

Pointwise Bias Error Bounds for Response Surface Approximations and Min-Max Bias Design

MELIH PAPILA

Department of Mechanical and Aerospace Engineering, University of Florida,
Gainesville, FL 32611-6250
papila@ufl.edu

RAPHAEL T. HAFTKA

Department of Mechanical and Aerospace Engineering, University of Florida,
Gainesville, FL 32611-6250
haftka@ufl.edu

LAYNE T. WATSON

Departments of Computer Science and Mathematics, Virginia Polytechnic Institute &
State University, Blacksburg, VA 24061-0106
ltw@cayuga.cs.vt.edu

Abstract. Two approaches addressing response surface approximation errors due to model inadequacy (bias error) are presented, and a design of experiments minimizing the maximal bias error is proposed. Both approaches assume that the functional form of the true model is known and seek, at each point in design space, worst case bounds on the absolute error. The first approach is implemented prior to data generation. This data independent error bound can identify locations in the design space where the accuracy of the approximation fitted on a given design of experiments may be poor. The data independent error bound can easily be implemented in a search for a design of experiments that minimize the bias error bound as it requires very little computation. The second approach is to be used posterior to the data generation and provides tightened error bound consistent with the data. This data dependent error bound requires the solution of two linear programming problems at each point. The paper demonstrates the data independent error bound for design of experiments of two-variable examples. Randomly generated polynomials in two variables are then used to validate the data dependent bias-error bound distribution.

NOMENCLATURE

A	Alias matrix
b	Vector of estimated coefficients of basis functions
b_j	Estimated coefficients of basis functions
$c^{(2)}$	Vector of bounds for coefficients of the basis functions
E_b	Vector of true prediction errors due to bias at the data points
$e_b(x)$	True bias error at design point x
$ e_b^D(x) $	Data dependent worst case bias error bound at design point x
$ e_b^I(x) $	Data independent worst case bias error bound at design point x
$ e_b^I _{av}$	Average of data independent bias error bounds over the design domain
$ e_b^I _{max}$	Data independent maximum absolute bias error bound over the design domain
$e_{es}(x)$	Estimated standard error at design point x
$[\bar{e}_{es}]_{max}$	Maximum of normalized estimated standard error over the design domain
$F(x), F^{(1)}(x), F^{(2)}(x)$	Vectors of basis functions at x
$f_j(x)$	Basis functions
N	Number of data points
n_1	Number of basis functions in the regression model
n_2	Number of missing basis functions in the regression model
s^2	Error mean square
x	Design point
$X, X^{(1)}, X^{(2)}$	Gramian (design) matrices
x_1, x_2, \dots, x_n	Design variables
$x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}$	Design variables for i^{th} design point
y	Vector of observed responses
$\hat{y}(x)$	Response surface approximation at a design point x
$\alpha_1, \alpha_2, \dots, \alpha_m$	Design of experiment parameter
$\beta, \beta^{(1)}, \beta^{(2)}$	Vectors of basis function coefficients
$\beta_j, \beta_j^{(1)}, \beta_j^{(2)}$	Coefficients of basis functions
$\eta(x)$	True mean response at x
σ^2	Noise variance

BACKGROUND

Response surface approximation (RSA) techniques have become popular in engineering optimization based on computer simulations. Kaufman et al. (1996), for instance, fitted quadratic polynomial RSAs to the structural weight of the high speed civil transport (HSCT), obtained with multiple structural optimizations via the commercial code GENESIS. Balabanov et al. (1996, 1999) investigated RSA construction for the wing bending material weight of the HSCT based on structural optimization results of a number of different configurations, and use of the RSA in the configuration optimization problem. Papila and Haftka (1999, 2000) also constructed RSAs for weight equations based on structural optimizations. Shyy et al. (2001) employed RSA methodology for rocket engine injector design optimization. Papila et al. (2001) constructed RSAs based on CFD simulations for supersonic turbine blade design optimization. Redhe et al. (2000) determined an efficient number of data points when using the RSA methodology in crashworthiness problems, where numerical simulations were carried out using LS-DYNA.

This popularity has brought attention to ways of increasing the accuracy of RSA and protecting designers from misleading predictions. The accuracy of RSA is mainly affected by the following three factors: (i) limitations on the number of data points due to the cost of data generation, (ii) noise in the data, and (iii) inadequacy of the fitting model (bias errors).

The prediction error of an RSA varies from point to point. There is an established statistical tool for characterizing noise-related prediction errors: prediction variance (Myers and Montgomery 1995, and Khuri and Cornell 1996). However, in many computer simulations, numerical noise is not a major problem, and RSA accuracy suffers mostly due to bias errors. There has been a fair amount of work on design of experiments for minimizing the mean squared error averaged over the design domain that combines variance and bias errors (Box and Draper, 1963, Draper and Lawrence, 1965, Kupper and Meydrech, 1973, and Welch, 1983). The bias component of the averaged or integrated mean squared error was also minimized to obtain so-called minimum bias designs. The fundamentals of minimizing integrated mean squared error and its components can be found in Myers and Montgomery (1995) and Khuri and Cornell (1996). Venter and Haftka (1997) developed an algorithm implementing a minimum-bias criterion, necessary for an irregularly shaped domain where no closed form solution exists for minimum-bias experimental design. They compared minimum-bias and D-optimal experimental designs for two problems with two and three variables. The minimum-bias experimental design was found to be more accurate than D-optimal for the two problems for the average error, but not for the maximum error. Montepiedra and Fedorov (1997) investigated experimental designs minimizing the bias component of the integrated

mean square error subject to a constraint on the variance component or vice versa. Federov et al. (1999) studied later design of experiments via weighted regression prioritizing regions where the approximation is desired to predict the response. Their approach considered both the variance and bias components of the estimation error. Palmer and Tsui (2001) studied minimum-bias Latin hypercube experimental design for sampling from deterministic process simulators. Recently Qu et al. (2004) have implemented Gaussian quadrature based minimum bias design. They have also presented minimum bias central composite designs for up to six variables.

OBJECTIVES

Bias error, averaged over the domain of interest, has been studied extensively in the approximation theory and numerical analysis literature. However, very little has been done on the point to point variation of bias errors. This variation is of interest for two reasons. First, when using an approximation, it can be important to identify regions where the bias error may be large. Second, when selecting a design of experiments, it may be desirable to minimize the maximal bias error (min-max bias error) rather than the average. An approach for estimating a bound on RSA bias errors was previously presented by Papila and Haftka (2001). The traditional decomposition of the mean squared error into variance and the square of the bias was followed, but point to point rather than averaged over the domain. The bound on the squared bias error does not depend on the response data. Such an error bound may be used to identify regions in design space where the accuracy of the RSA may be poor for a given design of experiments. Moreover, the error bound may provide a tool for designing experiments against the bias error prior to data generation. The first objective of this paper is to demonstrate how the error bound can be used to obtain designs of experiments that minimize the maximal bias error.

Once data is available, the error bound can be tightened, and the second objective of the paper is to develop a point to point error bound for a given set of data. This error bound can identify regions of possible large errors and may be used to adaptively refine a design of experiments, or to steer clear of such regions.

The next two sections, respectively, present the bias error expression at a given design point and derive the bias error bounds for a given design of experiment prior to data generation (data independent bounds). This is followed by demonstration of a min-max bias design of experiments. Next, the data dependent error bound is derived and demonstrated. The demonstrations are based on two-dimensional polynomial examples. The last section offers concluding remarks.

Response surface approximations, for which the fundamentals are reviewed in Appendix A, fit numerical or physical experimental data with an analytical model. The true function generating the data at a given design point x or the true mean response $\eta(x)$ is represented here as

$$\eta(x) = F^{(1)}(x)^T \beta^{(1)} + F^{(2)}(x)^T \beta^{(2)}, \quad (1)$$

where $F^{(1)}(x)$ and $F^{(2)}(x)$ are the vectors of basis functions f_j [see Appendix A, Eq. (A2)], and $\beta^{(1)}$ and $\beta^{(2)}$ are the coefficient vectors associated with the basis function vectors. The superscript (1) denotes the sets of basis functions in the linear regression procedure. The superscript (2) denotes terms on the true function Eq. (1) that are missing from the regression (often due to the cost of a comprehensive data matrix, in particular for high dimensional problems). As the present paper deals with the bias error due to partial characterization of the analytical model, the predicted response at a given design point x is given as

$$\hat{y}(x) = F^{(1)}(x)^T b, \quad (2)$$

where b is the estimate of the coefficient vector $\beta^{(1)}$, which is biased by the presence of the $\beta^{(2)}$ in the true function. The true error in the approximation at a design point x is

$$e_b(x) = \eta(x) - \hat{y}(x). \quad (3)$$

In this study the true function is assumed to be a polynomial. Polynomials are popular in RSA, and in addition, locally, assuming smoothness, all functions look like polynomials according to Taylor's Theorem. Since one usually does not know the functional form of the true response, it is often assumed to be a higher order polynomial. For instance, for a quadratic fitting model in two variables when the true function is cubic,

$$\begin{aligned} F^{(1)}(x) &= (1 \quad x_1 \quad x_2 \quad x_1^2 \quad x_1 x_2 \quad x_2^2)^T, \\ \beta^{(1)} &= (\beta_1^{(1)} \quad \beta_2^{(1)} \quad \beta_3^{(1)} \quad \beta_4^{(1)} \quad \beta_5^{(1)} \quad \beta_6^{(1)})^T, \\ F^{(2)}(x) &= (x_1^3 \quad x_1^2 x_2 \quad x_1 x_2^2 \quad x_2^3)^T, \\ \beta^{(2)} &= (\beta_1^{(2)} \quad \beta_2^{(2)} \quad \beta_3^{(2)} \quad \beta_4^{(2)})^T. \end{aligned} \quad (4)$$

Assuming that there will be no noise in the data y , and the error is only associated with the modelling, the true response at the data points is

$$\eta = y = X^{(1)} \beta^{(1)} + X^{(2)} \beta^{(2)}, \quad (5)$$

where $X^{(1)}$ and $X^{(2)}$ are the design matrices corresponding to the basis functions in $F^{(1)}$ and $F^{(2)}$. The coefficient vector b in Eq. (2), that is, the biased estimates of the coefficient vector $\beta^{(1)}$, becomes (after substitution of Eq.(5) into Eq.(A7) with $X = X^{(1)}$)

$$b = \beta^{(1)} + A\beta^{(2)}, \quad (6)$$

where $A = [(X^{(1)})^T X^{(1)}]^{-1} (X^{(1)})^T X^{(2)}$ is called the alias matrix. Substitution of Eq.(6) into Eq.(2) yields

$$\hat{y}(x) = F^{(1)}(x)^T (\beta^{(1)} + A\beta^{(2)}). \quad (7)$$

Unlike some previous work [Papila and Haftka 2001 and Papila et al. 2002], used here is the absolute bias error (rather than the squared error)

$$\begin{aligned} |e_b(x)| &= |F^{(1)}(x)^T \beta^{(1)} + F^{(2)}(x)^T \beta^{(2)} - F^{(1)}(x)^T (\beta^{(1)} + A\beta^{(2)})|, \\ |e_b(x)| &= |m^T \beta^{(2)}|, \end{aligned} \quad (8)$$

where

$$m = F^{(2)}(x) - A^T F^{(1)}(x). \quad (9)$$

Recall that one of the reasons why bias error variation is of interest is to be able to identify points where the bias error has the potential to be large. Given $\beta^{(2)}$, Eq. (8) is the exact bias error at a point of interest. Therefore, defining or assuming a bound on $\beta^{(2)}$, which is unknown, may provide a tool for comparing the worst magnitude of Eq. (8) at the design points, and determine relatively which regions may suffer the most due to an insufficient model.

POINT TO POINT BIAS ERROR BOUND PRIOR TO DATA GENERATION

The first approach dealing with the absolute bias error Eq. (8) makes use of the information in a given design of experiment (DOE) and the assumed true function form, but does not use the function data. That is, the error bound developed here may be used prior to data generation. For this reason, it is called the data independent worst case bias error bound.

It is assumed that whatever terms are missing from the fitting model are known, but there will not be enough data (for instance due to computational or experimental cost) to calculate the corresponding coefficients $\beta^{(2)}$. If one bounds the terms in $\beta^{(2)}$, it is possible to formulate a maximization problem for the largest absolute bias error, Eq.(8), that may be experienced at any given design point for the worst possible $\beta^{(2)}$ of the bound:

$$\max_{-c^{(2)} \leq \beta^{(2)} \leq c^{(2)}} |e_b(x)|, \quad (10)$$

where $c^{(2)} \geq 0$ is the bound on the $\beta^{(2)}$ terms. As the bias error, Eq. (8), is expressed by the scalar product of the two vectors m [Eq. (9)] and $\beta^{(2)}$, the solution of the maximization problem is

$$|e_b^I(x)| = \sum_{j=1}^{n_2} \text{sgn}(m_j) m_j c_j^{(2)}, \quad (11)$$

where superscript I labels the data independent approach, m_j is the j th component of the vector m calculated at x , and n_2 is the size of the vectors, i.e., the number of missing coefficients. Then the worst possible vector is $\beta^{(2)} = (\text{sgn}(m_1)c_1^{(2)}, \dots, \text{sgn}(m_{n_2})c_{n_2}^{(2)})^T$, that characterizes the coefficients of the missing basis functions resulting in the largest bias error when fitting only with b . The worst $\beta^{(2)}$ and therefore the true function corresponding to the worst possible bias error vary from point to point. The solution $|e_b^I(x)|$ reflects the assumed form of the true function with the basis function coefficients $\beta^{(2)}$ (among all the possible combinations such that $-c^{(2)} \leq \beta^{(2)} \leq c^{(2)}$) causing the largest error. It is viewed as a bound on the bias error given the assumptions on the true model and a bound $c^{(2)}$ on the terms in $\beta^{(2)}$. If the coefficient bound is scaled, $|e_b^I(x)|$ will simply scale up in proportion without changing the locations identified as having high error potential. Here, the bound vector is taken as $c^{(2)} = 1$. Equation (11) is easy to compute, although the number of coefficients that are variables for the optimization problem increases quadratically with the dimension of the problem.

USE OF ERROR BOUND AS A DESIGN OF EXPERIMENT CRITERION—MIN-MAX BIAS DESIGN

The data independent error measure is obtained prior to data generation for a given DOE. That means it may easily be used to compare different choices of DOE's. Ultimately, it can be used to search for the DOE minimizing the maximum point to point bias error bound, $|e_b^I|_{max}$ using Eq. (11),

$$|e_b^I|_{max} = \max_{-1 \leq x_1, x_2, \dots, x_n \leq 1} |e_b^I(x)|. \quad (12)$$

The experimental designs derived here minimize the maximum bound given by Eq. (12) instead of the average square error, as is done with minimum bias designs [Myers and Montgomery (1995), Khuri and Cornell (1996) and Venter and Haftka (1997)]. The difference is analogous to H^∞ control versus H^2 control.

Consider a design of experiments in an n -dimensional cube, $-1 \leq x_1, x_2, \dots, x_n \leq 1$, for which the data points are determined by parameters $0 \leq \alpha_1, \alpha_2, \dots, \alpha_m \leq 1$. For instance, two-level factorial points associated with the parameter α_m in the cube are 2^n combinations at $x_i = \pm\alpha_m$ for $i = 1, \dots, n$. Figure 1 shows two examples for DOE as a function of the parameters α_m in two dimensions.

The minimum of the maximum absolute bias error given by Eq. (12) is referred to as *min-max bias design*, and can be obtained by solving

$$\min_{0 \leq \alpha_1, \alpha_2, \dots, \alpha_m \leq 1} |e_b^I|_{max}. \quad (13)$$

The min-max bias design from Eq. (13) is compared with the standard DOE's: minimum variance design and minimum bias design that minimizes the average square error. The metrics or error measures for the comparison of the experimental designs are

- i) maximum absolute bias error, $|e_b^I|_{max}$, as given in Eq. (12) (obtained from Eq. (11) with $c^{(2)} = 1$);
- ii) average of absolute bias error, $|e_b^I|_{av}$ (from Eq. (11) with $c^{(2)} = 1$) over the domain $-1 \leq x_1, x_2, \dots, x_n \leq 1$,

$$|e_b^I|_{av} = \int |e_b^I(x)| dx; \quad (14)$$

- iii) maximum normalized estimated standard error (square root of the prediction variance which predicts sensitivity to noise), $[\bar{e}_{es}]_{max}$ (see the Appendix A, maximum of Eq. (A9) normalized by s),

$$[\bar{e}_{es}]_{max} = \max_{-1 \leq x_1, x_2, \dots, x_n \leq 1} e_{es}(x)/s. \quad (15)$$

Two polynomial examples are presented for demonstration: first order polynomial fit for a two-dimensional quadratic polynomial (linear-quadratic example), and quadratic fit for a two-dimensional cubic polynomial (quadratic-cubic example).

Example I: min-max bias design of linear fit for two-dimensional quadratic polynomial

A two-variable quadratic true function fitted by a first order polynomial is considered first:

$$\begin{aligned} \eta(x) &= \beta_1^{(1)} + \beta_2^{(1)}x_1 + \beta_3^{(1)}x_2 + \beta_1^{(2)}x_1^2 + \beta_2^{(2)}x_1x_2 + \beta_3^{(2)}x_2^2, \\ \hat{y}(x) &= b_1 + b_2x_1 + b_3x_2. \end{aligned} \quad (16)$$

The four data points (Figure 1a) are defined by a single variable α_1 that determines the experimental design within the coded square domain of $-1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1$. The design matrices are given as

$$X^{(1)} = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_1 \\ 1 & -\alpha_1 & \alpha_1 \\ 1 & \alpha_1 & -\alpha_1 \\ 1 & \alpha_1 & \alpha_1 \end{bmatrix}, \quad X^{(2)} = \begin{bmatrix} \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \end{bmatrix}. \quad (17)$$

The minimum noise variance design is at the four vertices of the square domain ($\alpha_1 = 1$), while the standard minimum bias design (that minimizes the average square error) is at $(x_1, x_2) \approx$

$(\pm 0.577, \pm 0.577)$ with $\alpha_1 = \frac{\sqrt{3}}{3}$ [Myers and Montgomery p. 415 (1995)]. In contrast, the min-max bias design, as the solution of Eq. (13), is found at $\alpha_1 = \frac{\sqrt{3}}{2} \approx 0.866$. In other words, the maximum potential bias error in a linear model for a true quadratic response is minimized if the experimental design is $(x_1, x_2) \approx (\pm 0.866, \pm 0.866)$. Table 1 compares the minimum noise variance, minimum bias, and minimum maximum bias error (min-max bias) experimental designs in terms of the three error measures. It shows that the maximum of the data independent bound $|e_b^I|_{max}$, Eq. (12) is highest for the minimum bias design and lowest for the design where it was minimized in the domain. Average absolute bias error $|e_b^I|_{av}$, Eq. (14) reported in Table 1 is lowest for the minimum bias design as it minimizes the bias error in an average sense. The contour plots of the data independent bound are presented in Figure 2. Comparisons of Figure 2b and Figure 2c shows that the data independent error bound levels are generally lower, but the maximum error can be high on the boundary. This is due to the data points being much farther from the boundary for the minimum average bias design. In addition, the min-max bias design is much less sensitive to noise variance, as measured by $[\bar{e}_{es}]_{max}$, Eq. (15), than the minimum bias design (similar to an H^∞ optimal controller)

Example II: min-max bias design of quadratic fit for two-dimensional cubic polynomial

The second two-variable problem is a cubic polynomial fitted by a quadratic polynomial:

$$\begin{aligned} \eta(x) &= \beta_1^{(1)} + \beta_2^{(1)}x_1 + \beta_3^{(1)}x_2 + \beta_4^{(1)}x_1^2 + \beta_5^{(1)}x_1x_2 + \beta_6^{(1)}x_2^2 \\ &\quad + \beta_1^{(2)}x_1^3 + \beta_2^{(2)}x_1^2x_2 + \beta_3^{(2)}x_1x_2^2 + \beta_4^{(2)}x_2^3, \\ \hat{y}(x) &= b_1 + b_2x_1 + b_3x_2 + b_4x_1^2 + b_5x_1x_2 + b_6x_2^2. \end{aligned} \quad (18)$$

The experimental design now includes, in addition to center point (0,0), 4 factorial and 4 axial points determined by parameters α_1 , and α_2 , respectively (Figure 1b). The design matrices are given as

$$X^{(1)} = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_1 & \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \\ 1 & -\alpha_2 & 0 & \alpha_2^2 & 0 & 0 \\ 1 & -\alpha_1 & \alpha_1 & \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ 1 & 0 & -\alpha_2 & 0 & 0 & \alpha_2^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \alpha_2 & 0 & 0 & \alpha_2^2 \\ 1 & \alpha_1 & -\alpha_1 & \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ 1 & \alpha_2 & 0 & \alpha_2^2 & 0 & 0 \\ 1 & \alpha_1 & \alpha_1 & \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \end{bmatrix}, \quad X^{(2)} = \begin{bmatrix} -\alpha_1^3 & -\alpha_1^3 & -\alpha_1^3 & -\alpha_1^3 \\ -\alpha_2^3 & 0 & 0 & 0 \\ -\alpha_1^3 & \alpha_1^3 & -\alpha_1^3 & \alpha_1^3 \\ 0 & 0 & 0 & -\alpha_2^3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2^3 \\ \alpha_1^3 & -\alpha_1^3 & \alpha_1^3 & -\alpha_1^3 \\ \alpha_2^3 & 0 & 0 & 0 \\ \alpha_1^3 & \alpha_1^3 & \alpha_1^3 & \alpha_1^3 \end{bmatrix}. \quad (19)$$

The minimum noise variance design is achieved at $\alpha_1 = \alpha_2 = 1$. The standard minimum bias design, minimizing the average square error, is with $\alpha_1 = 0.700$ and $\alpha_2 = 0.770$ [due to Qu et al.

(2004)]. The min-max bias design is now obtained as the solution of Eq. (13) at $\alpha_1 = \alpha_2 = 0.949$. Table 2 compares the minimum noise variance, minimum average bias, and minimum maximum bias error (min-max bias) experimental designs in terms of three error measures.

As in Example I, the maximum of the data independent bound $|e_b^I|_{max}$, Eq. (12) is the highest for the minimum bias design and lowest for the design where it was minimized in the domain. The contour plots for the data independent bound are presented in Figure 3. Comparisons of Figure 3b and Figure 3c reveal the fact that the minimum bias design minimizes the bias error in an average sense (see Table 2 for average $|e_b^I|_{av}$), and so the data independent error bound levels are generally lower, but the maximum error can be high. The two examples demonstrate that each of three design criteria produces a design that is optimal with respect to the criterion that produced it. The min-max bias experimental design, however, that was found via the data independent bias error bound appears to be a good compromise between the minimum noise variance and minimum bias designs: it is much less sensitive to noise variance as measured by $[\bar{e}_{es}]_{max}$ than the minimum bias design.

The squared error approach of Papila and Haftka (2001) was previously integrated in a design space windowing procedure (Papila et al. 2002) where the regions of interest were zoomed by the prediction of a global RSA. Then, additional design points in the zoomed area were selected using a squared error based data independent error bound. The present bias error bounds are equally applicable for such adaptive design of experiment applications where more data is progressively generated.

POINT TO POINT BIAS ERROR BOUND POSTERIOR TO DATA GENERATION

Once data is available, the error bounds can be tightened and the data dependent worst case bias error bounds can be obtained. That is, sharper bounds on the bias error can now be obtained by constraining $\beta^{(2)}$ to satisfy the data. This restricts consideration to all possible polynomials with coefficients that interpolate the given data points $(x^{(i)}, y_i)$ and result in the approximation $\hat{y}(x)$ associated with the data.

The bias error of the approximation $\hat{y}(x)$ at a design point x is given by Eq. (8). This relation along with Eq. (9) leads to the prediction error vector due to bias at the data points $x^{(i)}$,

$$E_b = y - \hat{Y} = (X^{(2)} - X^{(1)}A)\beta^{(2)}, \quad (20)$$

where \hat{Y} is the vector of predictions $\hat{y}(x^{(i)})$ by the response surface approximation at the data points.

The optimization problem now is to find the maximum absolute error at a design point subject to constraints on the magnitude of the coefficients and $\eta(x)$ matching the given data. As the latter constraint is equivalent to matching the prediction error vector at the data points, Eq. (20), the maximization of the error may be formulated as a linear program (LP),

$$\max_{-c^{(2)} \leq \beta^{(2)} \leq c^{(2)}} |e_b(x)| \quad \text{such that} \quad (X^{(2)} - X^{(1)}A)\beta^{(2)} = y - \hat{Y}. \quad (21)$$

The solution of the system Eq. (21) requires solving two LP problems, one for the minimum and one for the maximum error $e_b(x)$, and the solution is denoted as $|e_b^D(x)|$ for data y (superscript D labels the data dependent bound). Like the data independent error bound, this is also a worst case search (and hence a bound on the bias error) $|e_b^I(x)|$. The computational effort required for the data dependent bound may grow large as the number of coefficients $\beta^{(2)}$ that are variables for the optimization (LP) problems increases with the dimension of the problem.

Selection of the bounds on the coefficients is an important part of the present approach. The data independent error bound does not make use of information on the function values, and so it makes sense to have the same coefficient bounds $c^{(2)}$ [Eq. (10)] on all the coefficients, and the magnitude is immaterial in terms of a relative or qualitative comparison over the domain. In contrast, the choice of the bounds $c^{(2)}$ in the data dependent error bound [Eq. (21)] must be compatible with the data. If variables are normalized or coded so that all the data satisfy $\|x\|_\infty \leq 1$, one logical selection criterion is to use the error magnitudes at the data points as an indication of the magnitude of the missing coefficients. This fails, however, when the errors at the data points are all small. Another indication is the magnitude of the coefficients b . With normalized (coded) variables, one could assume that the coefficients $\beta^{(2)}$ have magnitudes that are a fraction of the largest coefficient of b . Finally, one could assume that the missing coefficients are a fraction of the maximum function value minus the value of the constant coefficient b_1 .

Note that the LP problem with the equality constraints as given in Eq. (21) may not always be feasible because for many standard design of experiments the rows of the matrix $[X^{(2)} - X^{(1)}A]$ may be linearly dependent and so the matrix is rank deficient. With such linear dependence, if the data is not exactly consistent with the assumed true model (due to noise or small deviations of the true function from the assumed true model) the equalities cannot be satisfied. This can be resolved by relaxing the equality constraint into a pair of inequality constraints with a small tolerance. Such rank deficient DOEs have the advantage that they provide information on the magnitude of deviation from the assumed true model.

The two polynomial examples of the previous section are used for demonstration along with the min-max bias designs developed. Note, however that such experimental designs are not prerequisite or essential for the use of data dependent error bounds. Also, recall that the data dependent error bounds $|e_b^D(x)|$ are worst case measures as they characterize the worst possible polynomials point to point subject to available data. Therefore, effective demonstration of the usefulness of the error bounds must consider a large number of quadratic and cubic polynomials in the first and second examples, respectively.

For the next two example cases, five different sets of data and 100 different polynomials for each set were randomly selected. This gives 100 randomly selected polynomials per data set to evaluate the data dependent bound. The actual RSA error at x for the l th polynomial with the k th data set is denoted by

$$|e_b(x)|_{kl} = |\eta(x) - \hat{y}(x)|_{kl}, \quad (22)$$

where the first index k denotes the data set (A, B, C, D, or E), and the second index $l = 1, 2, 3, \dots, 100$ denotes the polynomial. Then the worst case error at x for the k th data set and the overall worst case error at x , respectively, are

$$|e_b(x)|_{k \max} = \max_l (|\eta(x) - \hat{y}(x)|_{kl}), \quad (23)$$

$$|e_b(x)|_{\text{overall}} = \max_k (|e_b(x)|_{k \max}). \quad (24)$$

Example I: linear fit for two-dimensional quadratic polynomial using the min-max bias design

For this example, with only bias error, substituting Eq. (17) into Eq. (20), the absolute prediction error at the four data points depends on $\beta_2^{(2)}$ alone (equal to $\alpha_1^2 \beta_2^{(2)}$). As a result, selecting the data randomly is equivalent to random selection of coefficients b_1 , b_2 , b_3 and $\beta_2^{(2)}$. Once the data is selected, different polynomials that interpolate this data can be obtained by generating random $\beta_1^{(2)}$ and $\beta_3^{(2)}$ which also determine $\beta^{(1)}$ since the RSA coefficients, due to Eq. (6), are expressed as

$$\begin{aligned} b_1 &= \beta_1^{(1)} + \alpha_1^2 (\beta_1^{(2)} + \beta_3^{(2)}), \\ b_2 &= \beta_2^{(1)}, \\ b_3 &= \beta_3^{(1)}. \end{aligned} \quad (25)$$

Appendix B describes the details of the selection of data sets and associated random polynomials and the bounds $c^{(2)}$ on the coefficients.

The data dependent error bound $|e_b^D(x)|$ was computed using Eq. (21) for data sets A, ..., E (associated 100 random polynomials each) and compared to the actual error distribution. The contour plots for data sets C and E presenting the actual worst case errors (over associated 100 polynomials each) and the data dependent error bound are given in Figure 4 and Figure 5, respectively. They show good agreement and suggest that the error bound determines successfully the map of the worst case errors in the design space due to the available approximation. It suggests that the RSA may be vulnerable to prediction error at the center of the design domain. The corresponding coefficients of correlation between the error bounds and worst errors are calculated over a 11x11 uniform grid points and given in Table 3.

Note that the total of 500 polynomials (100 for each data set A–E) is indeed a subset of the quadratic polynomials of the $-c^{(2)} \leq \beta^{(2)} \leq c^{(2)}$ ($c^{(2)} = 1$ as described in Appendix B). Therefore, the maximum of the errors among all 500 polynomials may be characterized by the data independent error bound $|e_b^I(x)|$ for the min-max bias experimental design (from Eq. (11) as shown also in Figure 3c). Correlations of data independent error bounds with the worst, actual errors for the k th individual data set $|e_b(x)|_{k \max}$ and the worst of all 500 polynomials $|e_b(x)|_{\text{overall}}$, Eq. (24), are also presented in Table 3. As expected, coefficients of correlation of $|e_b^I(x)|$ with actual error for individual data sets are lower than the correlations obtained for the data dependent bound $|e_b^D(x)|$. The coefficient of correlation increases for the worst of all 500 polynomials, as the data matching constraint averages out. The contour plots for $|e_b(x)|_{\text{overall}}$ and the data independent bound $|e_b^I(x)|$ are shown in Figure 6.

Example II: quadratic fit for two-dimensional cubic polynomial using the min-max bias design

The example of DOE at $\alpha_1 = \alpha_2 = \alpha = 0.949$, with only bias error, Eq. (20), results in the prediction error at the data points as a function of only $\beta_2^{(2)}$ and $\beta_3^{(2)}$, and the RSA coefficients via Eq. (6) will be

$$\begin{aligned}
 b_1 &= \beta_1^{(1)}, \\
 b_2 &= \beta_2^{(1)} + \alpha^2(\beta_1^{(2)} + \frac{2}{3}\beta_3^{(2)}), \\
 b_3 &= \beta_3^{(1)} + \alpha^2(\frac{2}{3}\beta_2^{(2)} + \beta_4^{(2)}), \\
 b_4 &= \beta_4^{(1)}, \\
 b_5 &= \beta_5^{(1)}, \\
 b_6 &= \beta_6^{(1)}.
 \end{aligned} \tag{26}$$

Similar to Example I, randomly selecting the data is equivalent to random selection of the coefficients $b_1, b_2, b_3, b_4, b_5, b_6, \beta_2^{(2)}$, and $\beta_3^{(2)}$. The details of the selection of data sets and associated random polynomials and the bounds $c^{(2)}$ on the coefficients in this example are also presented in Appendix B.

The correlations between the data dependent error bound $|e_b^D(x)|$, Eq. (21), and the worst actual error distribution $|e_b(x)|_{k_{\max}}$, Eq. (23), on a 11×11 grid are given in Table 4. The data dependent error bound and actual worst error correlated very well. Figure 7 and Figure 8 present the contour plots for data sets B and D, respectively, presenting the actual worst errors (over associated 100 polynomials each) and the data dependent error bound. They also suggest that the data dependent error bound successfully determines the worst case error variation that may be considered as map of potential high bias error locations.

The maximum of the errors among all 500 polynomials (100 for each data set A–E) may again be characterized by the data independent error bound $|e_b^I(x)|$ for the min-max bias experimental design (from Eq. (11) as shown also in Figure 3c). Table 4 also presents the correlations for the data independent error bound. Lower correlations with the worst actual errors $|e_b(x)|_{k_{\max}}$ for the individual data sets demonstrate the benefit of taking data into account. As expected, correlation of the data independent error bound with the actual worst errors increases when polynomials for different data are included. It is about 0.97 for the worst of 500 polynomials $|e_b(x)|_{\text{overall}}$, Eq. (24). The contour plots for $|e_b(x)|_{\text{overall}}$ and the data independent bound $|e_b^I(x)|$ are shown in Figure 9.

The benefit of data available to tighten the bounds may be observed by comparing Figures 7b, 8b and 9b. The data independent error bounds in Figure 9b suggest, for instance, design points at $(-0.5, -0.5)$, $(-0.5, +0.5)$, $(0.5, -0.5)$ and $(0.5, 0.5)$ may experience high errors. This seems to be also true for the data set D as in Figure 8b. Figure 7b for data set B, however, shows the bounds at points $(-0.5, +0.5)$ and $(0.5, -0.5)$ are tightened. That is, much less danger is predicted, as also observed in 100 polynomials of data B.

Note that the consistency in selection of the bounds for the uniform distributions and the bounds for coefficients used in optimization provides a quantitative assessment by the error bounds in the present polynomial examples (see Appendix B). Their use, however, should be considered qualitative, in general.

CONCLUSIONS

Two error bounds were developed in order to characterize the point to point variation of error due to an incorrect data model (bias error). A data independent error bound is computed prior to data generation and depends only on the design of experiments, the true and least square fit models. It seeks the worst case absolute error, which varies point to point. The data independent bound may be used to obtain experimental designs that minimize the maximum bias error. It appears from the examples in this paper that such a min-max bias experimental design is a good compromise between traditional minimum noise variance and minimum bias designs.

The data dependent bound also considers the worst case absolute errors, but it is tightened by the available data. That is, the essential difference between the two error bounds is whether the coefficients of the permissible true functions are bounded without and with the data constraint. Comparisons of the actual worst errors and the error bounds for randomly selected polynomial examples showed that the data dependent error bound successfully determines the map of the worst case prediction errors in the design space due to the available approximation. This information may be used to adaptively refine a design of experiments, or to steer clear of the identified potential high bias error regions in the design domain.

The examples here are limited to polynomial true functions. The tools proposed, however, are still relevant. Even if the true function is not a polynomial, it is very well approximated locally by a cubic function, for instance, that is a better approximation than a quadratic to the true function. Therefore, these bias error bounds, locally in particular, may be a valuable source of information about the true bias, even when the correct form of the true function is unknown.

ACKNOWLEDGEMENTS

This research was supported by NASA Grant NAG1-02042 and AFOSR Grant F49620-02-1-0090.

REFERENCES

- Balabanov, V.O., Giunta, A.A., Golovidov, O., Grossman, B., Mason, W.H., Watson, L.T., and Haftka, R.T., "Reasonable Design Space Approach to Response Surface Approximation", *Journal of Aircraft* 36(1), pp. 308-315, 1999.
- Balabanov, V., Kaufman, M., Knill, D.L., Haim, D., Golovidov, O., Giunta, A.A., Haftka, R.T., Grossman, G., Mason W.H., and Watson, L.T., "Dependence of Optimal Structural Weight on Aerodynamic Shape for a High Speed Civil Transport", in *Proceedings, 6th AIAA/NASA/USAF Symposium on Multidisciplinary Analysis and Optimization*, Bellevue, WA, AIAA paper 96-4046, pp.599-612, September 1996.
- Box, G.E.P. and Draper, N.R., "The choice of a second order rotatable design", *Biometrika* 50 (3), pp. 335-352, 1963.
- Davis, P.J., *Interpolation and Approximation*, New York: Dover, 1975.
- Draper, N. R. and Lawrence, W. E., "Designs which minimize model inadequacies: cuboidal regions of interest", *Biometrika* 52 (1-2), pp.111-118, 1965.

- Fedorov V. V., Montepiedra, G. and Nachtsheim C. J., “Design of experiments for locally weighted regression”, *Journal of Statistical Planning and Inference* 81 , pp.363–383, 1999.
- Khuri, A.I. and Cornell, J.A., *Response Surfaces: Designs and Analyses, 2nd edition*, New York, Marcel Dekker Inc., pp.207–247, 1996.
- Kaufman, M., Balabanov, V., Burgee, S.L., Giunta, A.A., Grossman, B., Haftka, R.T., Mason, W.H. and Watson, L.T., “Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design”, *Computational Mechanics* 18, pp.112–126, 1996.
- Kupper, L. L. and Meydrecht, E. F., “A new approach to mean squared error estimation of response surfaces”, *Biometrika* 60 (3), pp.573–579, 1973.
- Montepiedra, G. and Fedorov V. V., “Minimum bias designs with constraints”, *Journal of Statistical Planning and Inference* 63 , pp.97–111, 1997.
- Myers, R.H., and Montgomery, D.C., *Response Surface Methodology—Process and Product Optimization Using Designed Experiments*, New York: John Wiley & Sons, Inc., pp.208–279, 1995.
- Palmer, K., and Tsui, K.L., “A minimum bias Latin hypercube design”, *Institute of Industrial Engineers Transactions* 33 (9), pp.793, 2001.
- Papila, M. and Haftka, R.T., “Uncertainty and Wing Structural Weight Approximations”, in *Proceedings, 40th AIAA/ASME/ASCE/ASC Structures, Structural Dynamics, and Material Conference*, St. Louis, MO, Paper AIAA-99-1312, pp.988–1002, April 1999.
- Papila, M. and Haftka, R.T., “Response Surface Approximations: Noise, Error Repair and Modeling Errors”, *AIAA Journal* 38(12), pp.2336–2343, 2000a.
- Papila, M. and Haftka, R.T., “Uncertainty and Response Surface Approximations”, in *Proceedings, 42nd AIAA/ASME/ASCE/ASC Structures, Structural Dynamics, and Material Conference*, Seattle, WA, Paper AIAA-01-1680, April 2001.
- Papila, M., Papila N., Shyy, W., Haftka, R.T and Fitz-coy, N., “Error-based Design Space Windowing”, in *Proceedings, 40th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, Paper AIAA-02-0539, January 2002.
- Papila, N., Shyy, W., Griffin, L., and Dorney, D.J., “Shape Optimization of Supersonic Turbines Using Response Surface and Neural Network Methods”, in *Proceedings, 39th AIAA Aerospace Sciences Meeting & Exhibit*, Paper No. 2001-1065, 2001.
- Qu, X., Venter, G, and Haftka, R.T., “New Formulation of a Minimum-Bias Experimental Design based on Gauss Quadrature”, *Structural and Multidisciplinary Optimization*, to appear 2004.
- Redhe, M., Forsberg, J., Jansson, T., Marklund, P.O., Nilsson, L., “Using the response surface methodology and the D-optimality criterion in crashworthiness related problems—An analysis of the surface approximation error versus the number of function evaluations”, *Structural and Multidisciplinary Optimization* 24(3), pp.185–194, 2002.
- Shyy, W., Tucker, P.K., and Vaidyanathan, R., “Response surface and neural network techniques for rocket engine injector optimization”, *Journal of Propulsion and Power* 17(2), pp.391–401, 2001.
- Venter, G., and Haftka R.T., “Minimum-Bias Based Experimental Design for Constructing Response Surfaces in Structural Optimization”, in *Proceedings, 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Kissimmee, FL, Paper AIAA-97-1053, Part 2, pp.1225–1238, April 1997.
- Welch, W. J., “A mean squared error criterion for the design of experiments”, *Biometrika* 70 (1), pp.205–213, 1983.

APPENDIX A: RESPONSE SURFACE METHODOLOGY

A scalar function $\eta(x)$, $x \in E^n$ (E^n is real n -dimensional Euclidean space) is observed at N points $x^{(i)}$, $i = 1, \dots, N$, and the observations y_i are contaminated by random variations ε_i such as measurement error, so

$$y_i = \eta(x^{(i)}) + \varepsilon_i. \quad (A1)$$

The random errors ε at different points $x \in E^n$ are assumed to be uncorrelated and normally distributed random variables with zero mean and variance σ^2 that is the same at all points. The true

mean response is assumed to be given in terms of basis functions $f_j(x)$ and associated coefficients β_j as

$$\eta(x) = \sum_{j=1}^{n_1} \beta_j f_j(x) = F(x)^T \beta, \quad (A2)$$

where $F(x)^T = (f_1(x), f_2(x), \dots, f_{n_1}(x))$ and $\beta^T = (\beta_1, \beta_2, \dots, \beta_{n_1})$. The n_1 basis functions f_j are often monomials. The least squares approximation based on $N > n_1$ observations is

$$\hat{y}(x) = \sum_{j=1}^{n_1} b_j f_j(x) = F(x)^T b, \quad (A3)$$

where the expected value of b_j (vector b) is the β_j (vector β). The difference (residual) between the data y and the estimate $\hat{y}(x)$ defined in Eq.(A3) at point $x^{(i)}$ is

$$e_i = y_i - \hat{y}(x^{(i)}). \quad (A4)$$

For N observations $y_i, i = 1, \dots, N$, the residuals can be written in matrix form

$$e = y - Xb, \quad (A5)$$

where $e^T = (e_1, \dots, e_N)$, and X is the Gramian matrix with terms $X_{i,j} = f_j(x^{(i)})$. For instance, for a quadratic model in two variables x_1, x_2 with N data points $(x_1^{(1)}, x_2^{(1)}), \dots, (x_1^{(N)}, x_2^{(N)})$,

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & (x_1^{(1)})^2 & x_1^{(1)}x_2^{(1)} & (x_2^{(1)})^2 \\ 1 & x_1^{(2)} & x_2^{(2)} & (x_1^{(2)})^2 & x_1^{(2)}x_2^{(2)} & (x_2^{(2)})^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(i)} & x_2^{(i)} & (x_1^{(i)})^2 & x_1^{(i)}x_2^{(i)} & (x_2^{(i)})^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} & (x_1^{(N)})^2 & x_1^{(N)}x_2^{(N)} & (x_2^{(N)})^2 \end{bmatrix}. \quad (A6)$$

The coefficient vector b in Eq.(A5) that minimizes $\|e\|_2^2$ is (assuming X has full rank)

$$b = (X^T X)^{-1} X^T y. \quad (A7)$$

When the fitting model Eqs.(A1–A2) is valid (no bias error), an unbiased estimator s^2 (error mean square) for variance σ^2 of the noise is given as

$$s^2 = MS_E = \frac{y^T y - b^T X^T y}{N - n_1}. \quad (A8)$$

The positive square root of the prediction variance is usually used as an estimate of the prediction error at a design point x , also called estimated standard error, and is defined by (Myers and Montgomery, 1995)

$$e_{es}(x) = s \sqrt{F(x)^T (X^T X)^{-1} F(x)}, \quad (A9)$$

where the estimator s [Eq. (A8)] is often called in the engineering literature the root-mean-square-error (RMSE) predictor. The estimated standard error depends on the location of the design point. Furthermore, as Eq.(A2) is, in general, only an assumption about the true function $\eta(x)$, s will contain not only noise error, but also modeling (bias) error.

APPENDIX B: SELECTION OF RANDOM POLYNOMIALS AND COEFFICIENT BOUNDS FOR THE EXAMPLES

In Example I, selecting the data randomly is equivalent to random selection of coefficients b_1 , b_2 , b_3 , and $\beta_2^{(2)}$. Consider a uniform distribution $U(-1, +1)$ for coefficients b_2 , b_3 , and $\beta^{(2)}$. For the intercept or constant coefficient b_1 , the $U(5, 10)$ distribution is selected in order to obtain reasonable relative errors at the data points. The coefficients and associated distributions corresponding to the linear-quadratic example are summarized in Table B1. The randomly selected coefficients for data sets A to E are presented in Table B2. It is seen that the maximum absolute error at the data points $|e^{(i)}|_{max}$ is equal to the absolute value of $\alpha_1^2 \beta_2^{(2)}$ ($\alpha_1 = 0.866$). The coefficients $\beta^{(2)}$ are bounded conservatively at ± 1 ($c^{(2)} = 1$) by considering $|e^{(i)}|_{max}$ from Table B2.

In Example II, randomly selecting the data is equivalent to random selection of the coefficients b_1 , b_2 , b_3 , b_4 , b_5 , b_6 , $\beta_2^{(2)}$, and $\beta_3^{(2)}$. Uniform distributions used in this example are also given in Table B1. The data sets A to E characterized by b , $\beta_2^{(2)}$, and $\beta_3^{(2)}$ are presented in Table B3. The maximum absolute errors at the data points $|e^{(i)}|_{max}$ are also shown in the last column of the table. Then selecting $\beta_1^{(2)}$ and $\beta_4^{(2)}$ randomly provides 100 polynomials for each data set satisfying the data. The coefficients $\beta^{(2)}$ are bounded conservatively at ± 1 ($c^{(2)} = 1$) by considering $|e^{(i)}|_{max}$ from Table B3.

Table 1: Design of experiments, Example I: linear-quadratic example (Figure 1a), maximum normalized standard error $[\bar{e}_{es}]_{max}$ [from Eq. (15)], average of absolute bias errors $|e_b^I|_{av}$ (from Eq. (14) for $c = 1$), and maximum absolute bias error (from Eq.(11) with $c = 1$) in $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$. First order polynomial fit at four data points $(-\alpha_1, -\alpha_1)$, $(-\alpha_1, +\alpha_1)$, $(+\alpha_1, -\alpha_1)$ and $(+\alpha_1, +\alpha_1)$.

Design of experiment	α_1	$[\bar{e}_{es}]_{max}$	$ e_b^I _{av}$ for $c = 1$	$ e_b^I _{max}$ for $c = 1$
Minimum noise variance	1.000	0.866	1.550	2.000
Minimum average bias	0.577	1.323	0.811	2.333
Minimum maximum bias	0.866	0.957	1.163	1.500

Table 2: Design of experiments, Example II: quadratic-cubic example (Figure 1b), Maximum normalized standard error $[\bar{e}_{es}]_{max}$ [from Eq. (15)], average of absolute bias errors $|e_b^I|_{av}$ (from Eq.(14) for $c = 1$), and maximum absolute bias error (from Eq. (11) with $c = 1$) in $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$.

Design of experiment	α_1, α_2	$[\bar{e}_{es}]_{max}$	$ e_b^I _{av}$ for $c = 1$	$ e_b^I _{max}$ for $c = 1$
Minimum noise variance	1.000,1000	0.898	0.860	1.147
Minimum average bias	0.700,0.770	1.944	0.656	2.331
Minimum maximum bias	0.949,0.949	0.994	0.793	1.000

Table 3: Linear-quadratic example: coefficients of correlation between error bounds and the actual worst errors calculated on a uniform 11×11 grid.

	$ e_b(x) _{A \max}$	$ e_b(x) _{B \max}$	$ e_b(x) _{C \max}$	$ e_b(x) _{D \max}$	$ e_b(x) _{E \max}$	$ e_b(x) _{\text{overall}}$
$ e_b^D(x) $	0.9953	0.9374	0.9206	0.9744	0.9966	N/A
$ e_b^I(x) $	0.7365	0.9417	0.9537	0.8877	0.5959	0.9650

Table 4: Quadratic-cubic example: coefficients of correlation between error bounds and the actual worst errors calculated on a 11×11 grid.

	$ e_b(x) _{A \max}$	$ e_b(x) _{B \max}$	$ e_b(x) _{C \max}$	$ e_b(x) _{D \max}$	$ e_b(x) _{E \max}$	$ e_b(x) _{\text{overall}}$
$ e_b^D(x) $	0.9786	0.9639	0.9869	0.9932	0.9826	N/A
$ e_b^I(x) $	0.7152	0.6743	0.7258	0.6403	0.7289	0.9701

Table B1: Description of uniformly distributed random coefficients for the example problems.

Uniform distribution	Examples	
	Linear-quadratic example	Quadratic-cubic example
$U(5, 10)$	b_1	b_1
$U(-1, +1)$	b_2, b_3	b_2, b_3, b_4, b_5, b_6
$U(-1, +1)$	$\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}$	$\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}$

Table B2: Linear-quadratic example: randomly selected coefficients for the data sets and the maximum absolute error at the data points.

Data Set	b_1	b_2	b_3	$\beta_2^{(2)}$	$ e^{(i)} _{\max}$
A	5.11	-0.40	0.79	0.37	0.28
B	8.57	0.53	0.89	0.94	0.71
C	5.10	-0.90	0.80	-0.84	0.63
D	9.88	-0.93	-0.84	-0.66	0.50
E	6.80	0.88	-0.08	-0.07	0.05

Table B3: Quadratic-cubic example: randomly selected coefficients for the data sets and the maximum absolute error at the data points.

Data Set	b_1	b_2	b_3	b_4	b_5	b_6	$\beta_2^{(2)}$	$\beta_3^{(2)}$	$ e^{(i)} _{\max}$
A	7.99	-0.45	-0.62	0.90	0.03	0.61	-0.96	0.64	0.54
B	7.72	0.64	-0.05	-0.78	-0.01	0.69	0.68	0.79	0.45
C	8.34	0.59	-0.81	0.62	0.38	0.67	-0.48	0.81	0.46
D	8.67	-0.83	-0.40	0.47	0.50	-0.22	0.13	0.88	0.50
E	6.78	-0.07	0.66	-0.18	-0.47	-0.01	-0.33	0.51	0.29

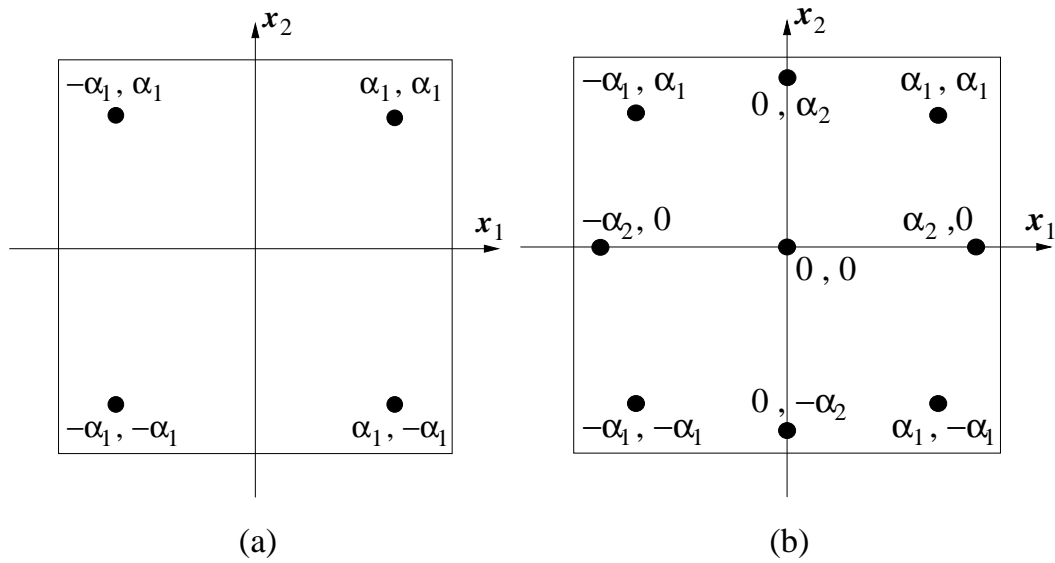
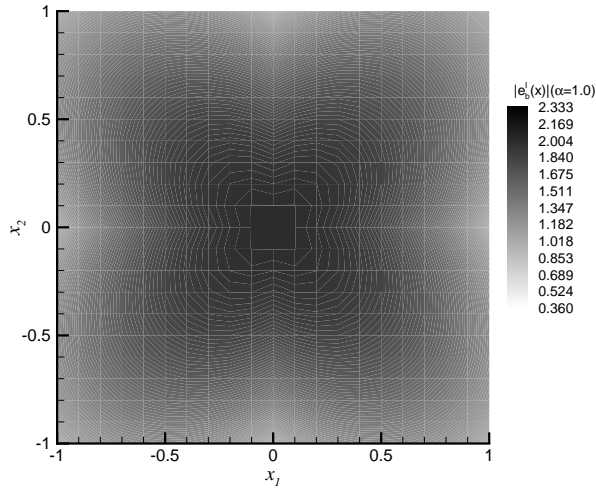
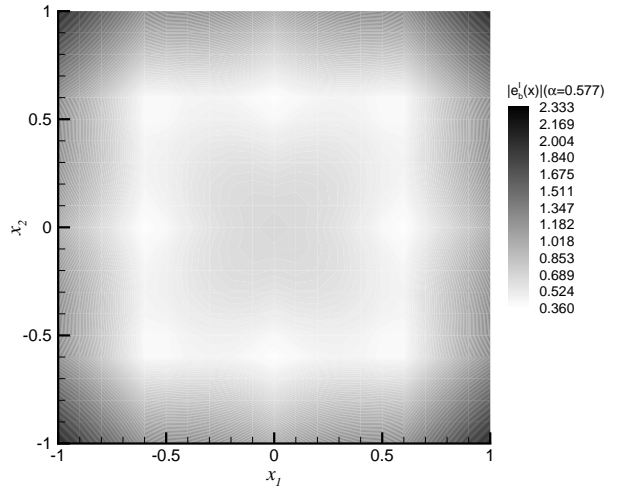


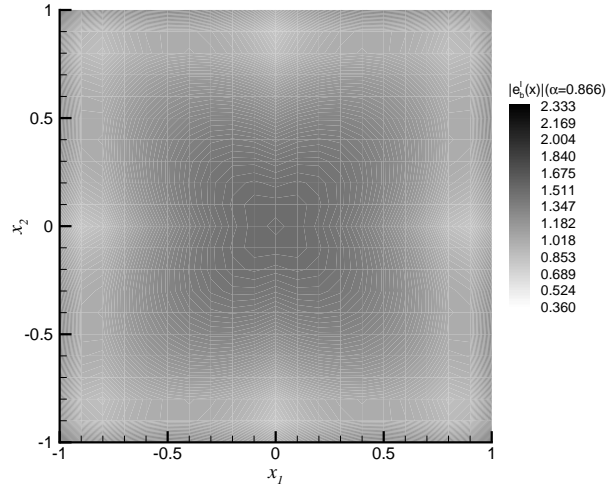
Figure 1: Design of experiment in coded square domain $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$. (a) Example 1: linear-quadratic example; (b) Example 2: quadratic-cubic example.



(a)



(b)



(c)

Figure 2: Data independent error bound in $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$ for linear-quadratic example (Figure 1a): (a) minimum noise variance design ($\alpha_1 = 1.0$), (b) minimum average bias design ($\alpha_1 = 0.577$), (c) minimum maximum bias design ($\alpha_1 = 0.866$).

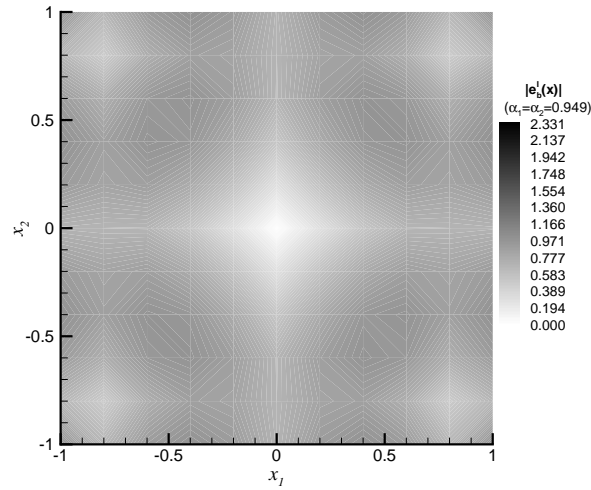
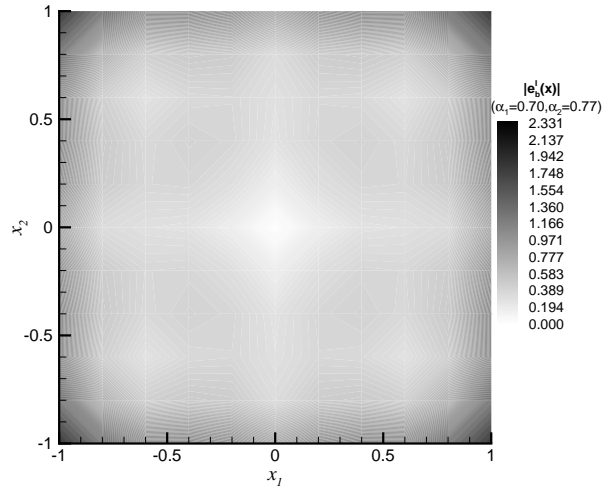
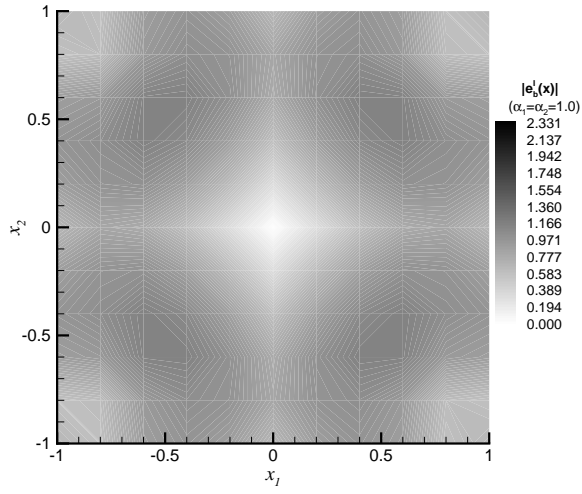
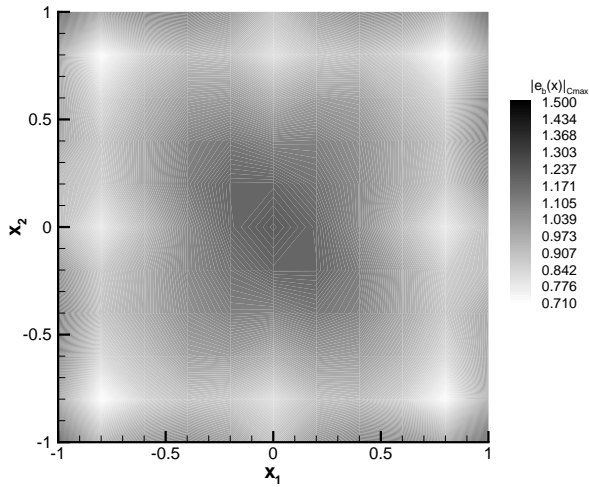
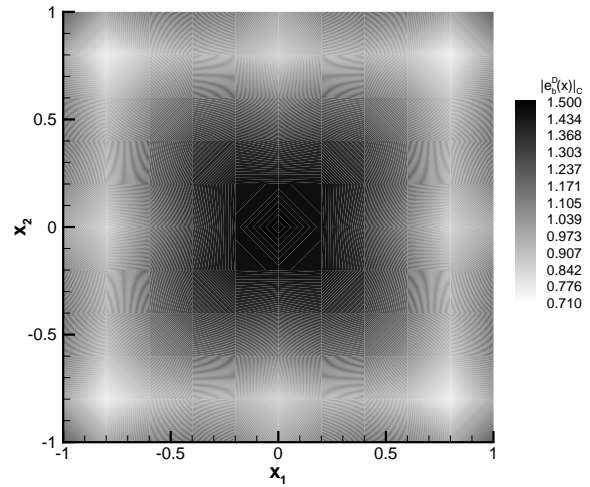


Figure 3: Data independent error bound in $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$ for quadratic-cubic example (Figure 1b): (a) minimum noise variance design ($\alpha_1 = \alpha_2 = 1.0$), (b) minimum average bias design ($\alpha_1 = 0.700, \alpha_2 = 0.770$), (c) minimum maximum bias design ($\alpha_1 = \alpha_2 = 0.949$).

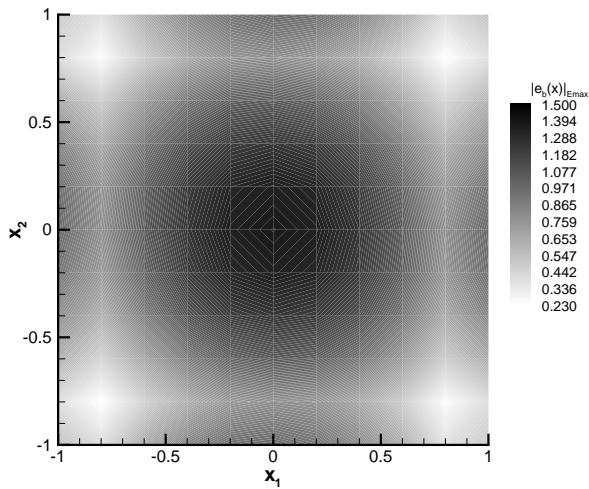


(a)

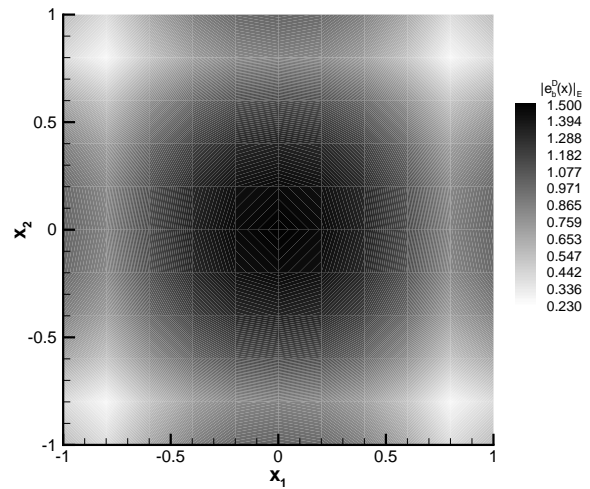


(b)

Figure 4: Comparison of absolute worst case errors for data set C in Example I: (a)worst of 100 polynomials, (b) data dependent bound, $|e_b^D(x)|$.

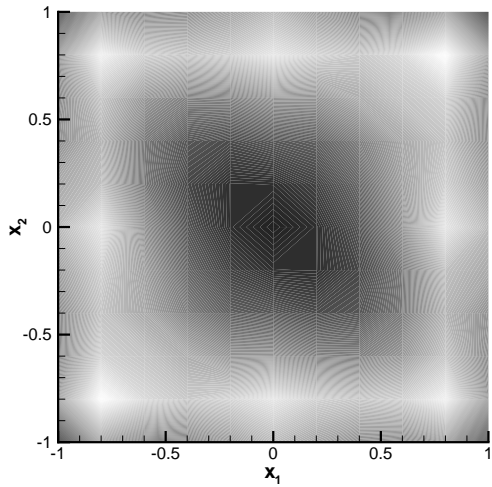


(a)

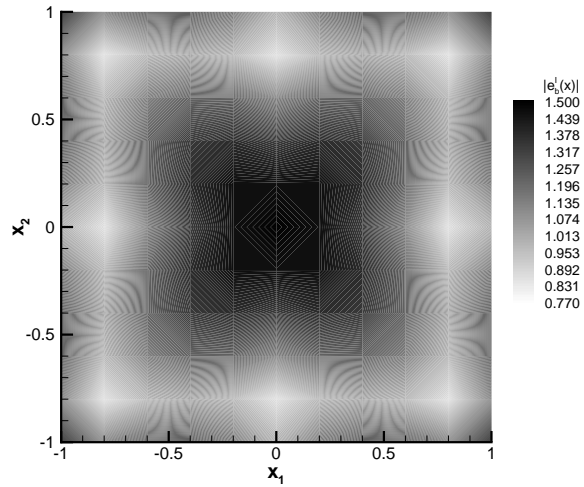


(b)

Figure 5: Comparison of absolute worst case errors for data set E in Example I: (a)worst of 100 polynomials, (b) data dependent bound, $|e_b^D(x)|$.

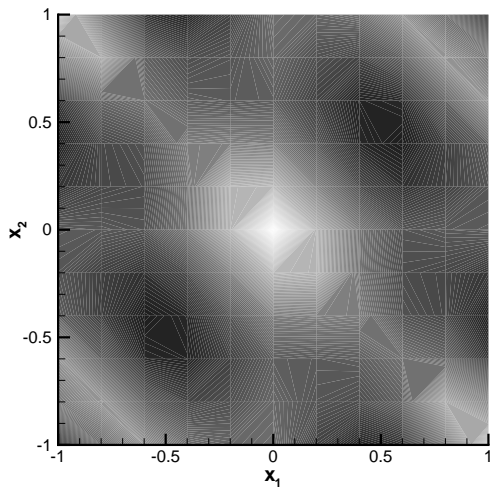


(a)

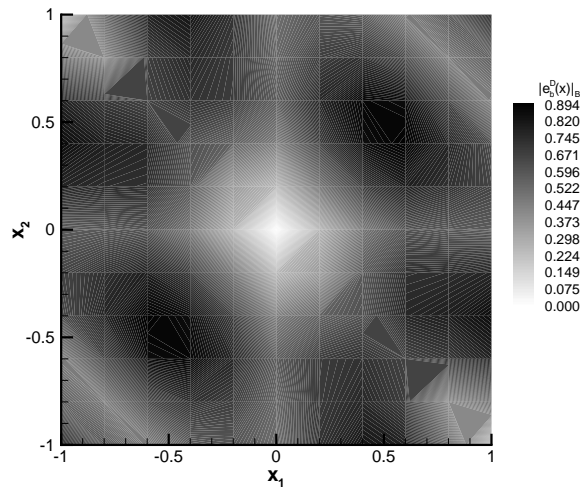


(b)

Figure 6: Comparison of absolute worst case errors for polynomials of data sets A through E in Example I: (a)worst of 500 polynomials, (b) data independent bound, $|e_b^I(x)|$.



(a)



(b)

Figure 7: Comparison of absolute worst case errors for data set B in Example II: (a)worst of 100 polynomials, (b) data dependent bound, $|e_b^D(x)|$.

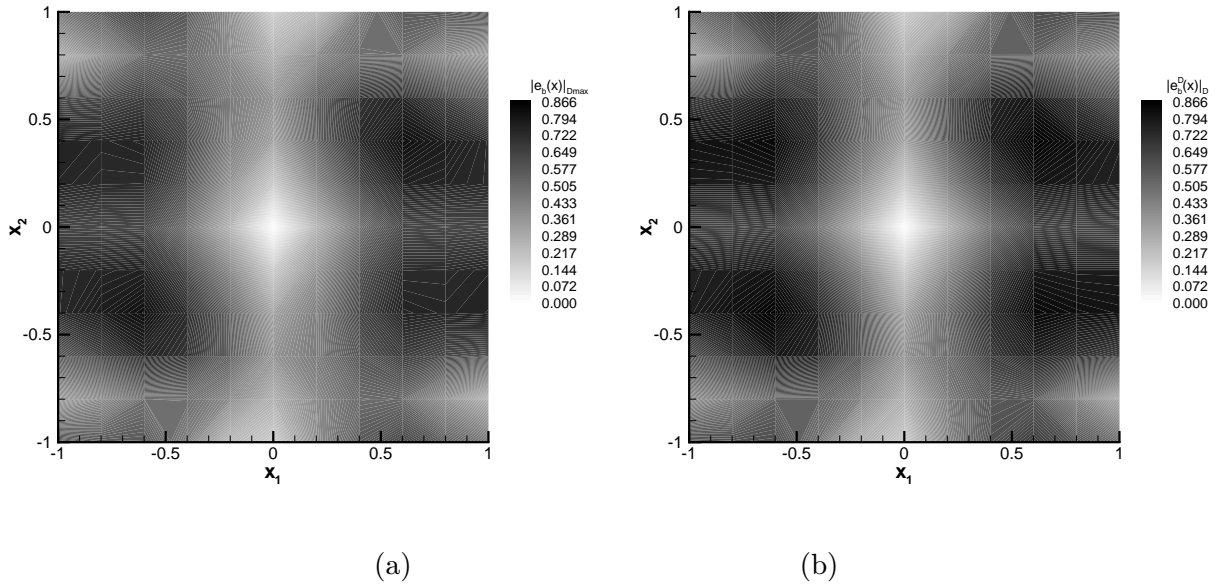


Figure 8: Comparison of absolute worst case errors for data set D in Example II: a) worst of 100 polynomials, b) data dependent bound, $|e_b^D(x)|$.

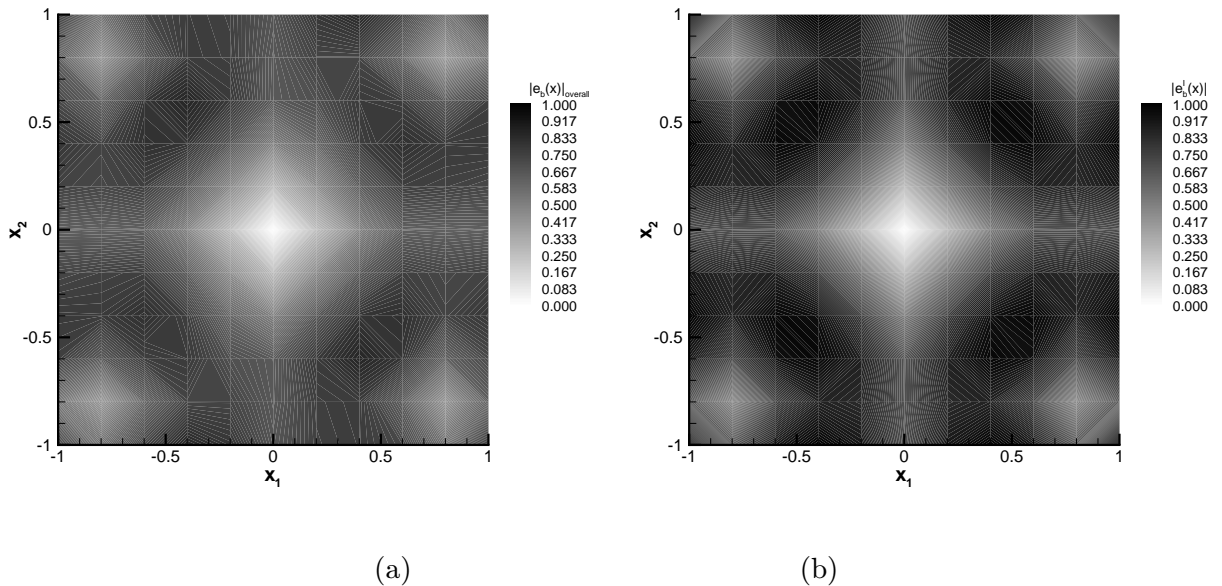


Figure 9: Comparison of absolute worst case errors for polynomials of data sets A through E in Example II: (a) worst of 500 polynomials, (b) data independent bound, $|e_b^I(x)|$.