used as basis functions in (1), the contribution of each term \( \beta_\alpha \psi_\alpha(u) \), \( \alpha \in \mathbb{N}_+(d, P) \) to the variance of the output is the square of its coefficient when \( u \) has density \( w \). This makes the magnitude of all \( \beta_\alpha \)’s comparable so that the prior (12) can be used.

In this article, we use independent priors for \( \mu, \sigma^2 \), and \( r_1, \ldots, r_d \). We set \( p(\mu) \propto 1, p(\sigma^2) \propto 1 \), and employ Beta priors for the \( r_i \)’s. Thus,

\[
p(\theta) \propto \prod_{i=1}^d r_i^{\alpha_i-1} (1 - r_i)^{\beta_i-1}.
\]

We have found that when data are scarce, it is better to use larger values of \( r \) so that the credible intervals have good coverage. Therefore, we choose \( p(\theta) \propto \prod_{i=1}^d r_i \) to pull the MAPEs of the \( r_i \)’s toward 1 when there is not much data. Note that the prior for \( r \) carries important influence when the data are scarce.

We now present a procedure for selecting \( P \). Assume that \( \theta \) and \( P \) are a priori independent, and that \( P \) is uniformly dis-
distributed on a finite set of values $\mathcal{Z} = \{P_{\text{min}}, P_{\text{min}} + 1, \ldots, P_{\text{max}}\}$. Then, the posterior probability of $(P, \theta)$ is given by

$$p(P, \theta | Y) \propto p(Y | \theta, P)p(\theta)|_{\mathcal{Z}(P)},$$

(14)

where $I_{\mathcal{Z}}(P) = 1$ if $P \in \mathcal{Z}$ and $I_{\mathcal{Z}}(P) = 0$ otherwise. As discussed in Section 2, the MAPE $\hat{\theta}(P)$ of $\theta$ given $P$ can be computed from (10) and (11). Thus, the MAPE of $\theta$ is obtained by the maximization of $\Lambda(P, \theta(P))$ over all $P \in \mathcal{Z}$, where $\Lambda(P, \theta) = -2 \ln[p(Y | \theta, P)p(\theta)]$. It is often difficult to specify $P_{\text{min}}$ and $P_{\text{max}}$ that reflects real prior knowledge, and it is time consuming to compute $\Lambda(P, \theta(P))$ for large values of $P$. We propose setting $P_{\text{min}}$ to be the smallest $P$ such that $q \geq 2n$ and $X$ has full row rank, and $P_{\text{max}}$ to be the largest value of $P$ that can be used without causing computational problems. In this article, we take $P_{\text{max}}$ to be the largest value of $P$ that gives $q \leq 10^4$. To reduce the need to compute many $\Lambda(P, \theta(P))$, we propose selecting $P$ to be $\hat{P}$, the smallest local minimum of $\Lambda(P, \theta(P))$, instead of the MAPE. To find $\hat{P}$, we compute $\Lambda(P_{\text{min}}, \theta(P_{\text{min}})), \Lambda(P_{\text{min}} + 1, \theta(P_{\text{min}} + 1))$, and so on until an increase in value of $P$ by 1 increases $\Lambda(P, \theta(P))$. This procedure frequently terminates quickly because $\hat{P}$ is often near $P_{\text{min}}$. Note that if this procedure is used, setting $P_{\text{min}}$ to be as small as possible, that is, setting $P_{\text{min}}$ to the smallest $P$ such that $q > n$, often do not give good results. If $P_{\text{min}}$ is set to the smallest $P$ such that $q > n$, and if there is no basis function of degree $P_{\text{min}} + 1$ with a strong effect, the estimation method would often choose $P = P_{\text{min}}$. However, if $P = P_{\text{min}}$, the degrees of freedom $q - n$ will often be small, which can easily lead to underestimation of the uncertainty about the true function. Finally, we point out that comparing different estimation methods or priors for $P$ is an area needing further research.

4. INITIAL DESIGNS

4.1 Designs for Uni-Dimensional Input Spaces

The Chebyshev points of the second kind in the interval $[-1, 1]$ are given by

$$u_i = \cos\left(\frac{i}{n - 1}\pi\right), i = 0, \ldots, n - 1.$$  

(15)

These points are recommended for Lagrange interpolation (Phillips 2003; Trefethen 2013) and can be used as a design for SIP. This design tends to give SIPs with small maximum prediction variance compared with SIPs obtained from a design with $n$ equally spaced points. To illustrate, Figure 1 plots the SIP point predictions and 95% credible intervals for the famous Runge function $y(u) = 1/(1 + 25u^2)$ obtained with 11 Chebyshev points $D_1$ (computed from (15)) (1(a) and 1(b)), the design $D_2 = [-1, -0.8, \ldots, 1]$ (1(c) and 1(d)), and the design $D_3 = [-\frac{10}{11}, -\frac{8}{11}, \ldots, \frac{10}{11}]$ (1(e) and 1(f)). Figure 1(a), 1(c), and 1(e) shows the predictions of SIPs constructed with orthonormal Legendre polynomials (Legendre-SIP), that is, the $\psi_n$’s in (1) are normalized Legendre polynomials. Figure 1(b), 1(d), and 1(f) shows the predictions of SIPs constructed with orthonormal Chebyshev polynomials (Chebyshev-SIP), that is, the $\psi_n$’s in (1) are normalized Chebyshev polynomials. Note that $w(u) = L_{-1,1}(u)/2$ and $w(u) = L_{-1,1}(u)/(\pi\sqrt{1 - u^2})$, where $L_{-1,1}(u) = 1$ if $u \in [-1, 1]$ and $L_{-1,1}(u) = 0$ otherwise, for the Legendre and Chebyshev polynomials, respectively. The estimate $\hat{P}$ of $P$ is 22 in all cases and the estimate $\hat{f}_i$ of $f_i$ ranges from 0.768 to 0.812. By comparing Figure 1(a), 1(c), and 1(e) and also Figure 1(b), 1(d), and 1(f), we see that the SIP point predictions (dashed line) are insensitive to the design. However, the predictions are less accurate and the credible interval width increases drastically near the endpoints ±1 when design $D_2$ or $D_3$ is employed instead of $D_1$. The Lagrange interpolating Lagrange polynomials, that is, polynomial of degree $P = n - 1$ that interpolates the data, for designs $D_1$, $D_2$, and $D_3$ are also plotted in Figure 1. We see that the Lagrange polynomial is very sensitive to the choice of design. It performs well when $D_1$ is employed but it gives wild predictions near the endpoints when $D_2$ or $D_3$ is used. This is the well-known Runge phenomenon. Thus, SIP is a good alternative to Lagrange interpolation when the choice of interpolation points is not fully under the control of the model builder. In Section S1 of the online supplement, we provide a heuristic explanation about why the Chebyshev points are better than equally spaced points for constructing SIPs.

4.2 Designs for Multi-Dimensional Input Spaces

In this section, we discuss the choice of an initial/exploratory design for multi-dimensional input spaces of the form $[-1, 1]^d$. We have demonstrated that Chebyshev points are good for one-dimensional interpolation. In a multi-dimensional input space, designs that are Cartesian products of Chebyshev points may be used. However, the number of design points can quickly become too large as $d$ increases. Space-filling designs are commonly used to build GP emulators in moderate to high-dimensional input spaces with small number of runs. These designs are model independent (not constructed to optimize criteria specific to a statistical model) and can achieve good spatial representativeness of the design region with relatively small number of runs compared to grid designs. Examples of popular space-filling designs include maximin Latin hypercube designs (maximin LHDS) (Morris and Mitchell 1995) and Sobol sequences (Lemieux 2009). However, when $d = 1$, it has been demonstrated in Section 4.1 that equally spaced designs, which are space-filling designs, are poor choices for constructing SIPs. Nonetheless, it is not clear whether space-filling designs are poor choices for $d \geq 2$. In this article, we shall propose a class of transformed space-filling design that combines some desirable features of Chebyshev points and space-filling designs. These designs can have any specified number $n \geq 2$ of runs in $d$ dimensions and they give one-dimensional projections that have distributions that converge to the asymptotic distribution of the Chebyshev points.

We propose the following general cosine transformed space-filling design. Let $D_U$ be an $n$-point space-filling design such that each one-dimensional projection of the design converges to a uniform distribution on $[0, 1]$ as $n \to \infty$, that is, the proportion of points in any subinterval of $[0, 1]$ is proportional to its length. Then, the cosine transformed space-filling design is given by

$$D_C = \cos(D_U \pi),$$  

(16)

where the multiplication by $\pi$ and cosine function are applied elementwise to the $D_U$ matrix. This design can be justified as follows. Let $U \sim \text{unif}([-1, 1])$ and $T = \cos(U \pi)$. Then, for
Figure 1. True Function, Data, Legendre-SIP Predictions, and Lagrange Polynomial obtained with Chebyshev points (a), design $D_2 = \{-1, -0.8, \ldots, 1\}$ (c), and design $D_3 = \{-10/11, -8/11, \ldots, 10/11\}$ (e); True Function, Data, Chebyshev-SIP Predictions and Lagrange Polynomial obtained with Chebyshev points (b), design $D_2 = \{-1, -0.8, \ldots, 1\}$ (d), and design $D_3 = \{-10/11, -8/11, \ldots, 10/11\}$ (f).

t \in [-1, 1],

$$P[\cos(U\pi) \in [-1, t]] = P[U \pi \in [\cos^{-1} t, \pi]]$$

$$= P[U \in [\pi^{-1} \cos^{-1} t, 1]]$$

$$= 1 - \pi^{-1} \cos^{-1} t.$$
The density given in (17) is the asymptotic density of the Chebyshev points, that is, when \( n \to \infty \), the proportion of Chebyshev points in each subinterval of \([-1, 1]\) converges to the value given by the integral of \( f(t) \) over the subinterval. It is known that if \( F: [-1, 1] \to \mathbb{R} \) is a function that can be analytically continued to a function that is analytic in a neighborhood of \([-1, 1]\), then any system of points with the asymptotic density given in (17) will give a Lagrange interpolator \( \xi \) that converges at a geometric rate or faster to \( F \) when \( n \to \infty \), that is, \( \| \xi - F \|_\infty = O(K^{-n}) \) for some \( K > 1 \) (Trefethen 2013).

In view of this and the discussion in Section S1 of the online supplement, we expect the proposed cosine transformed space-filling design to work well when the computer output can be analytically continued and the response variation is mainly due to a single input. In Section 5, we show that cosine transformed space-filling designs obtained by taking \( D_U \) as a maximin LHD, which we call cosine maximin LHDs, outperform maximin LHDs for the purpose of constructing SIPs. Note that cosine maximin LHDs have been proposed by Dette and Pepelyshev (2010) for fitting GP models.

Sequential designs can be very useful in applications where a given degree of interpolation accuracy needs to be achieved and the computing cost is to be kept as small as possible. In Section S2 of the online supplement, we propose a sequential design approach to achieve good prediction over the entire design region \( \chi \).

5. SIMULATION STUDIES

This section reports the results of a simulation study. The objectives of the simulation are to compare the prediction accuracy of SIPs and GP emulators, to compare the performance of cosine maximin LHD and maximin LHD for constructing SIPs and to assess the accuracy of SIP in estimating sensitivity indices.

5.1 Simulation Study 1

We perform a simulation with two main objectives. The first is to compare the accuracy of the SIP and the stationary GP emulator with Gaussian and Matérn correlation functions. The second is to compare the effectiveness of the maximin LHD and cosine maximin LHD.

Six test functions, that is, Rosenbrock, Tilden, Weld Shear Stress, Two-Bar Truss Safety Constraint, Product Peak, and Borehole Flow Rate, with input dimensions ranging from three to eight are employed. The Rosenbrock test functions are a popular class of test functions in the optimization literature. These functions are polynomial functions of degree 4. In this study, we work with a Rosenbrock function in three dimensions. The Tilden function, which has four input variables, describes the behavior of a chemical reaction system over time (Saltelli 2000). Both Weld Shear Stress and Two-Bar Truss Safety Constraint functions are given in Muller and Messac (2006). The former gives the logarithm of the shear stress in the welding joint of a structure while the latter is the amount by which the compressive stress exceeds the critical buckling stress in a Two-Bar Truss expressed as a fraction of the critical buckling stress. The Product Peak function is used as a test function in Barthelmann, Novak, and Ritter (2000) to numerically demonstrate the convergence of sparse grid interpolation. Finally, the Borehole Flow Rate function is introduced in Morris, Mitchell, and Ylvisaker (1993). Note that most of the test functions are real albeit simple physical models (Tilden, Weld Shear Stress, Two-Bar Truss Safety Constraint, and Borehole Flow Rate) while some of them are very popular test functions (Rosenbrock and Product Peak). All test functions are defined in Appendix A and their inputs are linearly transformed so that the design region is \( \chi = [-1, 1]^d \).

We run simulations with a design size of \( n = 7d \) for all the test functions. For the Rosenbrock and Tilden functions, we also run simulations with design size \( n = 10d \). Thus, there are a total of eight test cases. Two designs are tested: the maximin LHD and the cosine maximin LHD. In each simulation run, we first generate a good maximin LHD \( D_U \) on \([0, 1]^d\). This is done by using the Matlab function lhsdesign to generate 500 maximin LHDs and choosing the best (with respect to the maximin criterion) of the generated maximin LHDs. Note that the lhsdesign function generates maximin LHDs by randomly simulating a number of Latin hypercube designs and selecting the design that performs best. The maximin LHD \( D_U \) is then transformed into a cosine maximin LHD \( \cos[(1 - D_U) \pi] \) and a maximin LHD \( 2D_U - 1 \) on \( \chi = [-1, 1]^d \). The reason for constructing the two designs in this fashion is to induce positive correlation (the former design looks like the latter design with points pushed toward the boundary of \( \chi \)) in the simulation results, which will allow more precise comparison. These two designs, that is, the cosine maximin LHD and the maximin LHD on \( \chi \), are used to construct SIPs and GP emulators. We have found that they always give a model matrix \( X \) with full row rank if \( P \) is chosen such that \( q \geq 2n \). In each simulation run, two different SIP models, that is, Legendre-SIP and Chebyshev-SIP, are built using each of the two designs. The Legendre-SIPs are obtained by taking the \( \psi_j \)'s to be tensor products of orthonormal Legendre polynomials (similarly for Chebyshev-SIP). Stationary GP emulators with product Gaussian correlation function and product Matérn correlation function with smoothness parameter \( v = 2.5 \) (Santner, Williams, and Notz 2003) are also fitted using both designs. Thus, a total of eight metamodels are constructed in each simulation run, as indicated by the model-design combinations presented in Table 1. For the GP emulators, the mean, variance, and correlation parameters are estimated using the maximum likelihood method. Details of the optimization procedures used to obtain the MAPE of \( r \) for the SIPs and the MLE of the correlation parameters for the GP emulators are given in Appendix B.
For each test case, we run 100 simulations. As mentioned above, results for the eight model-design combinations given in Table 1 are expected to be positively correlated within a simulation because any two combinations use either the same or positively correlated designs. In other words, each simulation run is a random block. This allows more precise comparison of the performance of the eight model-design combinations. The simulation mean of the estimate $\hat{p}$ of $p$ is between 3 and 6.2 for the four SIP model-design combinations in all test cases. One of the performance measures computed in the simulations is the percent root mean squared prediction error (PRMSE). The PRMSE is the square root of the average squared prediction error $\text{PRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$, where $\hat{y}_i$ is the fitted value and $y_i$ is the true output value. The credible intervals for the SIPs and GP models are obtained by fitting the cosine maximin LHD over GP emulators can be quite large. The simulation means and standard deviations of the percent coverage of 95% credible intervals given by the SIPs and GP emulators are shown in Table 3. The percent coverage is the percentage of credible intervals in the test set that contains the true output value. The credible intervals for the SIPs are obtained from a normal distribution with mean and variance given by (6) and (7) while the credible intervals for the GP models are obtained from a $t$-distribution with $n - 1$ degrees of freedom, as described in Santner, Williams, and Notz (2003). Table 3 shows that the SLP and the GP model with Matérn correlation function have much closer-to-nominal coverage than the GP model with Gaussian correlation function (compare A, C, E, G for maximin LHD and B, D, F, H for cosine maximin LHD). The table also shows that in many cases, the SIP has closer-to-nominal coverage than the GP model with Matérn correlation function.

It is of interest to determine whether cosine maximin LHDs can give smaller maximum prediction variance than maximin LHDs, because it is demonstrated in Section 4.1 that Chebyshev points give smaller maximum prediction variance than evenly spaced points when $d = 1$. It is also of interest to compare the average prediction variance over the design region achieved with the two designs. Table 4 presents the simulation means of the maximum and average standard deviation in the test set.
Table 3. Simulation mean and standard deviation of percent coverage of credible intervals

| Function         | d  | n  | A   | B   | C   | D   | E   | F   | G   | H   |
|------------------|----|----|-----|-----|-----|-----|-----|-----|-----|-----|
| Rosenbrock       | 3  | 21 | Mean| 93.8| 93.3| 91.4| 95.1| 80.6| 86.6| 89.0| 94.5|
|                  |    |    | StD | 7.6 | 10.3| 8.8 | 8.4 | 12.4| 10.8| 9.3 | 7.3 |
| Rosenbrock       | 3  | 30 | Mean| 99.7| 99.8| 99.4| 99.9| 79.4| 76.2| 96.3| 98.2|
|                  |    |    | StD | 1.3 | 0.9 | 2.0 | 0.7 | 12.7| 11.6| 3.3 | 1.9 |
| Tilden           | 4  | 28 | Mean| 87.1| 87.0| 85.7| 87.2| 66.9| 69.9| 79.5| 80.9|
|                  |    |    | StD | 7.8 | 8.8 | 8.0 | 8.8 | 10.9| 11.9| 9.3 | 10.2|
| Tilden           | 4  | 40 | Mean| 87.5| 88.7| 86.2| 89.1| 71.1| 75.0| 83.0| 86.5|
|                  |    |    | StD | 7.3 | 8.8 | 7.7 | 8.8 | 9.8 | 11.4| 7.9 | 8.2 |
| Weld Shear Stress| 5  | 35 | Mean| 95.0| 94.0| 94.8| 94.3| 73.3| 68.1| 89.5| 86.5|
|                  |    |    | StD | 4.6 | 6.2 | 4.6 | 5.9 | 12.4| 13.4| 8.4 | 9.9 |
| Two-Bar Truss    | 6  | 42 | Mean| 81.8| 87.6| 78.6| 86.9| 70.0| 79.1| 73.8| 83.2|
| Safety Constraint|    |    | StD | 6.9 | 7.5 | 8.3 | 7.5 | 9.8 | 8.7 | 6.2 | 6.2 |
| Product Peak     | 7  | 49 | Mean| 92.4| 90.4| 92.4| 91.1| 85.7| 85.6| 90.2| 90.1|
|                  |    |    | StD | 4.3 | 5.0 | 5.3 | 4.8 | 6.0 | 6.0 | 5.1 | 5.1 |
| Borehole Flow    | 8  | 56 | Mean| 91.0| 91.9| 91.2| 92.1| 74.8| 75.9| 88.4| 90.9|
| Rate             |    |    | StD | 5.3 | 5.0 | 5.2 | 5.1 | 8.5 | 7.7 | 5.2 | 5.0 |

expressed as a fraction of the range of the true function values in the test set for cases A–D. Expressing the maximum and average standard deviation as a fraction of the range allows us to assess whether the standard deviation given by (7) is of reasonable magnitude. We see that the maximum and average standard deviations are small (around 0.1 or less) compared to the range of the function for most cases. We have observed that the maximum standard deviation over \( \chi \) often occurs at the corner points. Because the test set always include these points, the maximum standard deviations over the test set and \( \chi \) are likely to be nearly the same in most simulation runs. Table 4 also gives the \( t \)-statistics for testing the difference between cosine maximin LHD and maximin LHD under the columns headed A-B and C-D. For five of the test cases (Rosenbrock with \( n = 30 \), Tilden with \( n = 28, 40 \), Weld Shear Stress, and Product Peak), the maximum standard deviation is significantly smaller on the average when the Legendre-SIP or Chebyshev-SIP is constructed with a cosine maximin LHD instead of a maximin LHD. For two test cases, the opposite occurs. However, cosine maximin LHDs give a larger average standard deviation than maximin LHDs in many cases.

In summary, SIPs constructed with cosine maximin LHDs has better accuracy than the stationary GP models with Gaussian and Matérn correlation functions. In addition, SIPs give prediction credible intervals with better coverage than the GP model. The cosine maximin LHDs are better for constructing SIPs than maximin LHDs because they generally give more accurate SIPs and in some cases, they yield SIPs with smaller maximum prediction variance.

5.2 Estimation of Total Sensitivity Indices

Because estimation of variance decompositions and sensitivity indices is a key application of polynomial chaos mod-

Table 4. Simulation means of maximum (Max) and average (Ave) prediction standard deviation over range of function in test set

| Function         | d  | n  | A   | B   | C   | D   | A-B | C-D |
|------------------|----|----|-----|-----|-----|-----|-----|-----|
| Rosenbrock       | 3  | 21 | Max | 0.3626| 0.3330| 0.2164| 0.2328| 4.59 | −6.57|
|                  |    |    | Ave | 0.0537| 0.0566| 0.0453| 0.0586| −3.27| −13.95|
| Rosenbrock       | 3  | 30 | Max | 0.1923| 0.1413| 0.1551| 0.1190| 27.09| 26.40|
|                  |    |    | Ave | 0.0207| 0.0211| 0.0187| 0.0206| −2.57| −12.83|
| Tilden           | 4  | 28 | Max | 0.1239| 0.1128| 0.1106| 0.1027| 5.19 | 4.45 |
|                  |    |    | Ave | 0.0162| 0.0198| 0.0155| 0.0203| −10.57| −12.81|
| Tilden           | 4  | 40 | Max | 0.1046| 0.0898| 0.0913| 0.0820| 9.13 | 7.20 |
|                  |    |    | Ave | 0.0110| 0.0132| 0.0107| 0.0140| −12.66| −18.45|
| Weld Shear Stress| 5  | 35 | Max | 0.0259| 0.0227| 0.0233| 0.0204| 10.20| 10.82|
|                  |    |    | Ave | 0.0038| 0.0043| 0.0036| 0.0040| −8.94| −10.61|
| Two-Bar Truss    | 6  | 42 | Max | 0.0523| 0.0689| 0.0441| 0.0620| −11.51| −11.84|
| Safety Constraint|    |    | Ave | 0.0075| 0.0121| 0.0072| 0.0134| −18.49| −16.70|
| Product Peak     | 7  | 49 | Max | 0.1101| 0.0913| 0.1081| 0.0893| 12.62| 13.94|
|                  |    |    | Ave | 0.0249| 0.0223| 0.0257| 0.0230| 7.23 | 7.70 |
| Borehole Flow    | 8  | 56 | Max | 0.0108| 0.0112| 0.0110| 0.0115| −2.49| −2.87|
| Rate             |    |    | Ave | 0.0022| 0.0025| 0.0023| 0.0026| −7.91| −9.44|
els, we briefly provide some results concerning estimation of
total sensitivity indices (see Sudret 2008 for a definition)
with SIP. If \( Y \) denotes the true output function, \( y = y(0) + \sum_{i=1}^{d} y(i) + \sum_{i=d+1}^{d+q} y(i,j) + \cdots + y(1,\ldots,d) \)
denotes the functional ANOVA decomposition of \( y \) (and similarly for
the model \( Y \) in (1)), then it can be shown that

\[
\|Y - y\|_w^2 = (Y_0 - y_0)^2 + \sum_{k=1}^{d} \sum_{1 \leq i_1 < \cdots < i_k \leq d} \|Y_{i_1,\ldots,i_k}\|_w^2,
\]

where \( \|Y - y\|_w^2 = \int f(Y - y)^2 w(Y)dy. \) Thus, if \( Y \)
is a good approximation of \( y \), then all the functional ANOVA
components of \( Y \) given by (2) are good approximations of the
corresponding functional ANOVA component of \( y \). The good
accuracy of the SIP model \( Y \) for approximating \( y \) has already
been demonstrated in Section 5.1.

For model (1), the total sensitivity index of input \( i \) is

\[
T_i = \frac{\beta_i^2}{\sum_{\alpha \in N_i(d,P), \alpha \geq 0} \beta_\alpha^2}.
\]

Monte Carlo samples from the posterior distribution of \( \beta \) (8)
can be used to compute the posterior mean and quantiles of \( T_i \). We
use the data from the simulation reported in Section 5.1
for generating cosine maximin LHDs and for constructing
SIPs. It is demonstrated that SIPs constructed with these designs
generally have better prediction accuracy than SIPs constructed
with maximin LHDs. Simulation results also demonstrate that
SIPs constructed with cosine maximin LHDs tend to provide
more accurate predictions and credible interval coverage than GP
tools. A key computational advantage of SIP over GP
models is the computation of the functional ANOVA and vari-
ance decompositions for global sensitivity analysis. For this
application, we have demonstrated that the SIP gives reliable
estimates of sensitivity indices.

A disadvantage of the SIP is that it cannot model some func-
tions in high dimensions well. The computer memory and speed
limits the maximum number of basis functions \( q \) that can be
employed. Thus, the maximum \( P \) that can be used decreases
with \( d \). Consequently, large \( P \)'s cannot be used when \( d \) is large.
However, few, if any, metamodels can successfully approximate
complex high-dimension functions with small designs. Many
methods can be successful when a high-dimension function
has a handful of significant inputs. In these cases, the model-
ing problem essentially reduces to a low-dimensional modeling
problem. Further research is needed to determine how best to
model this kind of functions with SIPs.

### SUPPLEMENTARY MATERIALS

**Additional Material and Appendices.pdf:** This file contains
Supplementary Section 1: Heuristic Explanation of Good Performance of SIP Fitted with Chebyshev Points, Supplementary
Section 2: Design of Follow-Up Runs, and Appendices A, and B.

**Programs and Data.zip:** The zip file contains Matlab codes
for generating cosine maximin LHDs and for constructing
Legendre-SIPs and Chebyshev-SIPs. Matlab codes for generating
data from the test functions employed in this articles, and
for fitting the GP model with Gaussian correlation function are also
given.
ACKNOWLEDGMENTS

We thank the associate editor and two referees for comments that helped improve the article significantly. This research is supported by City University of Hong Kong Start-Up Grant 7200364 and Early Career Scheme (ECS) project No. 21201414 sponsored by the Research Grants Council of Hong Kong.

[Received November 2013. Revised July 2014.]

REFERENCES

Agarwal, N., and Aluru, N. R. (2009), “A Domain Adaptive Stochastic Collocation Approach for Analysis of MEMS Under Uncertainties,” Journal of Computational Physics, 228, 7662–7688. [458]
Apley, D. W., Liu, J., and Chen, W. (2006), “Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments,” Journal of Mechanical Design, 128, 945–958. [458]
Barthelmann, V., Novak, E., and Ritter, K. (2000), “High Dimensional Polynomial Interpolation on Sparse Grids,” Advances in Computational Mathematics, 12, 273–288. [458,463]
Bates, R. A., Giglio, B., and Wynn, H. P. (2003), “A Global Selection Procedure for Polynomial Interpolators,” Technometrics, 45, 246–255. [458]
Bates, R. A., Maruri-Aguilar, H., and Wynn, H. P. (2014), “Smooth Supersaturated Models,” Journal of Statistical Computation and Simulation, 84, 1–12. [458]
Box, G. E. P., and Draper, N. R. (2007), Response Surfaces, Mixtures, and Ridge Analyses (2nd ed.), Hoboken, NJ: Wiley. [457]
Chen, W., Jin, R., and Sudjianto, A. (2006). “Analytical Global Sensitivity Analysis and Uncertainty Propagation For Robust Design,” Journal of Quality Technology, 38, 333–348. [457]
Chung, K. C., and Yao, T. H. (1977), “On Lattices Admitting Unique Lagrange Interpolations,” SIAM Journal on Numerical Analysis, 14, 735–743. [458]
Crestaux, T., Le Maître, O. P., and Martinez, J. M. (2009), “Polynomial Chaos Expansion for Sensitivity Analysis,” Reliability Engineering and System Safety, 94, 1161–1172. [457]
Curran, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), “Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments,” Journal of the American Statistical Association, 86, 953–963. [458,460]
De Boor, C., and Ron, A. (1990), “On Multivariate Polynomial Interpolation,” Constructive Approximation, 6, 287–302. [458]
Dette, H., and Pepelyshev, A. (2010), “Generalized Latin Hypercube Design For Computer Experiments,” Technometrics, 52, 421–429. [458]
Gasca, M., and Sauer, T. (2000), “Polynomial Interpolation in Several Variables,” Advances in Computational Mathematics, 12, 377–410. [458]
Gautschi, W. (2004), Orthogonal Polynomials: Computation and Approximation, New York: Oxford University Press. [458]
Ghanem, R., and Spanos, P. D. (1991), Stochastic Finite Elements: A Spectral Approach, New York: Springer-Verlag. [457,460]
Holliday, T., Pistone, G., Riccomagno, E., and Wynn, H. P. (1999), “The Application of Computational Algebraic Geometry to the Analysis of Designed Experiments: A Case Study,” Computational Statistics, 14, 213–232. [458]
Joseph, V. R. (2006), “A Bayesian Approach to the Design and Analysis of Fractionated Experiments,” Technometrics, 48, 219–229. [459,460]
Lazarov, B. S., Schevenels, M., and Sigmund, O. (2012), “Topology Optimization Considering Material and Geometric Uncertainties Using Stochastic Collocation Methods,” Structural and Multidisciplinary Optimization, 46, 597–612. [458]
Le Maître, O. P., and Knio, O. M. (2010), Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics, New York: Springer-Verlag. [457,460]
Lemieux, C. (2009), Monte Carlo and Quasi-Monte-Carlo Sampling, New York: Springer-Verlag. [461]
Morris, M. D., and Mitchell, T. J. (1995), “Exploratory Designs for Computational Experiments,” Journal of Statistical Planning and Inference, 43, 381–402. [461]
Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993), “Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction,” Technometrics, 35, 243–255. [463]
Muirhead, R. I. (2005), Aspects of Multivariate Statistical Theory (2nd ed.), New York: Wiley. [459]
Muller, A. A., and Messac, A. (2006), “Metamodelling Using Extended Radial Basis Functions: A Comparative Approach,” Engineering with Computers, 21, 203–217. [463]
Narayan, A., and Xiu, D. (2012), “Stochastic Collocation Methods on Unstructured Grids in High Dimensions via Interpolation,” SIAM Journal on Scientific Computing, 34, 1729–1752. [458]
——— (2013), “Constructing Nested Nodal Sets for Multivariate Polynomial Interpolation,” SIAM Journal on Scientific Computing, 35, 2293–2315. [458]
Oakley, J. E., and O’Hagan, A. (2004), “Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach,” Journal of the Royal Statistical Society, Series B, 66, 751–769. [459]
Phillips, G. M. (2003), Interpolation and Approximation by Polynomials, New York: Springer-Verlag. [458,461]
Pistone, G., and Wynn, H. P. (1996), “Generalised Confounding With Gröbner Bases,” Biometrika, 83, 653–666. [458]
Rasmussen, C. E., and Williams, C. K. I. (2006), Gaussian Processes for Machine Learning, Cambridge: MIT Press. [460]
Robert, C. F. (2007), The Bayesian Choice (2nd ed.), New York: Springer. [460]
Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), “Design and Analysis of Computer Experiments,” Statistical Science, 4, 409–423. [458]
Saltelli, A. (2000), “What is Sensitivity Analysis?,” in Sensitivity Analysis, eds. A. Saltelli, K. Chan, and E. M. Scott, New York: Wiley. [463]
Santrner, T. J., Williams, B. J., and Notz, W. I. (2003), The Design and Analysis of Computer Experiments, New York: Springer-Verlag. [463,464]
Sauer, T., and Xu, Y. (1995), “On Multivariate Lagrange Interpolation,” Mathematics of Computation, 64, 1147–1170. [458]
Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F. (1998), “Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization,” AIAA Paper 98, 4758, 1–11. [457]
Steinberg, D. M., and Burszyn, D. (2004), “Data Analytic Tools for Understanding Random Field Regression Models,” Technometrics, 46, 411–420. [460]
Sudret, B. (2008), “Global Sensitivity Analysis Using Polynomial Chaos Expansions,” Reliability Engineering and System Safety, 93, 964–979. [457,459,466]
Tan, M. H. Y., and Wu, C. F. J. (2012), “Robust Design Optimization With Quadratic Loss Derived From Gaussian Process Models,” Technometrics, 54, 51–63. [458]
Trefethen, L. N. (2013), Approximation Theory and Approximation Practice, Philadelphia: SIAM. [458,461,463]
Wang, G. G., and Shan, S. (2007), “Review of Metamodeling Techniques in Support of Engineering Design Optimization,” Journal of Mechanical Design, 129, 370–380. [457]
Welch, W., Yu, T., Kang, S. M., and Sacks, J. (1990), “Computer Experiments for Quality Control by Parameter Design,” Journal of Quality Technology, 22, 15–22. [457]
Xiu, D. (2010), Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton, NJ: Princeton University Press. [457,460]
Xiu, D., and Hesthaven, J. S. (2005), “High-Order Collocation Methods for Differential Equations With Random Inputs,” SIAM Journal on Scientific Computing, 27, 1118–1139. [458]
Xiu, D., and Karniadakis, G. E. (2002), “The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations,” SIAM Journal on Scientific Computing, 24, 619–644. [457]