Regularized random-sampling high dimensional model representation (RS-HDMR)

Genyuan Li and Herschel Rabitz*

Department of Chemistry, Princeton University, Princeton, NJ 08544, USA
E-mail: hrabitz@princeton.edu

Jishan Hu

Department of Mathematics, Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong

Zheng Chen and Yiguang Ju

Department of Mechanical and Aerospace Engineering, Princeton University, Princeton,
NJ 08544, USA

High Dimensional Model Representation (HDMR) is under active development as a set of quantitative model assessment and analysis tools for capturing high-dimensional input–output system behavior. HDMR is based on a hierarchy of component functions of increasing dimensions. The Random-Sampling High Dimensional Model Representation (RS-HDMR) is a practical approach to HDMR utilizing random sampling of the input variables. To reduce the sampling effort, the RS-HDMR component functions are approximated in terms of a suitable set of basis functions, for instance, orthonormal polynomials. Oscillation of the outcome from the resultant orthonormal polynomial expansion can occur producing interpolation error, especially on the input domain boundary, when the sample size is not large. To reduce this error, a regularization method is introduced. After regularization, the resultant RS-HDMR component functions are smoother and have better prediction accuracy, especially for small sample sizes (e.g., often few hundred). The ignition time of a homogeneous H_2/air combustion system within the range of initial temperature, $1000 < T_0 < 1500$ K, pressure, $0.1 < P < 100$ atm and equivalence ratio of H_2/O_2, $0.2 < R < 10$ is used for testing the regularized RS-HDMR.

KEY WORDS: high dimensional model representation (HDMR), orthonormal polynomials, regularization, smoothing, combustion, ignition

AMS(MOS) subject classifications: 33C50, 41A10, 57R12

*Corresponding author.

1. Introduction

High Dimensional Model Representation (HDMR) [1–22] is under development as a set of quantitative model assessment and analysis tools for capturing high-dimensional input–output system behavior. The HDMR techniques have been successfully applied in a variety of applications including semiconductor formulation [2], amino acid mutations of proteins [6], atmospheric chemistry [7], atmospheric solar radiation transport [8], molecular dynamics simulations [17], rate constant determination from concentration observations [18], and optimal control of molecular motion [19]. Recently, the capabilities of the HDMR technique were extended through the introduction of Random Sampling (RS)-HDMR, which is a practical procedure based on RS of the input variables. RS-HDMR is very efficient for treating high dimensional input–output mapping problems and has been successfully utilized in several modeling applications, e.g., atmospheric chemistry [9], environmental metal bioremediation [11], integrated exposure and dose studies [12], bio-kinetics modeling [21].

As the impact of the multiple input variables on the output can be independent and cooperative, HDMR expresses the output $f(\mathbf{x})$ as a finite hierarchical correlated function expansion in terms of the input variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$:

$$
\begin{aligned}
f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j) + \cdots \\
+ \sum_{1 \le i_1 < \cdots < i_l \le n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots \\
+ f_{12 \ldots n}(x_1, x_2, \ldots, x_n).
\end{aligned}
\tag{1}
$$

The HDMR component functions for normalized input variables ($0 \le x_i \le 1, i = 1, 2, \ldots, n$) are defined as follows [22]:

$$
f_0 = \int_{K^n} w(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x},
\tag{2}
$$

$$
f_i(x_i) = \int_{K^{n-1}} w_{\mathbf{x}^i|x_i}(\mathbf{x}^i) f(\mathbf{x}) \mathrm{d}\mathbf{x}^i - f_0,
\tag{3}
$$

$$
f_{ij}(x_i, x_j) = \int_{K^{n-2}} w_{\mathbf{x}^{ij}|x_i, x_j}(\mathbf{x}^{ij}) f(\mathbf{x}) \mathrm{d}\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0,
\tag{4}
$$

$$
\ldots,
$$

where $\mathbf{x}^i, \mathbf{x}^{ij}$ are $(x_1, x_2, \ldots, x_n)$ without elements x_i; x_i, x_j, respectively. $w(\mathbf{x})$ is the probability density function (pdf) for $\mathbf{x}$ in the n-dimensional hypercube K^n satisfying the conditions

$$w(\mathbf{x}) \geq 0, \quad \mathbf{x} \in K^n,$$

$$\int_{K^n} w(\mathbf{x})\mathrm{d}\mathbf{x} = 1. \tag{5}$$

$w_{\mathbf{x}^i|x_i}(\mathbf{x}^i)$, $w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij})$ denote the conditional pdf's of $\mathbf{x}$ for a fixed value of x_i and x_i, x_j, i.e.,

$$w_{\mathbf{x}^i|x_i}(\mathbf{x}^i) = w(x_i, \mathbf{x}^i)/w_i(x_i), \tag{6}$$

$$w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij}) = w(x_i, x_j, \mathbf{x}^{ij})/w_{ij}(x_i, x_j), \tag{7}$$

$$w_i(x_i) = \int_{K^{n-1}} w(x_i, \mathbf{x}^i)\mathrm{d}\mathbf{x}^i, \tag{8}$$

$$w_{ij}(x_i, x_j) = \int_{K^{n-2}} w(x_i, x_j, \mathbf{x}^{ij})\mathrm{d}\mathbf{x}^{ij}. \tag{9}$$

Different, but formally equivalent HDMR expansions, all of the same form as equation (1) have been constructed [1–6]. To reduce the sampling effort, the RS-HDMR component functions may be approximated by optimal weighted orthonormal polynomials $\{\varphi\}$ as [22]

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i), \tag{10}$$

$$f_{ij}(x_i, x_j) \approx \sum_{r=1}^{k} \left[\alpha_r^{(ij)i} \varphi_r^i(x_i) + \alpha_r^{(ij)j} \varphi_r^j(x_j) \right]$$

$$+ \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i) \varphi_q^j(x_j), \tag{11}$$

$$f_{ijk}(x_i, x_j, x_k) \approx \sum_{r=1}^{k} \left[\alpha_r^{(ijk)i} \varphi_r^i(x_i) + \alpha_r^{(ijk)j} \varphi_r^j(x_j) + \alpha_r^{(ijk)k} \varphi_r^k(x_k) \right]$$

$$+ \sum_{p=1}^{l} \sum_{q=1}^{l'} \left[\beta_{pq}^{(ijk)ij} \varphi_p^i(x_i) \varphi_q^j(x_j) \right.$$

$$+ \left. \beta_{pq}^{(ijk)ik} \varphi_p^i(x_i) \varphi_q^k(x_k) + \beta_{pq}^{(ijk)jk} \varphi_p^j(x_j) \varphi_q^k(x_k) \right]$$

$$+ \sum_{p=1}^{m} \sum_{q=1}^{m'} \sum_{r=1}^{m''} \gamma_{pqr}^{(0)ijk} \varphi_p^i(x_i) \varphi_q^j(x_j) \varphi_r^k(x_k), \tag{12}$$

$$\cdots,$$

where k, l, l', m, m', m'' are integers. In many cases the number of basis functions is no larger than three. $\{\varphi\}$ are referred to as *optimal weighted orthonormal*

polynomials tailored to a particular set of samples, and they are defined as follows:

$$\varphi_1^i(x_i) = a_1 x_i + a_0, \tag{13}$$
$$\varphi_2^i(x_i) = b_2 x_i^2 + b_1 x_i + b_0, \tag{14}$$
$$\varphi_3^i(x_i) = c_3 x_i^3 + c_2 x_i^2 + c_1 x_i + c_0, \tag{15}$$
$$\cdots,$$

where the constant coefficients $a_0, a_1, b_0, \ldots, c_3$ are chosen in such a way that for a given set of random samples the orthonormality of $\{\varphi\}$ is forced to be satisfied:

$$\int_0^1 w_i(x_i)\varphi_r^i(x_i)\mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^N \varphi_r^i(x_i^{(s)}) = 0, \quad r = 1, 2, \ldots, \tag{16}$$

$$\int_0^1 w_i(x_i)[\varphi_r^i(x_i)]^2\mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^N [\varphi_r^i(x_i^{(s)})]^2 = 1, \quad r = 1, 2, \ldots, \tag{17}$$

$$\int_0^1 w_i(x_i)\varphi_p^i(x_i)\varphi_q^i(x_i)\mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^N \varphi_p^i(x_i^{(s)})\varphi_q^i(x_i^{(s)}) = 0, \quad p \neq q, \tag{18}$$

where $x_i^{(s)}$ is the value of x_i of the sth sample, and N is the total number of samples.

Using the formulas in equations (10)–(12), the third-order RS-HDMR expansion of an n-variate function $f(\mathbf{x})$ can be expressed as

$$
\begin{aligned}
f(\mathbf{x}) \approx f_0 &+ \sum_{i=1}^n \sum_{r=1}^k \left[\alpha_r^{(0)i} + \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_r^{(ij)i} + \sum_{\substack{j<k=1 \\ j,k \neq i}}^n \alpha_r^{(ijk)i} \right] \varphi_r^i(x_i) \\
&+ \sum_{1 \leq i < j \leq n} \sum_{p=1}^l \sum_{q=1}^{l'} \left[\beta_{pq}^{(0)ij} + \sum_{\substack{k=1 \\ k \neq i,j}}^n \beta_{pq}^{(ijk)ij} \right] \varphi_p^i(x_i)\varphi_q^j(x_j) \\
&+ \sum_{1 \leq i < j < k \leq n} \sum_{p=1}^m \sum_{q=1}^{m'} \sum_{r=1}^{m''} \gamma_{pqr}^{(0)ijk} \varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k).
\end{aligned} \tag{19}
$$

The expansion coefficients $\{\alpha, \beta, \gamma\}$ may be determined by stepwise least squares regression from zeroth-order to higher orders [22]. First, we set

$$f_0 = \int_{K^n} w(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x} \approx \frac{1}{N} \sum_{s=1}^{N} f(\mathbf{x}^{(s)}), \tag{20}$$

$$y_1(\mathbf{x}) = f(\mathbf{x}) - f_0, \tag{21}$$

$$y_2(\mathbf{x}) = y_1(\mathbf{x}) - \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i), \tag{22}$$

$$y_3(\mathbf{x}) = y_2(\mathbf{x}) - \sum_{1 \leq i < j \leq n} \left\{ \sum_{r=1}^{k} [\alpha_r^{(ij)i} \varphi_r^i(x_i) + \alpha_r^{(ij)j} \varphi_r^j(x_j)] \right.$$
$$\left. + \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i) \varphi_q^j(x_j) \right\}, \tag{23}$$

$$\cdots$$

and solve the following equations by stepwise least squares regression

$$y_1(\mathbf{x}^{(s)}) = \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i^{(s)}) + \varepsilon_s, \quad s = 1, 2, \ldots, N, \tag{24}$$

$$y_2(\mathbf{x}^{(s)}) = \sum_{1 \leq i < j \leq n} \left\{ \sum_{r=1}^{k} [\alpha_r^{(ij)i} \varphi_r^i(x_i^{(s)}) + \alpha_r^{(ij)j} \varphi_r^j(x_j^{(s)})] \right.$$
$$\left. + \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \right\} + \varepsilon_s, \quad s = 1, 2, \ldots, N, \tag{25}$$

$$y_3(\mathbf{x}^{(s)}) = \sum_{1 \leq i < j < k \leq n} \left\{ \sum_{r=1}^{k} [\alpha_r^{(ijk)i} \varphi_r^i(x_i^{(s)}) + \alpha_r^{(ijk)j} \varphi_r^j(x_j^{(s)}) + \alpha_r^{(ijk)k} \varphi_r^k(x_k)^{(s)}] \right.$$
$$+ \sum_{p=1}^{l} \sum_{q=1}^{l'} [\beta_{pq}^{(ijk)ij} \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})$$
$$+ \beta_{pq}^{(ijk)ik} \varphi_p^i(x_i^{(s)}) \varphi_q^k(x_k^{(s)}) + \beta_{pq}^{(ijk)jk} \varphi_p^j(x_j^{(s)}) \varphi_q^k(x_k^{(s)})]$$
$$\left. + \sum_{p=1}^{m} \sum_{q=1}^{m'} \sum_{r=1}^{m''} \gamma_{pqr}^{(0)ijk} \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \varphi_r^k(x_k^{(s)}) \right\}$$
$$+ \varepsilon_s, \quad s = 1, 2, \ldots, N \tag{26}$$
$$\cdots,$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is a Gaussian white noise vector, and σ^2 is unknown.

Since the RS-HDMR component functions are approximated by an orthonormal polynomial expansion with the expansion coefficients determined by least squares regression, insufficient data and significant errors in the data prevent a unique solution for the expansion coefficients. Moreover, high order polynomial fitting can produce oscillations in the evaluation of the expansion, especially on the input domain boundary when the sample size N is not large. Regularization of the RS-HDMR component functions can reduce the resultant prediction error. In regularization, both the total squared residuals (the difference between the model value and the prediction of RS-HDMR) and the total squared second-order derivatives of the RS-HDMR expansions for all the data are minimized in least squares regression. A regularization parameter is used to trade off fidelity to the data (i.e., a small sum of the squared residuals) against smoothness (i.e., low values of the total squared second-order derivatives). Different methods have been proposed for choosing the regularization parameter. Among these methods, generalized cross-validation (GCV) is commonly used [23, 24]. However, the GCV method needs to treat a matrix with dimension of N. For large N, the computational effort is significant. In the calculation of chemical kinetics models with RS-HDMR, it is common that $N \sim 10^3$ or more, and it is difficult to use the GCV method.

In this paper, an efficient method to choose the regularization parameter for the RS-HDMR component functions is presented. The optimal value of the regularization parameter divides the basis functions $\{\varphi\}$ of RS-HDMR into two groups, and the regularization parameter may be estimated from this division without significant computational effort.

The paper is organized as follows. Section 2 presents the methodology of regularized RS-HDMR and the determination of the regularization parameter. Section 3 presents illustrations of the regularized RS-HDMR method for the estimation of the ignition time of a homogeneous H_2/air combustion system. Finally, section 4 contains conclusions.

2.   Regularized RS-HDMR

Consider the determination of the first-order RS-HDMR component functions. Suppose $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$ are N points in the input domain in which the output observations $y(\mathbf{x}^{(s)}) = f(\mathbf{x}^{(s)})$ $(s = 1, 2, \ldots, N)$ are taken, and the first-order approximation for $y_1(\mathbf{x}^{(s)}) = f(\mathbf{x}^{(s)}) - f_0$ satisfies

$$y_1(\mathbf{x}^{(s)}) = \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i^{(s)}) + \varepsilon_s. \tag{27}$$

Least squares regression with regularization is performed by minimizing the functional $\mathcal{F}$

$$\mathcal{F} = \frac{1}{N} \sum_{s=1}^{N} \left[y_1(\mathbf{x}^{(s)}) - \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i^{(s)}) \right]^2 \\ + \lambda \sum_{i=1}^{n} \sum_{r=1}^{k} \lambda_r^i \frac{1}{N} \sum_{s=1}^{N} \left[\alpha_r^{(0)i} \frac{\mathrm{d}^2 \varphi_r^i(x_i^{(s)})}{\mathrm{d}x_i^2} \right]^2, \tag{28}$$

where $\lambda > 0$ is the regularization parameter used to trade off fidelity to the data (i.e., the first term in equation (28) being the sum of squared residuals) against smoothness (i.e., the second term in equation (28) being of the squared second-order derivatives), and $\lambda_r^i > 0$ is the relative weight for regularization of $\varphi_r^i(x_i)$ defined as

$$\lambda_r^i = \frac{\displaystyle\sum_{s=1}^{N} \left[\frac{\mathrm{d}^2 \varphi_r^i(x_i^{(s)})}{\mathrm{d}x_i^2} \right]^2}{\displaystyle\sum_{j=1}^{n} \sum_{t=1}^{k} \sum_{s=1}^{N} \left[\frac{\mathrm{d}^2 \varphi_t^j(x_j^{(s)})}{\mathrm{d}x_j^2} \right]^2}. \tag{29}$$

Thus, only one free parameter λ is used to determine the amount by which the data are smoothed to produce regularized RS-HDMR functions.

Cross-validation may be used to choose λ. The basic principle of cross-validation is to leave the data points out one at a time and to choose the value of λ for which the missing data points are best predicted by the remainder of the data. The GCV method needs to treat a matrix of dimension N, and for large N, the computational effort is significant even though some modifications of cross-validation have been proposed [23, 24].

Here we present an easy approach to determine λ in regularized RS-HDMR. In order to develop the procedure, first suppose that λ is already chosen. The least squares regression equations are obtained by differentiating $\mathcal{F}$ with respect to $\alpha_t^{(0)j}$ ($j = 1, 2, \ldots, n; t = 1, 2, \ldots, k$) and setting the derivatives to zero to produce

$$\frac{1}{N} \sum_{s=1}^{N} y_1^{(s)} \varphi_t^j(x_j^{(s)}) = \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \frac{1}{N} \sum_{s=1}^{N} \varphi_r^i(x_i^{(s)}) \varphi_t^j(x_j^{(s)}) \\ - \lambda \lambda_t^j \alpha_t^{(0)j} \frac{1}{N} \sum_{s=1}^{N} \left[\frac{\mathrm{d}^2 \varphi_t^j(x_j^{(s)})}{\mathrm{d}x_j^2} \right]^2. \tag{30}$$

Using the condition given in equation (16), the above equation becomes

$$\frac{1}{N} \sum_{s=1}^{N} f(\mathbf{x}^{(s)}) \varphi_t^j(x_j^{(s)}) = \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \frac{1}{N} \sum_{s=1}^{N} \varphi_r^i(x_i^{(s)}) \varphi_t^j(x_i^{(s)})$$

$$- \lambda \lambda_t^j \alpha_t^{(0)j} \frac{1}{N} \sum_{s=1}^{N} \left[\frac{\mathrm{d}^2 \varphi_t^j(x_j^{(s)})}{\mathrm{d}x_j^2} \right]^2 . \tag{31}$$

Equation (31) can be represented in matrix form

$$[A - \lambda B]\alpha = \mathbf{b}, \tag{32}$$

where A and B are nk-dimensional symmetric and diagonal matrices, respectively, with elements

$$A_{(ir)(jt)} = \frac{1}{N} \sum_{s=1}^{N} \varphi_r^i(x_i^{(s)}) \varphi_t^j(x_j^{(s)}), \tag{33}$$

$$B_{(jt)(jt)} = \lambda_t^j \frac{1}{N} \sum_{s=1}^{N} \left[\frac{\mathrm{d}^2 \varphi_t^j(x_j^{(s)})}{\mathrm{d}x_j^2} \right]^2 \tag{34}$$

and α and $\mathbf{b}$ are nk-dimensional vectors with

$$\alpha_{(jt)} = \alpha_t^{(0)j}, \tag{35}$$

$$b_{(jt)} = \frac{1}{N} \sum_{s=1}^{N} f(\mathbf{x}^{(s)}) \varphi_t^j(x_j^{(s)}). \tag{36}$$

The solution α of equation (32) depends on the value of the regularization parameter λ. The basic concept behind the proposed method is that *there is no significant difference between the results of least squares regression for the N and $N-1$ data points when N is sufficiently large.* Therefore, equation (32) can be used as an approximation of the corresponding equation obtained from $N-1$ data points.

The second-order regularized RS-HDMR component functions can be treated similarly. For

$$y_2(\mathbf{x}^{(s)}) = \sum_{1 \le i < j \le n} \left\{ \sum_{r=1}^{k} [\alpha_r^{(ij)i} \varphi_r^i(x_i^{(s)}) + \alpha_r^{(ij)j} \varphi_r^j(x_j^{(s)})] \right.$$

$$\left. + \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \right\} + \varepsilon_s \tag{37}$$

the functional subjected to minimization for determining $\alpha_r^{(ij)i}$, $\alpha_r^{(ij)j}$, $\beta_{pq}^{(0)ij}$ with regularization is

$$
\mathcal{F} = \frac{1}{N} \sum_{s=1}^{N} \left\{ y_2(\mathbf{x}^{(s)}) - \sum_{1 \leq i < j \leq n} \left[\sum_{r=1}^{k} (\alpha_r^{(ij)i} \varphi_r^i(x_i) + \alpha_r^{(ij)j} \varphi_r^j(x_j)) \right. \right.
$$

$$
\left. \left. + \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i) \varphi_q^j(x_j) \right] \right\}^2
$$

$$
+ \sum_{1 \leq i < j \leq n} \frac{1}{N} \sum_{s=1}^{N} \left\{ \sum_{r=1}^{k} \left[\lambda_r^{(ij)i} \left(\alpha_r^{(ij)i} \frac{d^2 \varphi_r^i(x_i^{(s)})}{dx_i^2} \right)^2 \right. \right.
$$

$$
\left. + \lambda_r^{(ij)j} \left(\alpha_r^{(ij)j} \frac{d^2 \varphi_r^j(x_j^{(s)})}{dx_j^2} \right)^2 \right] + \sum_{p=1}^{k} \sum_{q=1}^{k} \lambda_{pq}^{ij} \left[\left(\beta_{pq}^{(0)ij} \frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_i^2} \right)^2 \right.
$$

$$
\left. \left. + \left(\beta_{pq}^{(0)ij} \frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_i \partial x_j} \right)^2 + \left(\beta_{pq}^{(0)ij} \frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_j^2} \right)^2 \right] \right\}, \tag{38}
$$

where

$$
\lambda_r^{(ij)i} = \Lambda_r^{(ij)i}/\Lambda, \tag{39}
$$

$$
\lambda_r^{(ij)j} = \Lambda_r^{(ij)j}/\Lambda, \tag{40}
$$

$$
\lambda_{pq}^{ij} = \Lambda_{pq}^{ij}/\Lambda, \tag{41}
$$

$$
\Lambda_r^{(ij)i} = \frac{1}{N} \sum_{s=1}^{N} \left(\frac{d^2 \varphi_r^i(x_i^{(s)})}{dx_i^2} \right)^2, \tag{42}
$$

$$
\Lambda_r^{(ij)j} = \frac{1}{N} \sum_{s=1}^{N} \left(\frac{d^2 \varphi_r^j(x_j^{(s)})}{dx_j^2} \right)^2, \tag{43}
$$

$$
\Lambda_{pq}^{ij} = \frac{1}{N} \sum_{s=1}^{N} \left[\left(\frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_i^2} \right)^2 + \left(\frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_i \partial x_j} \right)^2 \right.
$$

$$
\left. + \left(\frac{\partial^2 \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)})}{\partial x_j^2} \right)^2 \right], \tag{44}
$$

$$
\Lambda = \sum_{1 \leq i < j \leq n} \left[\sum_{r=1}^{k} (\Lambda_r^{(ij)i} + \Lambda_r^{(ij)j}) + \sum_{p=1}^{k} \sum_{q=1}^{k} \Lambda_{pq}^{ij} \right]. \tag{45}
$$

The equations obtained from differentiating $\mathcal{F}$ in equation (38) with respect to its parameters $\{\alpha, \beta\}$ and setting the derivatives to zero also can be represented in matrix form as in equation (32) when $\{\alpha, \beta\}$ are properly arranged.

The regularization parameters λ for the first- and second-order RS-HDMR component functions are chosen in such a way that the errors of cross-validation

$$
\text{er1} = \frac{1}{N} \sum_{s=1}^{N} e_s^2 = \frac{1}{N} \sum_{s=1}^{N} \left[y_1(\mathbf{x}^{(s)}) - \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i}(s) \varphi_r^i(x_i^{(s)}) \right]^2, \tag{46}
$$

$$
\text{er2} = \frac{1}{N} \sum_{s=1}^{N} e_s^2 = \frac{1}{N} \sum_{s=1}^{N} \left\{ y_2(\mathbf{x}^{(s)}) - \sum_{1 \le i < j \le n} \left[\sum_{r=1}^{k} (\alpha_r^{(ij)i}(s) \varphi_r^i(x_i^{(s)}) \right. \right.
$$
$$
\left. \left. + \alpha_r^{(ij)j}(s) \varphi_r^j(x_j^{(s)})) + \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{(0)ij}(s) \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \right] \right\}^2, \tag{47}
$$

are minimized. Here the parameters $\alpha(s), \beta(s)$ should be different for distinct values of s and determined from the remaining $N - 1$ data points, but they are approximated by α, β obtained from the total set of N data points.

A further approximation is needed to determine er1 and er2. Since $\{\varphi\}$ are optimal orthonormal polynomials, the orthonormality given by equations (16)–(18) is forced to be satisfied. For $\varphi_p^i(x_i)\varphi_q^j(x_j)$ with different variables the orthonormality may not be strictly satisfied, but the diagonal elements of A are very close to unity, and off-diagonal elements of A are small numbers close to zero, i.e.,

$$
A \approx I. \tag{48}
$$

Then an approximate solution for λ in equation (32) can be obtained by setting $A = I$ which gives

$$
\alpha = (I - \lambda B)^{-1} \mathbf{b}. \tag{49}
$$

As B is diagonal, the inverse $(I - \lambda B)^{-1}$ can be readily obtained. The error er1 can then be represented as

$$
\begin{aligned}
\text{er1} &= \frac{1}{N} \sum_{s=1}^{N} e_s^2 = \frac{1}{N} \sum_{s=1}^{N} \left[y_1(\mathbf{x}^{(s)}) - \sum_{i=1}^{n} \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i^{(s)}) \right]^2 \\
&= \sigma^2 - 2\mathbf{b}^T \alpha + \alpha^T A \alpha \\
&\approx \sigma^2 - 2\mathbf{b}^T (I - \lambda B)^{-1} \mathbf{b} + \mathbf{b}^T [(I - \lambda B)^{-1}]^2 \mathbf{b}
\end{aligned}
$$

$$= \sigma^2 - 2\mathbf{b}^T \begin{bmatrix} \frac{1}{1-\lambda B_{11}} & & \\ & \ddots & \\ & & \frac{1}{1-\lambda B_{(nk)(nk)}} \end{bmatrix} \mathbf{b}$$

$$+ \mathbf{b}^T \begin{bmatrix} \left(\frac{1}{1-\lambda B_{11}}\right)^2 & & \\ & \ddots & \\ & & \left(\frac{1}{1-\lambda B_{(nk)(nk)}}\right)^2 \end{bmatrix} \mathbf{b}$$

$$= \sigma^2 - 2\sum_{i=1}^{nk} \frac{b_i^2}{1-\lambda B_{ii}} + \sum_{i=1}^{nk} \frac{b_i^2}{(1-\lambda B_{ii})^2}, \tag{50}$$

where σ^2 is the output total variance, i.e.,

$$\sigma^2 \approx \frac{1}{N}\sum_{s=1}^{N}[y_1(\mathbf{x}^{(s)})]^2 = \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - f_0]^2. \tag{51}$$

The value of λ is obtained by setting the derivative of er1 with respect to λ to zero, which gives

$$\sum_{i=1}^{nk} \frac{b_i^2 B_{ii}}{(1-\lambda B_{ii})^2} - \sum_{i=1}^{nk} \frac{b_i^2 B_{ii}}{(1-\lambda B_{ii})^3} = 0, \tag{52}$$

where B_{ii} is the nonzero ith diagonal element of B. Simplification of equation (52) yields

$$\lambda \sum_{i=1}^{nk} \frac{b_i^2 B_{ii}^2}{(1-\lambda B_{ii})^3} = 0. \tag{53}$$

The first estimate for λ in equation (53) is

$$\lambda = 0, \tag{54}$$

which corresponds to no regularization. Other estimates are obtained by solving the algebraic equation

$$\sum_{i=1}^{nk} \frac{b_i^2 B_{ii}^2}{(1-\lambda B_{ii})^3} = 0. \tag{55}$$

Since the numerators are all positive, and the denominators are cubic functions, the solutions for λ should be located between the roots of the denominators, i.e., $1/B_{ii}$, and can be readily determined. The cross-validation may have only one minimum. The extra multiple solutions given in equation (55) are produced by

the approximations above. However, if the minimum of cross-validation is located between the roots of equation (55), only a few values of λ need to be tested, and the direct calculation of cross-validation for these points is then not expensive. As shown in the example below, the regularization parameter λ can be easily estimated from $1/B_{ii}$. For er2 the treatment is the same.

3.   Application to ignition time of an H_2/air model

The regularized RS-HDMR technique is applied here for estimation of the ignition time of an H_2/air combustion model with eight species (H_2, O_2, H_2O, H, O, OH, HO_2, and H_2O_2) and 19 reactions. The detail of the H_2/air combustion model can be found in [25]. The initial temperature, $1000 < T_0 < 1500$ K, pressure, $0.1 < P < 100$ atm, and H_2/O_2 equivalence ratio, $0.2 < R < 10.0$, are chosen as three inputs, and the homogeneous ignition time t_{ign} (defined as the time lapse needed to get an increase of 400 K from the initial temperature) is the output. 4699 random data points over the three inputs were sampled with a uniform distribution within the above domain, and their corresponding output values were obtained from the model.

The relationship of ignition time t_{ign} to T_0, P and R is given in figure 1. The figure shows that at low pressure the ignition time is not a monodrome function. Therefore, the data set was separated into three subsets according to $P < 1$ atm (1579 points), $1 \leq P < 10$ atm (1538 points) and $P \geq 10$ atm (1582 points). The second-order RS-HDMR expansions were constructed for the entire data set as well as the three subsets.

3.1.   Accuracy of the second-order RS-HDMR

RS-HDMR can be used as a Fully Equivalent Operational Model (FEOM) [4, 7, 8] to replace the original time-consuming H_2/air model for approximate estimation of ignition time. For the entire data set and its three subsets, different sample sizes ($N = 100, 300, 500, 1000, ...$ referred to as "used data") were employed to construct the second-order RS-HDMR expansion whose component functions are approximated by third-order optimal orthonormal polynomials. The remaining $M - N$ data points (referred to as "test data"), where M is the total number of samples, were used to test the interpolation accuracy of the RS-HDMR. The results are given in tables 1–4. The accuracy of different order RS-HDMR expansions is represented by the data portion whose relative errors are not larger than a given value (5, 10, and 20%).

The results in tables 1–4 show that the accuracy is much improved when the data are separated into three subsets. The used data and test data have almost the same accuracy for all sample sizes. Most data ($>90\%$) have relative errors less than 5%. This implies that a few hundred samples are sufficient to construct
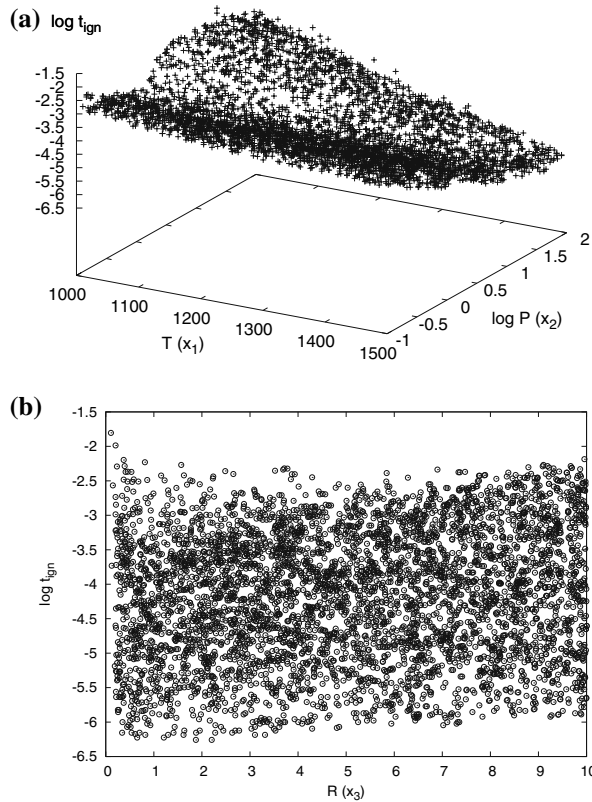
Figure 1. The dependence of ignition time on the inputs: initial temperature T_0, pressure P, and H_2/O_2 equivalence ratio R. Their designations in terms of x_1, x_2, and x_3 are indicated.

the second-order RS-HDMR expansion which can satisfactorily interpolate all points over the input domains.

Figure 2 compares the model values with those estimated by the second-order RS-HDMR approximation of the output obtained from the entire data set and its three subsets.

3.2. Second-order regularized RS-HDMR

When the sample size is small, the accuracy of the second-order standard RS-HDMR is not fully satisfactory, and regularization can improve the accuracy. The last subset of data ($P \geqslant 10$ atm) was employed as a test example. The first 100 samples were used to construct the second-order regularized RS-HDMR expansion. For the first-order RS-HDMR component functions, the matrix B in equation (32) is

Table 1
The relative errors of different order RS-HDMR expansions obtained from the entire data set.

| Sample size N (M − N) | Relative error (%) | Data portion | | | |
| | | Used data | | Test data | |
		1st order	2nd order	1st order	2nd order
100	5	0.300	0.880	0.349	0.767
(4599)	10	0.600	0.980	0.652	0.931
	20	0.870	1.000	0.898	0.993
300	5	0.367	0.897	0.386	0.879
(4399)	10	0.653	0.973	0.673	0.980
	20	0.893	0.997	0.889	0.999
500	5	0.400	0.884	0.390	0.876
(4199)	10	0.694	0.980	0.679	0.984
	20	0.898	0.998	0.891	0.999
1000	5	0.397	0.888	0.389	0.877
(3699)	10	0.695	0.985	0.677	0.984
	20	0.900	0.999	0.890	0.999
2000	5	0.398	0.897	0.381	0.888
(2699)	10	0.681	0.985	0.671	0.986
	20	0.892	0.999	0.893	1.000
3000	5	0.397	0.893	0.383	0.897
(1699)	10	0.675	0.986	0.680	0.991
	20	0.892	0.999	0.890	0.999
4000	5	0.388	0.896	0.402	0.906
(699)	10	0.672	0.987	0.682	0.990
	20	0.888	0.999	0.908	1.000

$$B = \begin{bmatrix} 0 & & & & & & & \\ & 19.5 & & & & & & \\ & & 3555 & & & & & \\ & & & 0 & & & & \\ & & & & 16.6 & & & \\ & & & & & 2443 & & \\ & & & & & & 0 & \\ & & & & & & & 30.8 \\ & & & & & & & & 2427 \end{bmatrix}. \tag{56}$$

The algebraic equation in equation (55) for the determination of λ is

$$f(x) = \frac{5.08}{(1 - 19.5x)^3} + \frac{998}{(1 - 3555x)^3} + \frac{0.011}{(1 - 16.6x)^3} + \frac{1609}{(1 - 2443x)^3}$$
$$+ \frac{9.13}{(1 - 30.8x)^3} + \frac{31351}{(1 - 2427x)^3} = 0. \tag{57}$$

Table 2
The relative errors of different order RS-HDMR expansions obtained from the data subset ($0.1 \leqslant P < 1$ atm).

| Sample size N ($M - N$) | Relative error (%) | Data portion | | | |
| | | Used data | | Test data | |
		1st order	2nd order	1st order	2nd order
100	5	0.880	1.000	0.853	0.968
(1479)	10	0.990	1.000	0.983	0.995
	20	1.000	1.000	0.999	0.999
300	5	0.857	0.993	0.873	0.979
(1279)	10	0.993	1.000	0.986	1.000
	20	0.997	1.000	1.000	1.000
500	5	0.862	0.996	0.874	0.986
(1079)	10	0.986	0.998	0.984	1.000
	20	0.996	1.000	1.000	1.000
1000	5	0.873	0.992	0.871	0.986
(579)	10	0.983	0.999	0.988	1.000
	20	0.998	1.000	1.000	1.000

Table 3
The relative errors of different order RS-HDMR expansions obtained from the data subset ($1 \leq P < 10$ atm).

| Sample size N ($M - N$) | Relative error (%) | Data portion | | | |
| | | Used data | | Test data | |
		1st order	2nd order	1st order	2nd order
100	5	0.750	0.960	0.736	0.919
(1438)	10	0.880	1.000	0.873	0.978
	20	0.990	1.000	0.951	0.998
300	5	0.737	0.970	0.738	0.943
(1238)	10	0.910	1.000	0.866	0.986
	20	0.983	1.000	0.946	0.998
500	5	0.722	0.970	0.724	0.928
(1038)	10	0.896	0.994	0.876	0.989
	20	0.976	1.000	0.950	1.000
1000	5	0.709	0.949	0.695	0.946
(538)	10	0.884	0.987	0.877	0.994
	20	0.969	1.000	0.952	1.000

Table 4
The relative errors of different order RS-HDMR expansions obtained from the data subset
$(10 \leq P < 100 \, \text{atm})$.

| Sample size N $(M - N)$ | Relative error (%) | Data portion | | | |
| | | Used data | | Test data | |
		1st order	2nd order	1st order	2nd order
100	5	0.880	0.980	0.841	0.907
(1482)	10	0.980	1.000	0.982	0.970
	20	0.990	1.000	1.000	0.996
300	5	0.897	0.997	0.862	0.967
(1282)	10	0.983	0.997	0.982	0.994
	20	0.997	1.000	1.000	1.000
500	5	0.868	0.988	0.879	0.977
(1082)	10	0.978	0.996	0.986	0.997
	20	0.998	0.998	1.000	1.000
1000	5	0.887	0.992	0.881	0.990
(582)	10	0.976	0.996	0.985	0.996
	20	0.999	0.999	1.000	1.000

The roots of the denominators, $1/B_{ii}$, in equation (57) are given below in increasing order:

$$0.000281 \quad 0.000409 \quad 0.000412 \quad 0.0325 \quad 0.0513 \quad 0.0602$$

The roots of equation (57) are located between the adjacent roots of the denominators, which correspond to the approximate values of the regularization parameter λ giving local minima of the first-order regularized RS-HDMR approximation. Figure 3 gives the profile of $f(x)(\times 10^6)$ from equation (57), where some of the roots can be observed. As the mesh of x used to plot $f(x)$ is of fixed size, some rapid jumps across the x axis cannot be observed in the figure.

To demonstrate that the best value of λ, which gives the minimum cross-validation error, is located between the roots of equation (57), the cross-validation for different values of the regularization parameter λ was performed for a set of 100 data points. The result for the first-order regularized RS-HDMR is given in figure 4. The minimum is located at $\lambda \sim 0.0006$, i.e., between the roots 0.000412 and 0.0325 of the denominators in equation (57) where a sharp variation occurs. Importantly, there is a stable domain around 0.0006 where the error does not change much; if the choice 0.000412 $(1/B_{99})$ instead of 0.0006 is made for λ, the error is still acceptable.

The value of $\lambda \sim 0.0006$ means that the polynomials, whose $1/B_{ii}$ values are smaller than λ, may produce oscillations and will be smoothed out. In contrast, other polynomials, whose $1/B_{ii}$ values are larger than λ, do not need heavy
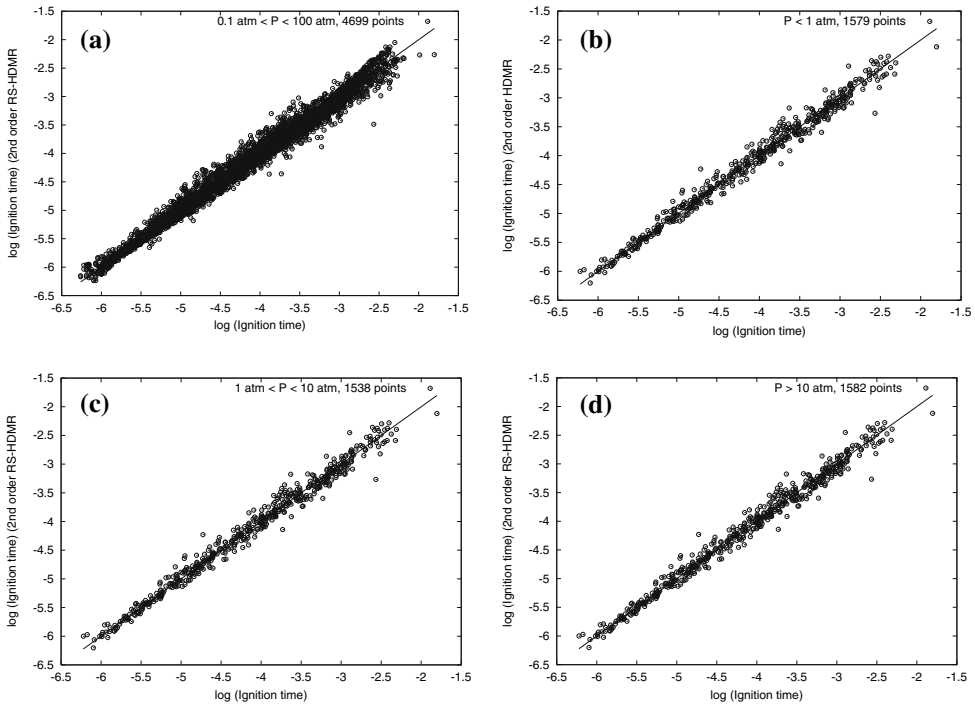
Figure 2. The comparison of the output model values and their second-order RS-HDMR approximation whose component functions are approximated by third-order optimal orthonormal polynomials.

regularization. It was found that these two categories of behavior were associated with third-order and second-order polynomials, respectively (linear functions naturally need no regularization). This behavior is reasonable because high order polynomials will more likely cause oscillation.

Let us consider this separation generally, and suppose that $\mathbf{x}^\alpha = (x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_n^{\alpha_n})$ represents a monomial in $x_1, x_2, \ldots, x_n$ where all of the exponents $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ are non-negative integers. The *total degree* of the monomial is defined as the sum $|\alpha| = \alpha_1 + \cdots + \alpha_n$. A polynomial $f(\mathbf{x})$ in $x_1, x_2, \ldots, x_n$ with coefficients a_α is

$$f(\mathbf{x}) = \sum_\alpha a_\alpha \mathbf{x}^\alpha.$$

The total degree of f denoted as $\deg(f)$ is the maximum $|\alpha|$ with nonzero coefficient a_α. The above result implies that if the nonlinear polynomials are arranged in decreasing order of their $\deg(f)$ and within each group with the same $\deg(f)$ the polynomials are arranged in decreasing order of their B_{ii} values or in increasing order of their $1/B_{ii}$ values, then one often can readily find the best value of λ by properly dividing the nonlinear polynomials into two groups. For the
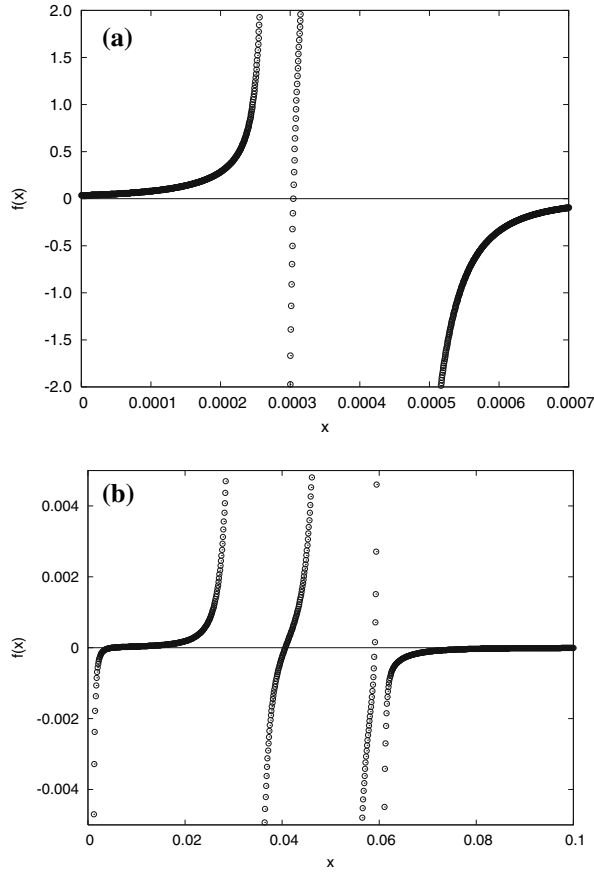
Figure 3. The behavior of the algebraic equation, equation (57), for the regularization parameter x associated with the first-order approximation of RS-HDMR.

first-order approximation of RS-HDMR which is approximated by third-order orthonormal polynomials, there are at most three possible divisions: (a) group 1: the third- and second-order polynomials, group 2: empty, i.e., all the third- and second-order polynomials need to be regularized; (b) group 1: the third-order polynomials, group 2: the second-order polynomials, and only the third-order polynomials need to be regularized; (c) group 1: empty, group 2: all the third- and second-order polynomials, i.e., none needs to be regularized. The corresponding λ value in each case is: (a) a value smaller than the reciprocal of the largest B_{ii} for the third-order polynomials; (b) a value between the reciprocals of the smallest B_{ii} for the third-order polynomials and the largest B_{ii} for the second-order polynomials; and (c) a value larger than the reciprocal of the smallest B_{ii} for the second-order polynomials. We may simply pick λ as the reciprocal of the largest B_{ii} for the third-order polynomials for case (a), the reciprocal of the smallest B_{ii}
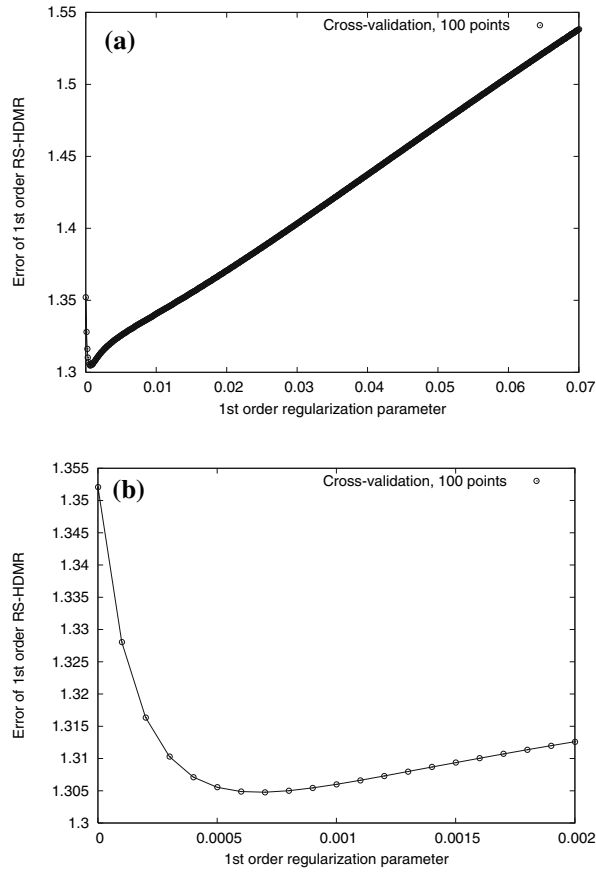
Figure 4. Direct estimation of regularization parameter by cross-validation for the first-order RS-HDMR component functions approximated by third-order orthonormal polynomials.

for the third-order polynomials for case (b), and the reciprocal of the smallest B_{ii} for the second-order polynomials for case (c) without significantly reducing the accuracy of the regularized RS-HDMR. In the H_2/air model, λ is chosen from case (b).

We used the remaining 1482 data points to test the interpolation error of the first-order regularized RS-HDMR approximation with respect to different values of λ. The results given in figure 5 are consistent with the cross-validation from the 100 points. Only the minimum moves slightly toward 0.0004. This behavior again suggests that one can simply choose $\lambda = 1/B_{99}$ without any direct performance of cross-validation.

Similarly, the algebraic equation to determine λ for the second-order regularized RS-HDMR component functions was constructed. The roots of its denominators are given below in increasing order:
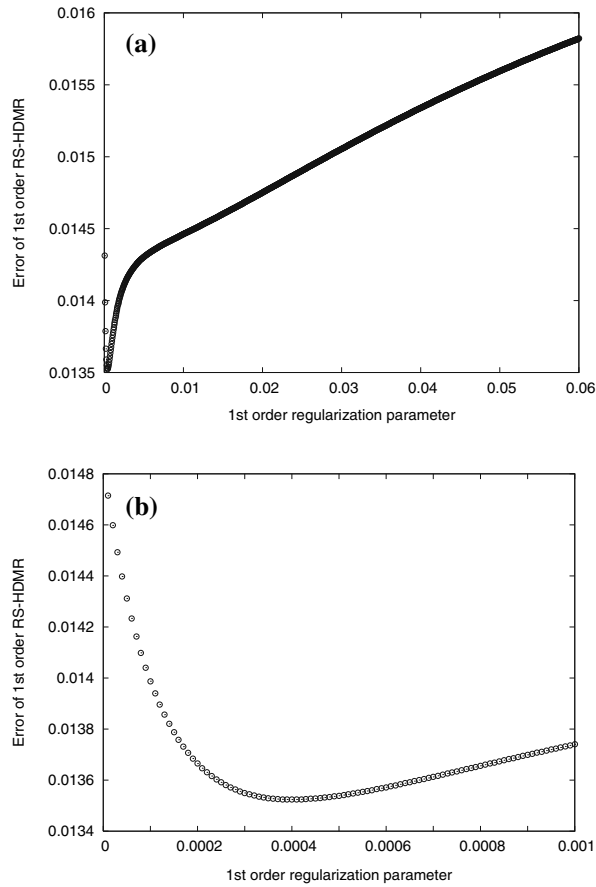
Figure 5. The relationship between the error $\sum e_i^2/N$ for 1482 test data and the regularization parameter for the first-order regularized RS-HDMR component functions approximated by third-order orthonormal polynomials.

0.0001	0.0003	0.0005	0.0005	0.0007	0.0009
0.0010	0.0013	0.0014	0.0016	0.0017	0.0019
0.0022	0.0033	0.0048	0.0048	0.0104	0.0111
0.0119	0.0133	0.0215	0.0670	0.1072	0.1094
0.1134	0.2304	0.3473	0.3775	0.5962	0.7012
11.2570	14.3229	15.3172			

Figure 6 shows the profile of $f(x)$, and some of the roots can be observed. Direct determination of the regularization parameter λ for the second-order regularized RS-HDMR component functions by cross-validation was also performed with the 100 data points, and the result is given in figure 7. The minimum is located at $\lambda \sim 0.02$. Similarly, this value of λ divides the basis functions in equation (11) into two groups. One group consists of the orthonormal
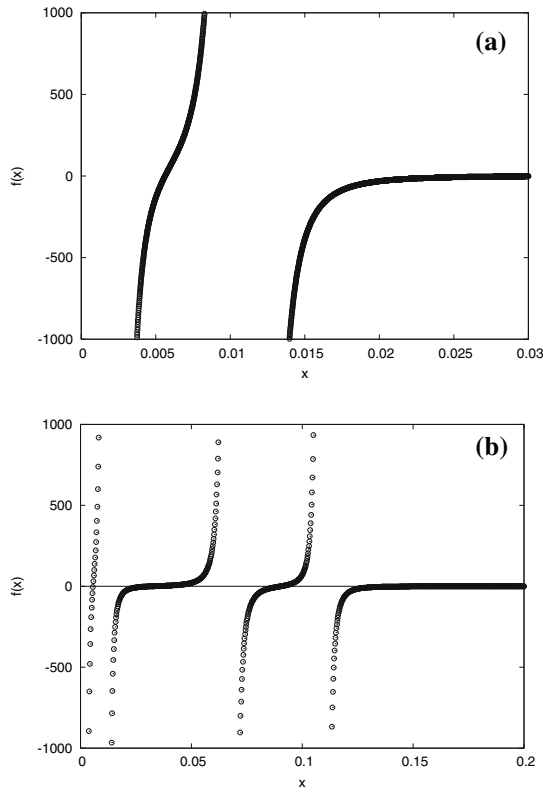
Figure 6. The behavior of the algebraic equation for the regularization parameter x associated with the second order RS-HDMR component functions.
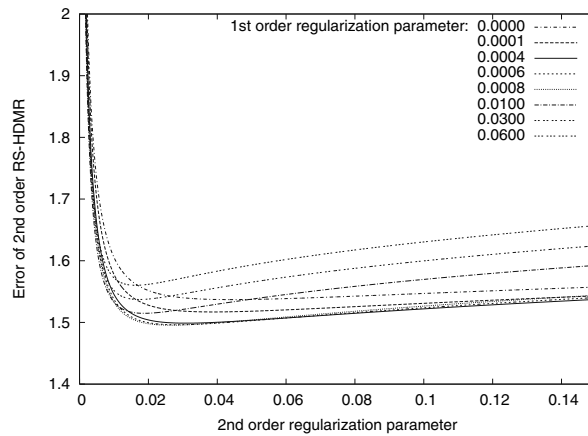

Figure 7. Direct estimation of the regularization parameter by cross-validation for the second-order RS-HDMR component functions approximated by third-order orthonormal polynomials for various values of the first order regularization parameters.
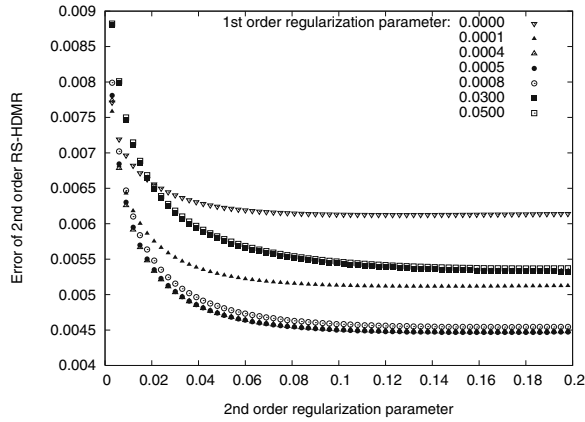
Figure 8. The relationship between the error $\sum e_i^2/N$ for $N = 1482$ test data and the regularization parameter for the second-order RS-HDMR component functions.

Table 5
The relative errors of different order RS-HDMR expansions obtained from the 1482 test data ($10 \leq P < 100$ atm), with and without regularization.

	Relative error (%)	Data portion			
		Used data		Test data	
		1st order	2nd order	1st order	2nd order
	5	0.880	0.980	0.841	0.907
Non-regularization	10	0.980	1.000	0.982	0.970
	20	0.990	1.000	1.000	0.996
	5	0.880	0.980	0.850	0.958
Regularization	10	0.990	0.990	0.985	0.993
	20	0.990	0.990	1.000	1.000

polynomials $\varphi_3^i(x_i)$ and $\varphi_r^i(x_i)\varphi_t^j(x_j)$ with $r + t \geq 4$. The other group consists of $\varphi_2^i(x_i)$ and $\varphi_r^i(x_i)\varphi_t^j(x_j)$ with $r + t \leq 3$. The boundary of the first group is $\varphi_1^2(x_2)\varphi_3^3(x_3)$, whose $1/B_{ii}$ value is 0.0215. This is also reasonable, i.e., the first group have large oscillations and need to be regularized; in contrast, the second group have small oscillations and need to be only slightly regularized.

The remaining 1482 data were used to test the interpolation error of the second-order regularized RS-HDMR approximation with respect to different values of λ. The results are given in figure 8, which are also consistent with cross-validation from the 100 points. The only difference is that the regularization parameter λ with the smallest error for the second-order RS-HDMR is larger than 0.0215, and there is no difference when $\lambda > 0.1$. These results suggest that one can choose either one of the two boundary $1/B_{ii}$ values as λ (i.e., the
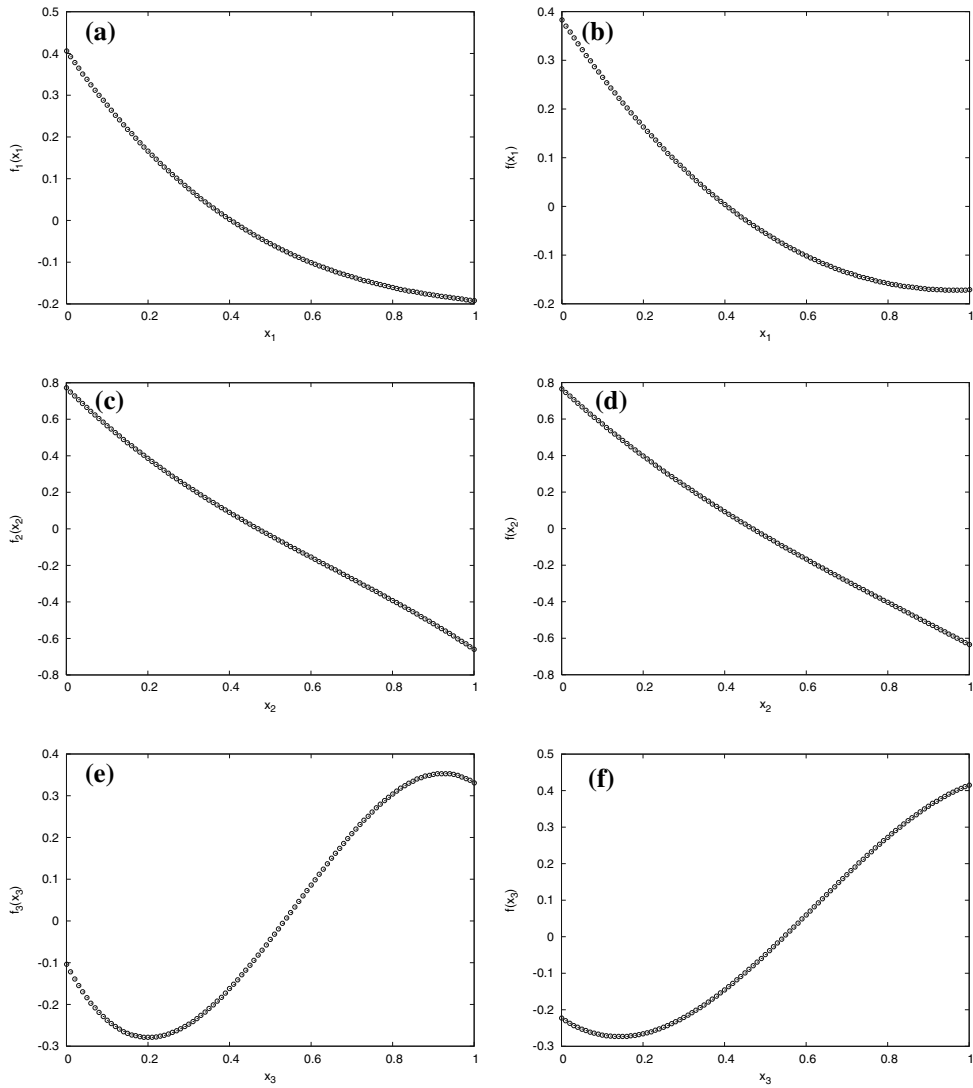
Figure 9. The comparison between $f_i(x_i)$ where $x_1 = T_0$, $x_2 = \log P$, $x_3 = R$ obtained from 100 data points with (right panels) and without (left panels) regularization for the $P \geqslant 10$ atm data.

largest $1/B_{ii}$ of group 1 or the smallest $1/B_{ii}$ of group 2. In our case, they are 0.0215 and 0.0670). If necessary, the cross-validation may be calculated only for the two values, and the computational cost is small.

The regularization parameters λ for the first- and second-order regularized RS-HDMR component functions were set to 0.000412 and 0.0215, respectively, simply based on separation of the orthonormal polynomials without any cross validation calculations. Since the accuracy of the second-order regularized
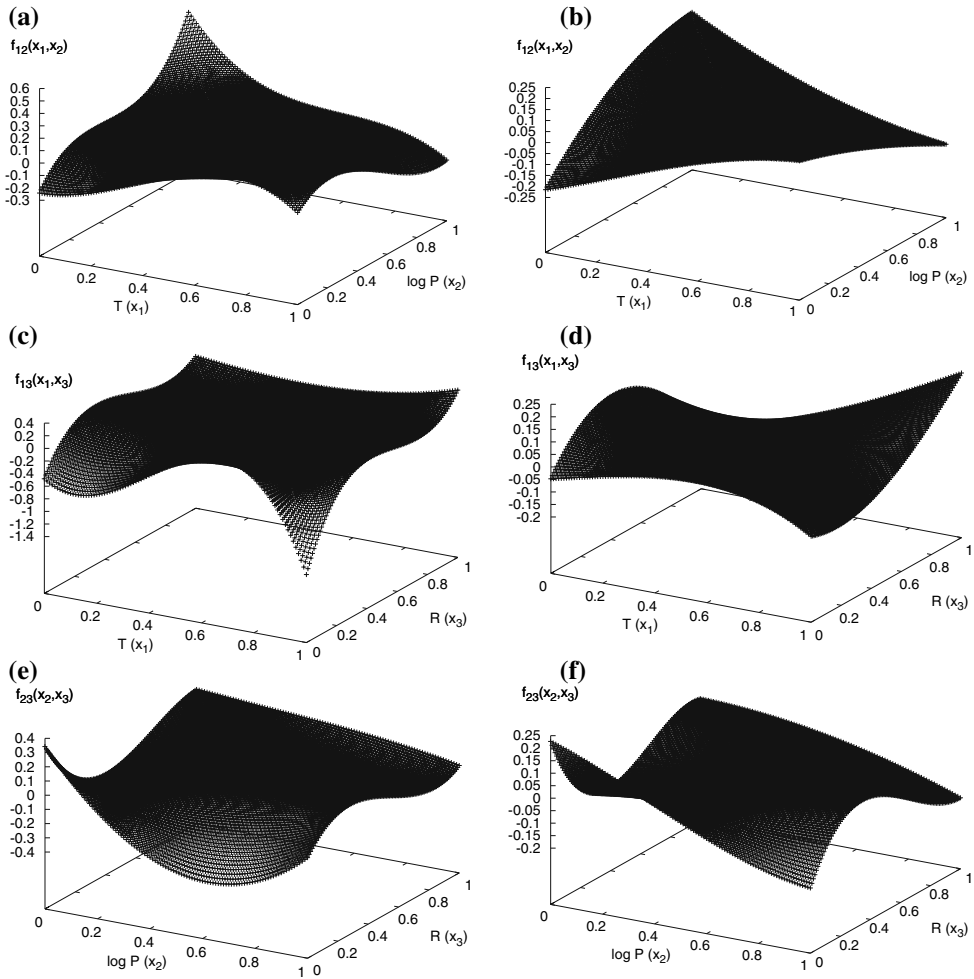
Figure 10. The comparison between $f_{ij}(x_i, x_j)$ where $x_1 = T_0$, $x_2 = \log P$, $x_3 = R$ obtained from 100 data points with (right panels) and without (left panels) regularization for the $P \geqslant 10$ atm data.

RS-HDMR does not change much around the above chosen values of λ, the method of dividing the basis function $\{\varphi\}$ into two groups and determining the values of λ from the corresponding values of $1/B_{ii}$ is very efficient.

The component functions $f_i(x_i)$, $f_{ij}(x_i, x_j)$ obtained from the 100 data points, with and without regularization, are given in figures 9 and 10. Using regularization, the accuracy for the used data was reduced because of the trade off between fidelity against regularization, but the accuracy for the test data was improved. The resultant second-order regularized RS-HDMR expansion was used to test the remaining 1482 test data. Comparison between the accuracy of the regularized and nonregularized RS-HDMR's for the used and test data is

given in table 5. The data portion with relative error not larger than 5% increases from 90.7 to 95.8%.

4. Conclusion

RS-HDMR is a practical approach for interpolating multi-dimensional functions based on randomly sampling the input variables. To reduce the sampling effort, the RS-HDMR component functions are approximated by expansions in terms of a suitable set of basis functions, for instance, orthonormal polynomials. Since the expansion coefficients are determined using either Monte Carlo integration or stepwise least squares regression, there can be oscillation of the resultant orthonormal polynomial expansion, especially on the input domain boundary when the sample size is not large. To reduce this error due to overfitting, regularization of the RS-HDMR component functions can be helpful. Cross-validation may be used to choose the regularization parameter, but it is computationally demanding. In this paper, an efficient approach to choose the regularization parameter for the RS-HDMR component functions is presented. The polynomial basis functions $\{\varphi\}$ of RS-HDMR are first divided into two groups according to their polynomial orders and total degrees, then a trial value of the regularization parameter λ is located between the values of $1/B_{ii}$ for the adjacent basis functions of the two groups. In RS-HDMR approximated by third-order orthonormal polynomials, the separation between the second- and third-order polynomials is often the right choice for the first-order RS-HDMR, and the separation of $\varphi_r^i(x_i)\varphi_t^j(x_j)$ with $r + t \geq 4$ is also proper for the second-order RS-HDMR. The $1/B_{ii}$ of the dividing boundary polynomials can be used as the regularization parameter. This method is not computationally demanding and is easy to perform. If B is not diagonal, then B_{ii} may be replaced by the eigenvalues of B, and a similar method can be established for choosing λ. The ignition time of a homogeneous H_2/air combustion model was used for successful illustration of this method. The same technique could be applied to a wide variety of other multi-dimensional interpolation problems.

References

[1] H. Rabitz, O.F. Alis, J. Shorter and K. Shim, Comput. Phys. Commun. 117 (1999) 11–20.

[2] K. Shim and H. Rabitz, Phys. Rev. B 58 (1998) 1940–1946.

[3] O.F. Alis and H. Rabitz, J. Math. Chem. 25 (1999) 197–233.

[4] J. Shorter and H. Rabitz, J. Phys. Chem. A 103 (1999) 7192–7198.

[5] O.F. Alis and H. Rabitz, J. Math. Chem. 29 (2001) 127–142.

[6] G. Li, C. Rosenthal and H. Rabitz, J. Phys. Chem. A 105 (2001) 7765–7777.

[7] S.W. Wang, H. Levy, II., G. Li and H. Rabitz, J. Geophys. Res. 104 (1999) No.D23 30417–30426.

[8] J. Shorter and H. Rabitz, Geophys. Res. Let. 27 (2000) 3485–3488.

[9] G. Li, S.W. Wang and H. Rabitz, J. Phys. Chem. 106 (2002) 8721–8733.

[10] G. Li, S.W. Wang, H. Rabitz, S.K. Wang and P. Jáffe, Chem. Eng. Sci. 57 (2002) 4445–4460.

[11] S. Wang, P.R. Jaffé, G. Li, S.W. Wang and H. Rabitz, J. Contam. Hydrol. 64 (2003) 283–307.

[12] S.W. Wang, P.G. Georgopoulos, G. Li and H. Rabitz, J. Phys. Chem. A 107 (2003) 4707–4718.

[13] G. Li, S.W. Wang, C. Rosenthal and H. Rabitz, J. Math. Chem. 30 (2001) 1–30.

[14] G. Li, M. Artamonov, H. Rabitz, S.W. Wang, P.G. Georgopoulos and M. Demiralp, J. Comput. Chem. 24 (2003) 647–656.

[15] G. Li, H. Rabitz, S.W. Wang and P.G. Georgopoulos, J. Comput. Chem. 24 (2003) 277–283.

[16] G. Li and H. Rabitz, J. Comput. Chem. 27 (2006) 1112–1118.

[17] J.M. Geremia, H. Rabitz and C. Rosenthal, J. Chem. Phys. 114 (2001) 9325–9336.

[18] N. Shenvi, J.M. Geremia and H. Rabitz, J. Phys. Chem. A 106 (2002) 12315–12323.

[19] J.M. Geremia and H. Rabitz, Phys. Rev. A 64(2) (2001) 022710.

[20] G. Li, H. Rabitz, S.W. Wang and P.G. Georgopoulos, IJRAM 5 (2005) 387–406.

[21] X. J. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss and H. Rabitz, Biophys. J. 87 (2004) 2195–2202.

[22] G. Li, J.S. Hu, S.W. Wang and P.G. Georgopoulos, J. Schoendorf and H. Rabitz, J. Phys. Chem. 110 (2006) 2474–2485.

[23] G. Wahba, Ann. Statist. 13 (1985) 1378–1402.

[24] G. Wahba, Y.D. Wang, C. Gu, R. Klein and B. Klein, Ann. Statist. 23 (1995) 1865–1895.

[25] J. Li, Z.W. Zhao, A. Kazakov and F.L. Dryer, Int. J. Chem. Kinet. 36 (2004) 566–575.