High-dimensional model representations generated from low order Terms – lp-RS-HDMR

ARTICLE *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · APRIL 2003

Impact Factor: 3.59 · DOI: 10.1002/jcc.10232 · Source: PubMed

6 AUTHORS, INCLUDING:

Genyuan Li
Princeton University
54 PUBLICATIONS **1,548** CITATIONS
SEE PROFILE

Herschel Rabitz
Princeton University
947 PUBLICATIONS **23,653** CITATIONS
SEE PROFILE

Panos Georgopoulos
Rutgers, The State University of New Jersey
165 PUBLICATIONS **2,515** CITATIONS
SEE PROFILE

High-Dimensional Model Representations Generated from Low Order Terms—lp-RS-HDMR

GENYUAN LI,[1] MAXIM ARTAMONOV,[1] HERSCHEL RABITZ,[1] SHENG-WEI WANG,[2]
PANOS G. GEORGOPOULOS,[2] METIN DEMIRALP[3]

[1]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544*
[2]*Environmental and Occupational Health Sciences Institute, 170 Frelinghuysen Road, Piscataway, New Jersey 08854*
[3]*Computational Science and Engineering Division, Informatics Institute, Istanbul Technical University, Maslak, 80626, Istanbul, Turkey*

Received 6 August 2002; Accepted 7 November 2002

Abstract: High-dimensional model representation (HDMR) is a general set of quantitative model assessment and analysis tools for improving the efficiency of deducing high dimensional input–output system behavior. RS-HDMR is a particular form of HDMR based on random sampling (RS) of the input variables. The component functions in an HDMR expansion are optimal choices tailored to the n-variate function $f(\mathbf{x})$ being represented over the desired domain of the n-dimensional vector $\mathbf{x}$. The high-order terms (usually larger than second order, or equivalently beyond cooperativity between pairs of variables) in the expansion are often negligible. When it is necessary to go beyond the first and the second order RS-HDMR, this article introduces a modified *low-order term product* (lp)-RS-HDMR method to approximately represent the high-order RS-HDMR component functions as products of low-order functions. Using this method the high-order truncated RS-HDMR expansions may be constructed without directly computing the original high-order terms. The mathematical foundations of lp-RS-HDMR are presented along with an illustration of its utility in an atmospheric chemical kinetics model.

 J Comput Chem 24: 647–656, 2003

Key words: HDMR; random sampling; Monte Carlo integration; atmospheric modeling

Introduction

High-dimensional model representation (HDMR) as a general set of quantitative model assessment and analysis tools for capturing high-dimensional input–output system behavior has been introduced recently.[1–4] HDMR expresses the model output $f(\mathbf{x})$ as a finite hierarchical cooperative function expansion in terms of its input variables:

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

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$; $f_0, f_i(x_i), f_{ij}(x_i, x_j) \cdots$ are zeroth, first, second, $\cdots$ order component functions, respectively; i.e., $f(\mathbf{x})$ can be decomposed as a finite sum of constant, 1-variate, 2-variate, . . . functions.

Distinct, but formally equivalent HDMR expansions, all of the same structure as eq. (1), may be constructed. When data are randomly sampled, this produces the RS(Random Sampling)-HDMR expansion. For RS-HDMR, we first rescale the variables x_i by some suitable transformations such that $0 \le x_i \le 1$ for all i. The output function $f(\mathbf{x})$ is then defined in the unit hypercube $K^n = \{(x_1, x_2, \ldots, x_n) | 0 \le x_i \le 1, i = 1, 2, \ldots, n\}$. The component functions of RS-HDMR have the following forms[2]:

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

Correspondence to: H. Rabitz; e-mail: hrabitz@princeton.edu

Contract/grant sponsor: Petroleum Research Fund of the American Chemical Society

Contract/grant sponsor: U.S. Environmental Protection Agency; contract/grant number: EPAR-827033

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

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

$$f_{ijk}(x_i, x_j, x_k) = \int_{K^{n-3}} f(\mathbf{x}) d\mathbf{x}^{ijk} - f_{ij}(x_i, x_j) - f_{ik}(x_i, x_k) - f_{jk}(x_j, x_k) - f_i(x_i) - f_j(x_j) - f_k(x_k) - f_0 , \tag{5}$$

$$\ldots ,$$

where $d\mathbf{x}^i$, $d\mathbf{x}^{ij}$ and $d\mathbf{x}^{ijk}$ are the products $dx_1 dx_2 \cdots dx_n$ without dx_i, $dx_i dx_j$ and $dx_i dx_j dx_k$, respectively. Finally, the last term $f_{12\cdots n}(x_1, x_2, \ldots, x_n)$ is determined from the difference between $f(\mathbf{x})$ and all the other component functions in eq. (1). The component functions of RS-HDMR satisfy the condition that the integral of a component function with respect to any of its own variables is zero, i.e.,

$$\int_0^1 f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) dx_s = 0, \quad s \in \{i_1, i_2, \ldots, i_l\}, \tag{6}$$

which defines the orthogonality relation between two RS-HDMR component functions as

$$\int_{K^n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) f_{j_1 j_2 \ldots j_k}(x_{j_1}, x_{j_2}, \ldots, x_{j_k}) d\mathbf{x} = 0,$$

$$\{i_1, i_2, \ldots, i_l\} \neq \{j_1, j_2, \ldots, j_k\}. \tag{7}$$

The component functions of RS-HDMR also can be obtained through application of a suitably defined set of linear operators $\wp_0$, $\wp_i$ $(i = 1, 2, \ldots, n)$, $\wp_{ij}$ $(1 \leq i < j \leq n)$, ... acting on a linear space $\mathcal{F}$ composed of all n-variate functions $f(\mathbf{x})$ as

$$\wp_0 f(\mathbf{x}) = f_0, \tag{8}$$

$$\wp_i f(\mathbf{x}) = f_i(x_i), \tag{9}$$

$$\wp_{ij} f(\mathbf{x}) = f_{ij}(x_i, x_j), \tag{10}$$

$$\ldots .$$

It has been proven that all the operators are commutative projection operators, and they are mutually orthogonal to one another.[1,2] These properties imply that

$$\wp_p^2 = \wp_p, \tag{11}$$

$$\wp_p \wp_q = 0, \quad (p \neq q), \tag{12}$$

where $\wp_p$, $\wp_q$ are members of the operators defined above. As these projectors are mutually orthogonal, adding more terms in the RS-HDMR expansion will improve the accuracy of the representation of $f(\mathbf{x})$ on $\mathcal{F}$, i.e., a higher order truncated RS-HDMR expansion is better than any lower order ones where the error is measured under the integration metric used in eqs. (2)–(5).

To save sampling effort, the RS-HDMR component functions are approximated by orthonormal polynomial expansions[5]:

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i \varphi_r(x_i), \tag{13}$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j), \tag{14}$$

$$f_{ijk}(x_i, x_j, x_k) \approx \sum_{p=1}^{m} \sum_{q=1}^{m'} \sum_{r=1}^{m''} \gamma_{pqr}^{ijk} \varphi_p(x_i) \varphi_q(x_j) \varphi_r(x_k), \tag{15}$$

where α_r^i, β_{pq}^{ij}, γ_{pqr}^{ijk} are constant expansion coefficients, k, l, l', m, m', m'' are integers often not larger than 3 for many systems, and

$$\varphi_1(x) = \sqrt{3}\,(2x - 1), \tag{16}$$

$$\varphi_2(x) = 6\sqrt{5}\,(x^2 - x + \tfrac{1}{6}), \tag{17}$$

$$\varphi_3(x) = 20\sqrt{7}\,(x^3 - \tfrac{3}{2}x^2 + \tfrac{3}{5}x - \tfrac{1}{20}), \tag{18}$$

$$\ldots ,$$

which are obtained from the orthonormality condition

$$\int_0^1 \varphi_r(x) dx = 0, \quad \forall\, r \tag{19}$$

$$\int_0^1 \varphi_r^2(x) dx = 1, \quad \forall\, r \tag{20}$$

$$\int_0^1 \varphi_p(x) \varphi_q(x) dx = 0, \quad (p \neq q). \tag{21}$$

In the approximations of eqs. (13)–(15), the orthonormality property of the set $\{\varphi\}$ preserves the mutual orthogonality of the RS-HDMR component functions given by eq. (7), and eq. (1) is equivalent to

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \alpha_r^i \varphi_r(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij}\varphi_p(x_i)\varphi_q(x_j) + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{ijk}\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k) + \cdots. \tag{22}$$

Using the orthonormality property of $\{\varphi\}$ the expansion coefficients $\{\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk}, \cdots\}$ can be obtained and approximately represented by Monte Carlo integration

$$\alpha_r^i = \int_{K^n} f(\mathbf{x})\varphi_r(x_i)d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\varphi_r(x_i^{(s)}), \tag{23}$$

$$\beta_{pq}^{ij} = \int_{K^n} f(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}), \tag{24}$$

$$\gamma_{pqr}^{ijk} = \int_{K^n} f(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)})\varphi_r(x_k^{(s)}), \tag{25}$$

$$\cdots,$$

where $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \ldots, x_n^{(s)})$ $(s = 1, 2, \ldots, N)$ is the sth sample, and N is the total number of random samples.

When Monte Carlo integration approximation is used, for a given set of N samples the orthogonality of the elements of $\{\varphi\}$ no longer holds, for example,

$$\int_{K^n} \varphi_r(x_i)\varphi_p(x_i)\varphi_q(x_j)d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} \varphi_r(x_i^{(s)})\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}) \neq 0. \tag{26}$$

Then all the coefficients $\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk}, \cdots$ in eq. (22) are coupled with one another, and we may determine them simultaneously by solving a system of linear algebraic equations. However, for a high dimension n, the number of linear algebraic equations can be very large, and solving the equations is not computationally efficient.[9]

The theoretical foundation of Monte Carlo integration is the following.[8] If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are independent random variables with a uniform probability density function and $F(\mathbf{x})$ is a function of $\mathbf{x}$, then the random variable

$$F_N = \frac{1}{N}\sum_{s=1}^{N} F(\mathbf{x}^{(s)}) \tag{27}$$

has the expectation

$$\langle F_N\rangle = \int_{K^n} F(\mathbf{x})d\mathbf{x}, \tag{28}$$

the variance

$$\mathrm{var}\{F_N\} = \frac{1}{N}\mathrm{var}\{F(\mathbf{x})\}, \tag{29}$$

and the standard deviation (standard error)

$$\sigma\{F_N\} = (\mathrm{var}\{F_N\})^{1/2} = \frac{1}{\sqrt{N}}\sigma\{F(\mathbf{x})\}. \tag{30}$$

Therefore, F_N can be used as an estimate of the integral $\int_{K^n} F(\mathbf{x})d\mathbf{x}$ with a standard error $\sigma\{F_N\}$ proportional to the standard error $\sigma\{F(\mathbf{x})\}$ of the integrand random variable $F(\mathbf{x})$. The error of Monte Carlo integration can be reduced either by increasing the sample size N or decreasing the variance of $F(\mathbf{x})$ in K^n. Monte Carlo integration error becomes troublesome when the random data of the integrand $F(\mathbf{x})$ has a large variance, i.e., $F(\mathbf{x})$ has rapid changes in the desired domain, especially in sign. This behavior is expected to arise when considering integrands in eqs. (23)–(25) with more $\{\varphi\}$s such as $f(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)$ and $f(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)$ because the functions $\{\varphi\}$ are fixed regardless of the form of $f(\mathbf{x})$. The determination of the expansion coefficients of high order RS-HDMR component functions by Monte Carlo integration generally requires additional samples. For example, to determine α_r^i by eq. (23) a few hundred samples may give a good accuracy, but for β_{pq}^{ij} in eq. (24) to achieve the same accuracy thousands of samples may be needed, and for γ_{pqr}^{ijk} in eq. (25) thousands of samples are still not enough. However, the sample size is often restricted by time and cost. The other way to improve the accuracy of Monte Carlo integration is to reduce the variance of the integrand.

To improve the accuracy of Monte Carlo integration, the correlation method may be employed to reduce the variance of the integrand.[9] Consider an integral for any coefficient in eqs. (23)–(25), for example

$$\alpha_r^i = \int_{K^n} f(\mathbf{x})\varphi_r(x_i)d\mathbf{x}. \tag{31}$$

The variance of the integrand $f(\mathbf{x})\varphi_r(x_i)$ in K^n can be reduced if one can find a reference function $h(\mathbf{x})$ satisfying two conditions: (1) $f(\mathbf{x}) - h(\mathbf{x})$ is almost constant or vanishes in the whole domain, and (2) the integral

$$\int_{K^n} h(\mathbf{x})\varphi_r(x_i)d\mathbf{x} = c_r^i \tag{32}$$

is known analytically. Then

$$\alpha_r^i = \int_{K^n} [f(\mathbf{x}) - h(\mathbf{x})]\varphi_r(x_i)d\mathbf{x} + \int_{K^n} h(\mathbf{x})\varphi_r(x_i)d\mathbf{x}. \tag{33}$$

Now the variance comes only from the first term in eq. (33). As $f(\mathbf{x}) - h(\mathbf{x})$ is almost constant or zero everywhere, we expect that

$$\mathrm{var}\{[f(\mathbf{x}) - h(\mathbf{x})]\varphi_r(x_i)\} < \mathrm{var}\{f(\mathbf{x})\varphi_r(x_i)\}. \tag{34}$$

α_r^i may be approximated by Monte Carlo integration

$$\alpha_r^i \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_r(x_i^{(s)}) + c_r^i \tag{35}$$

with a better accuracy than that given by eq. (23). Similarly, we also have

$$\beta_{pq}^{ij} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}) + c_{pq}^{ij}, \tag{36}$$

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)})\varphi_r(x_k^{(s)}) + c_{pqr}^{ijk}, \tag{37}$$

where

$$c_{pq}^{ij} = \int_{K^n} h(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)d\mathbf{x}, \tag{38}$$

$$c_{pqr}^{ijk} = \int_{K^n} h(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)d\mathbf{x}, \tag{39}$$

and formulas for other high-order constant expansion coefficients.

A truncated RS-HDMR expansion of eq. (22) satisfies these two conditions and can be used as $h(\mathbf{x})$, for instance the third-order expansion

$$h(\mathbf{x}) = f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k}\bar{\alpha}_r^i\varphi_r(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'}\bar{\beta}_{pq}^{ij}\varphi_p(x_i)\varphi_q(x_j) + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''}\bar{\gamma}_{pqr}^{ijk}\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k), \tag{40}$$

where the coefficients $\{\bar{\alpha}_r^i, \bar{\beta}_{pq}^{ij}, \bar{\gamma}_{pqr}^{ijk}\}$ are determined by direct Monte Carlo integration given in eqs. (23)–(25). The difference $f(\mathbf{x}) - h(\mathbf{x})$ should be small if the truncated RS-HDMR expansion is a good approximation of $f(\mathbf{x})$. Moreover, the second condition holds by using the orthonormality property of $\{\varphi\}$

$$\begin{aligned}\int_{K^n} h(\mathbf{x})\varphi_r(x_i)d\mathbf{x} &= \int_{K^n}\Bigg[f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k}\bar{\alpha}_r^i\varphi_r(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'}\bar{\beta}_{pq}^{ij}\varphi_p(x_i)\varphi_q(x_j) \\ &\quad + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''}\bar{\gamma}_{pqr}^{ijk}\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)\Bigg] \\ &\quad \times \varphi_r(x_i)d\mathbf{x} \\ &= \bar{\alpha}_r^i.\end{aligned} \tag{41}$$

Then we have

$$\alpha_r^i \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_r(x_i^{(s)}) + \bar{\alpha}_r^i. \tag{42}$$

Similarly, we also have

$$\beta_{pq}^{ij} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}) + \bar{\beta}_{pq}^{ij}, \tag{43}$$

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)})\varphi_r(x_k^{(s)}) + \bar{\gamma}_{pqr}^{ijk}. \tag{44}$$

Equations (42)–(44) show that the first terms in these equations are corrections for the initial values $\bar{\alpha}_r^i$, $\bar{\beta}_{pq}^{ij}$ and $\bar{\gamma}_{pqr}^{ijk}$. The resultant α_r^i, β_{pq}^{ij} and γ_{pqr}^{ijk} may be reused as new initial values for the construction of a new $h(\mathbf{x})$ with even smaller values of $f(\mathbf{x}) - h(\mathbf{x})$ to repeat the calculation again. Then, eqs. (42)–(44) become an iteration procedure. The iteration procedure should be convergent if the initial $h(\mathbf{x})$ is close to $f(\mathbf{x})$ and the sample size N is large enough.

If one chooses the second-order expansion as a reference function

$$h(\mathbf{x}) = f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k}\bar{\alpha}_r^i\varphi_r(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'}\bar{\beta}_{pq}^{ij}\varphi_p(x_i)\varphi_q(x_j), \tag{45}$$

then

$$\begin{aligned}c_{pqr}^{ijk} &= \int_{K^n} h(\mathbf{x})\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)d\mathbf{x} \\ &= \Bigg[f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k}\bar{\alpha}_r^i\varphi_r(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'}\bar{\beta}_{pq}^{ij}\varphi_p(x_i)\varphi_q(x_j)\Bigg]\varphi_p(x_i)\varphi_q(x_j)\varphi_r(x_k)d\mathbf{x} \\ &= 0,\end{aligned} \tag{46}$$

and γ_{pqr}^{ijk} may be determined without iteration:

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)})\varphi_r(x_k^{(s)}). \quad (47)$$

The correlation method achieves improved accuracy through the enhanced quality of the Monte Carlo integration. Thus, for a particular choice of basis $\{\varphi\}$, reference function $h(\mathbf{x})$, and sample size N convergence of the iteration is expected to give the best possible results. This behavior was confirmed later, where the correlation method using only a few hundred samples gave good accuracy for the determination of α_r^i, β_{pq}^{ij} and γ_{pqr}^{ijk}.

The so called Cut-HDMR formulation[1,2] is based on a special ordered sampling (i.e., cuts) through the input space $\mathbf{x}$ along lines, planes, etc. A modified Cut-HDMR method referred to as *monomial based preconditioned* (mp)-Cut-HDMR was introduced[6] whereby the high-order component functions of Cut-HDMR were approximately represented in a similar fashion as those for the zeroth-, first-, and second-order component functions. This procedure permitted the inclusion of the higher order Cut-HDMR approximations without dramatically increasing the number of experiments or model runs as well as reducing the computer storage requirements. In this article a method denoted as *low-order term product* (lp)-RS-HDMR is presented for this purpose within RS-HDMR. Using this method the high-order RS-HDMR component functions are approximated as products of low order ones. The theoretical foundation of lp-RS-HDMR is presented along with an illustrative application to an atmospheric chemical kinetics model.

The article is organized as follows. The next section introduces orthonormal bases consisting of normalized low-order RS-HDMR component functions and their products. Any function $f(\mathbf{x})$ can be approximately expanded in these bases. Then we present an illustration of this method to an atmospheric chemical kinetics model. Finally, we conclude.

Orthonormal Bases Composed of Normalized Low-Order RS-HDMR Component Functions

We define a set of functions for $l = 1, 2, \ldots, n$

$$g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = \begin{cases} \prod_{s=1}^{l} f_{i_s}(x_{i_s}) \Big/ \prod_{s=1}^{l} \|f_{i_s}(x_{i_s})\|, \\ 0, \quad \text{if any } \|f_{i_s}(x_{i_s})\| = 0, \end{cases} \quad (48)$$

where $f_{i_s}(x_{i_s})$ are the first order RS-HDMR functions and

$$\|f_{i_s}(x_{i_s})\| = \left[\int_0^1 f_{i_s}^2(x_{i_s})dx_{i_s}\right]^{1/2}, \quad i_s \in \{1, 2, \ldots, n\}. \quad (49)$$

For completeness we set

$$g_0 = \frac{f_0}{\|f_0\|}, \quad (50)$$

and the set of functions may be expressed as

$$g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = \prod_{s=1}^{l} g_{i_s}(x_{i_s}), \quad (51)$$

where $g_{i_s}(x_{i_s}) = f_{i_s}(x_{i_s})/\|f_{i_s}(x_{i_s})\|$ for $\|f_{i_s}(x_{i_s})\| \neq 0$. Considering that the functions $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ are a separable product and using the property of RS-HDMR component functions given by eq. (6), it can be readily proved that $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ $(l = 0, 1, \ldots, n)$ are orthonormal, i.e.,

$$\int_0^1 g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})dx_s = 0, \quad s \in \{i_1, i_2, \ldots, i_l\} \quad (52)$$

$$\int_{K^n} g_{i_1 i_2 \ldots i_l}^2(x_{i_1}, x_{i_2}, \ldots, x_{i_l})d\mathbf{x} = 1, \quad (53)$$

$$\int_{K^n} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})\, g_{j_1 j_2 \ldots j_k}(x_{j_1}, x_{j_2}, \ldots, x_{j_k})d\mathbf{x} = 0,$$

$$\{i_1, i_2, \ldots, i_l\} \neq \{j_1, j_2, \ldots, j_k\}. \quad (54)$$

The set of functions $\{g\}$ may be used as a basis to expand $f(\mathbf{x})$,

$$\begin{aligned} f(\mathbf{x}) \approx \alpha_0 g_0 &+ \sum_{i=1}^{n} \alpha_i g_i(x_i) + \sum_{1 \le i < j \le n} \alpha_{ij} g_{ij}(x_i, x_j) + \cdots \\ &+ \sum_{1 \le i_1 < \cdots < i_l \le n} \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots + \cdots \\ &+ \alpha_{12 \ldots n} g_{12 \ldots n}(x_1, x_2, \ldots, x_n), \end{aligned} \quad (55)$$

where $\{\alpha\}$ are constant coefficients that can be obtained by using the orthonormality property of $\{g\}$, i.e.,

$$\alpha_{i_1 i_2 \ldots i_l} = \int_{K^n} f(\mathbf{x})\, g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})d\mathbf{x}, \quad (l = 0, 1, \ldots, n). \quad (56)$$

As $\{g\}$ is not a complete set, the expansion of an n-variate function $f(\mathbf{x})$ in this basis is likely only an approximation. Considering that

$$\alpha_0 = \|f_0\|, \quad (57)$$

$$\alpha_i = \|f_i(x_i)\|, \quad (i = 1, 2, \ldots, n), \quad (58)$$

eq. (55) becomes

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} \alpha_{ij} g_{ij}(x_i, x_j) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n} \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots + \cdots + \alpha_{12\ldots n} g_{12\ldots n}(x_1, x_2, \ldots, x_n), \quad (59)$$

which implies the following approximations upon comparison with eq. (1)

$$f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) \approx \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}), \quad (l = 2, 3, \ldots, n). \quad (60)$$

Notice that all $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ $(l \ge 2)$ are products of normalized first-order function $f_i(x_i)$. Hence, eq. (59) only requires computing the first-order RS-HDMR component functions, and the high order component functions are no longer explicitly computed. The coefficient $\alpha_{i_1 i_2 \ldots i_l}$ for $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ is determined by eq. (56) whose Monte Carlo integration approximation still involves the product of l one variable functions. At first sight, the sampling effort and achieved accuracy might appear similar to the $\{\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk}, \cdots\}$ for the orthonormal polynomial $\{\varphi\}$ expansions given by eqs. (13)–(15). However, the $f_i(x_i)$s are obtained by averaging $f(\mathbf{x})$ over K^{n-1} with respect to all x_j except x_i, and this basis reflects the behavior of $f(\mathbf{x})$. In contrast, $\{\varphi\}$ is a fixed set of functions without specific knowledge of the particular n-variate function $f(\mathbf{x})$ of interest, the variance of $f(\mathbf{x})\ \Pi\ \varphi_i$ should be larger than those for $f(\mathbf{x})\ \Pi\ g_i$. Therefore, $\{g\}$ is a natural basis for $f(\mathbf{x})$, and using the Monte Carlo integration approximation with a given sample size, $\{g\}$ should yield better accuracy compared to $\{\varphi\}$. Moreover, compared to the $\{\varphi\}$ expansion given by eq. (22), the $\{g\}$ expansion given by eq. (59) has many fewer terms, and consequently, less unknown coefficients to be determined by Monte Carlo integration. For instance, the second- and third-order component functions $f_{ij}(x_i, x_j)$ and $f_{ijk}(x_i, x_j, x_k)$ have 9 and 27 terms, respectively, in eq. (22) if the third-order orthonormal polynomial expansion is used. In contrast, eq. (59) has only one term $g_{ij}(x_i, x_j)$, $g_{ijk}(x_i, x_j, x_k)$ for each. The *reduction* in the number of terms to calculate with the $\{g\}$ expansion increases exponentially with the order l of RS-HDMR component functions. The reduction in the number of Monte Carlo integration with the basis $\{g\}$ should also reduce the computational error.

The above treatment may be viewed from another perspective. Define new operators

$$\begin{aligned} \tilde{p}_{i_1 i_2 \ldots i_l} f(\mathbf{x}) &= \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) \quad (61)\\ &= \left(\int_{K^n} f(\mathbf{x})\, g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) d\mathbf{x} \right) \times g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}), \quad (l = 0, 1, \ldots, n). \end{aligned}$$

It is easy to prove that $\tilde{p}_{i_1 i_2 \ldots i_l}$ are projectors:

$$\begin{aligned} \tilde{p}^2_{i_1 i_2 \ldots i_l} f(\mathbf{x}) &= \tilde{p}_{i_1 i_2 \ldots i_l} [\alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})]\\ &= \left(\int_{K^n} \alpha_{i_1 i_2 \ldots i_l} g^2_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) d\mathbf{x} \right) \times g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})\\ &= \left(\alpha_{i_1 i_2 \ldots i_l} \int_{K^n} g^2_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) d\mathbf{x} \right) \times g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})\\ &= \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})\\ &= \tilde{p}_{i_1 i_2 \ldots i_l} f(\mathbf{x}). \quad (62) \end{aligned}$$

As $f(\mathbf{x})$ is arbitrary, this implies that

$$\tilde{p}^2_{i_1 i_2 \ldots i_l} = \tilde{p}_{i_1 i_2 \ldots i_l}, \quad (63)$$

i.e., $\tilde{p}_{i_1 i_2 \ldots i_l}$ is a projector. Using eq. (52), it can be also proven that all the $\tilde{p}_{i_1 i_2 \ldots i_l}$ $(l = 0, 1, \ldots, n)$ are mutually orthogonal. This means that the higher order truncated expansions in eq. (59) will yield better accuracy in representing functions $f(\mathbf{x})$ in $\mathcal{F}$ compared to the lower order expansions. The maximum projector $\mathcal{M}$ for the lattice generated by the orthogonal projectors $\{\tilde{p}\}$ is the sum of all the projectors[6,7]

$$\mathcal{M} = \sum_{l=0}^{n} \sum_{\substack{\{i_1 i_2 \ldots i_l\}\\ \subseteq\{1,2,\ldots,n\}}} \tilde{p}_{i_1 i_2 \ldots i_l}, \quad (64)$$

and the algebraically best approximation to $f(\mathbf{x})$ in $\mathcal{F}$ given by the projectors in this lattice is just $\mathcal{M}f(\mathbf{x})$.

$$\begin{aligned} f(\mathbf{x}) &\approx \mathcal{M} f(\mathbf{x})\\ &= \sum_{l=0}^{n} \sum_{\substack{\{i_1 i_2 \ldots i_l\}\\ \subseteq\{1,2,\ldots,n\}}} \tilde{p}_{i_1 i_2 \ldots i_l} f(\mathbf{x})\\ &= f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} \alpha_{ij} g_{ij}(x_i, x_j) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n} \alpha_{i_1 i_2 \ldots i_l} g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots + \cdots + \alpha_{12\ldots n} g_{12\ldots n}(x_1, x_2, \ldots, x_n). \quad (65) \end{aligned}$$

However, as the coefficients $\alpha_{i_1 i_2 \ldots i_l}$ are approximately determined by Monte Carlo integration, the higher order truncated expansions in eq. (65) may not always be better than the lower order ones, especially when l is very large.

The above treatment can be extended further by considering the products of the first- and second-order RS-HDMR component functions $f_i(x_i)$ and $f_{ij}(x_i, x_j)$. We define new functions for $l = 1, 2, \ldots, n$ and $p = 0, 1, \ldots, l$

$$g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = \begin{cases} \dfrac{\prod_{s=1}^{p} f_{i_s}(x_{i_s}) \prod_{r=1}^{(l-p)/2} f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})}{\prod_{s=1}^{p} \|f_{i_s}(x_{i_s})\| \prod_{r=1}^{(l-p)/2} \|f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})\|}, \\ 0, \quad \text{if any } \|f_{i_s}(x_{i_s})\| \text{ and/or } \|f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})\| = 0, \end{cases} \tag{66}$$

where

$$\|f_{ij}(x_i, x_j)\| = \left[\int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j \right]^{1/2}. \tag{67}$$

All the new functions $\{g\}$ defined in eq. (66) combined with g_0 given by eq. (50) are orthonormal. The only restriction is that any two functions in $\{g\}$ cannot have the same set of variables. For instance, if $\{g\}$ contains the function

$$g_{ijk}(x_i, x_j, x_k) = \frac{f_i(x_i) f_{jk}(x_j, x_k)}{\|f_i(x_i)\| \|f_{jk}(x_j, x_k)\|}, \tag{68}$$

it cannot also have

$$g_{ijk}(x_i, x_j, x_k) = \frac{f_i(x_i) f_j(x_j) f_k(x_k)}{\|f_i(x_i)\| \|f_j(x_j)\| \|f_k(x_k)\|} \tag{69}$$

and vice versa because the two functions are not orthogonal. The new set $\{g\}$ is another orthonormal basis to expand $f(\mathbf{x})$ as eq. (59) and the coefficients $\alpha_{i_1 i_2 \ldots i_l}$ are also determined by eq. (56). The extention to the products of even higher order RS-HDMR component functions is straight forward.

The special class of basis functions $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = 0$ in eqs. (48) and (66) impose restrictions in the expression in eq. (65). Although this is a deficiency, it is not expected to be significant, as this behavior in the basis functions is just a reflection of $f(\mathbf{x})$, from which they come. A reasonable expectation is that the higher order RS-HDMR terms will be small when associated with lower order terms that are identically zero. Thus, this choice of expansion functions is anticipated to be optimally suited to each particular application.

Table 1. The Variable Ranges.

	Range	
	Lower limit	Upper limit
Relative humidity, (%), x_1	10	100
CO (ppb), x_2	20	200
NO_x (ppt), x_3	1	950
O_3 (ppb), x_4	10	200

Table 2. The Relative Errors of the Different Order RS-HDMR Expansions Given by eq. (22) for Output D Determined by Direct Monte Carlo Integration.

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		First order	Second order	Third order
300	5	33.7	24.2	14.6
	10	55.0	43.4	28.1
	20	70.0	68.5	49.7
500	5	37.0	34.6	19.3
	10	55.7	59.6	36.8
	20	69.2	80.2	61.8
1000	5	38.6	58.6	36.8
	10	56.7	81.4	61.6
	20	69.7	93.6	83.8
3000	5	38.4	86.4	65.5
	10	55.4	95.5	84.0
	20	68.8	99.2	95.3
5000	5	38.4	91.8	75.9
	10	55.7	97.6	90.6
	20	68.9	99.6	97.1

[a]The percentage of 53,312 data with a relative error not larger than a given value.

Illustration: A Photochemical Box Model

A zero-dimensional photochemical box model designed to treat the ozone chemistry in the background troposphere is being used to study three-dimensional global chemical transport.[10] The box model consists of 63 reactions and 28 chemical species used to calculate the tendency, which is the difference of the rates of ozone production P and destruction D for incorporation into the overall three-dimensional model. The details of this process are not relevant here, but the box model provides a good testing ground for lp-RS-HDMR.

The rates of ozone production P, destruction D and the tendency $P - D$ are used as three output variables of the box model. The input variables $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$ here are the concentrations of the four precursors: H_2O, CO, NO_x, and O_3. In the following, we will apply the method discussed above to this model. The RS-HDMR component functions with respect to the different outputs are approximated by third-order orthonormal polynomial expansions. The expansion coefficients for $\{\varphi\}$ and $\{g\}$ were determined by Monte Carlo integration with different random sample sizes (300, 500, 1000, 3000, and 5000). Then the accuracy of the different approaches discussed above were examined using 53,312 samples, which uniformly cover the whole domain of $\mathbf{x}$. The ranges of the four variables are shown in Table 1.

Orthonormal Polynomial Expansion

Table 2 gives the accuracy of different order RS-HDMR expansions for output D whose component functions were approximated by the third order orthonormal polynomial expansions and the $\{\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk}\}$ were determined by direct Monte Carlo integration in eqs. (23)–(25) with different sample sizes.

As α_r^i can be accurately determined by direct Monte Carlo integration with a few hundred samples, the accuracy of the resultant first-order RS-HDMR approximations given in Table 2 does not much depend on the sample size. In contrast, the accurate determination of β_{pq}^{ij} needs thousands of samples. The third-order RS-HDMR expansion is worse than the second-order one, especially for small samples where it is even worse than the first-order one. This behavior means that even if thousands of samples are used, the direct Monte Carlo integration is still inaccurate for the determination of γ_{pqr}^{ijk}. The results for outputs P and $P - D$ are similar (not shown here).

The correlation method in eqs. (42)–(44) and (47) was used to determine the coefficients α_r^i, β_{pq}^{ij}, and γ_{pqr}^{ijk}. When $h(\mathbf{x})$ was chosen to be the third-order orthonormal polynomial expansion in eq. (40), the iterations were convergent for $N = 3000, 5000$, but divergent when $N \leq 1000$. Convergent results for the three outputs P, D and $P - D$ were obtained for all sample sizes when $h(\mathbf{x})$ was chosen to be the second-order orthonormal polynomial expansion in eq. (45). The coefficients α_r^i and β_{pq}^{ij} were determined by eqs. (42) and (43). The coefficients γ_{pqr}^{ijk} are determined by eq. (47). The accuracy of the resultant third-order RS-HDMR expansions for the three outputs are given in Table 3. Compared to Table 2, the results are satisfactory even for small sample sizes. This behavior shows that correlation method reduces the variance of the integrand, and hence improves the accuracy of Monte Carlo integration. Therefore, using the correlation method the third-order RS-HDMR expansions can be accurately constructed for the three outputs.

Table 3. The Relative Errors of the Third-Order RS-HDMR Expansions in eq. (22) Constructed by the Correlation Method with Monte Carlo Integration [$h(\mathbf{x})$ in eqs. (42)–(44) and (47) is the Second-Order RS-HDMR Expansion in eq. (45)].

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		P	D	$P - D$
300	5	93.6	92.7	98.9
	10	97.2	98.4	99.8
	20	99.6	100	100
500	5	95.6	96.5	98.0
	10	98.7	99.6	99.4
	20	99.8	100	99.8
1000	5	97.3	97.9	99.1
	10	99.1	99.7	99.7
	20	99.8	100	100
3000	5	99.2	99.6	99.5
	10	99.9	100	99.9
	20	100	100	100
5000	5	99.6	99.9	99.7
	10	100	100	100
	20	100	100	100

[a]The percentage of 53,312 data with a relative error not larger than a given value.

Table 4. The Relative Errors of the Approximation Given by the Basis $\{g\}$ Composed of the Normalized First-Order RS-HDMR Functions in eq. (59) Whose Expansion Coefficients $\alpha_{i_1 i_2 \ldots i_l}$ Were Obtained by Direct Monte Carlo Integration with Different Sample Sizes.

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		P	D	$P - D$
300	5	68.6	54.0	80.3
	10	85.8	78.3	97.4
	20	94.2	90.1	99.8
500	5	71.3	67.0	85.0
	10	90.1	85.4	95.7
	20	95.2	93.8	99.1
1000	5	71.7	93.9	94.7
	10	90.8	98.5	99.2
	20	95.9	99.7	100
3000	5	72.0	96.9	92.5
	10	91.9	99.3	98.9
	20	97.2	99.9	100
5000	5	71.6	97.3	93.0
	10	91.6	99.4	99.1
	20	97.1	100	100

[a]The percentage of 53,312 data with a relative error not larger than a given value.

Normalized $\mathbf{f_i(x_i)}$ *Expansion*

First, we apply the method whereby $f(\mathbf{x})$ is expanded in the basis $\{g\}$ composed of the normalized first-order RS-HDMR functions in eq. (59). The results obtained by direct Monte Carlo integration are given in Table 4. Compared to Table 2 one can see that the accuracy given by this basis is much better than the original first-order RS-HDMR approximation even though both of them only use $f_i(x_i)$. When $N \geq 1000$, the accuracy for D and $P - D$ are quite good, and better than those for the original second-order RS-HDMR approximation (see Table 2) despite the determination of the coefficients $\alpha_{i_1 i_2 \ldots i_l}$ [eq. (56)] by Monte Carlo integration involving the product of $l(\geq 2)$ one variable functions. These results show that the variance of $f(\mathbf{x}) \prod g_i$ is smaller than that for $f(\mathbf{x}) \prod \varphi_i$, and hence, $\{g\}$ is a better basis to expand $f(\mathbf{x})$ than $\{\varphi\}$. However, the results are not fully satisfactory, especially for small sample sizes and output P.

The correlation method with Monte Carlo integration where $h(\mathbf{x})$ is taken as an expansion up to fourth-order in the basis $\{g\}$ was used for the determination of the coefficients $\alpha_{i_1 i_2 \ldots i_l}$. The iterations for all outputs and sample sizes were convergent. Compared to direct Monte Carlo integration (see Table 4), the results are better, especially for outputs D and $P - D$ with a sample size equal to or larger than 1000. The results are listed in Table 5. However, compared to the results of the third-order RS-HDMR expansion with the basis $\{\varphi\}$ obtained by the correlation method with Monte Carlo integration (see Table 3), the accuracy of the normalized first-order RS-HDMR function basis expansion with the correlation method is still not fully satisfactory. It appears that for this atmospheric model the normalized first-order RS-HDMR function basis is far from complete. Thus, the next section will

explore going to second-order RS-HDMR component functions to form the basis $\{g\}$ in eq. (66).

Normalized $\mathbf{f_i(x_i)}$ *and* $\mathbf{f_{ij}(x_i, x_j)}$ *Expansion*

The basis $\{g\}$ composed of the first and second RS-HDMR component functions was explored. First, the coefficients α_r^i, β_{pq}^{ij}, and $\alpha_{i_1 i_2 \cdots i_l}$ were all obtained by direct Monte Carlo integration. Here, only the results obtained from the basis functions composed of $f_i(x_i)$ $(i = 1, 2, 3, 4)$ and $f_{12}(x_1, x_2)$, $f_{34}(x_3, x_4)$ are shown in Table 6. The results in Table 6 are generally better than those in Table 4. This is consistent with the conclusion that a projection with ρ_i is not better than that with ρ_i and ρ_{ij}. Some results, especially for D and $P - D$ at small sample sizes ($N \leq 1000$) in Table 6 are worse than those in Table 4. This behavior may be caused by the error in determining the coefficients β_{pq}^{ij} of the orthonormal polynomial expansion of $f_{ij}(x_i, x_j)$ by direct Monte Carlo integration, which needs more samples than for $f_i(x_i)$.

The correlation method was used for the basis $\{g\}$ composed of the normalized first- and second-order RS-HDMR functions. First, the orthonormal polynomial expansion coefficients α_r^i and β_{pq}^{ij} of $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ were determined by the correlation method with Monte Carlo integration and $h(\mathbf{x})$ being the second-order RS-HDMR expansion in eq. (45) for the construction of the basis $\{g\}$. Then the expansion coefficients $\alpha_{i_1 i_2 \cdots i_l}$ were obtained by the correlation method with Monte Carlo integration using different second-order functions $f_{ij}(x_i, x_j)$ and different order $h(\mathbf{x})$ of $\{g\}$ expansions. The best results are given in Table 7 with $h(\mathbf{x})$ being the third-order expansion in terms of the basis $\{g\}$. The accuracy is very satisfactory, and almost the same as that for the third-order RS-HDMR expansion with $\{\varphi\}$ obtained by the correlation method with Monte Carlo integration (see Table 3). A few hundred samples are sufficient to construct an accurate approximate expansion for $f(\mathbf{x})$.

Table 5. The Relative Errors of the Approximation Given by the Basis $\{g\}$ Composed of the Normalized First-Order RS-HDMR Functions in eq. (59) Whose Expansion Coefficients $\alpha_{i_1 i_2 \cdots i_l}$ Are Obtained by the Correlation Method with Monte Carlo Integration from Different Sample Sizes.

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		P	D	$P - D$
300	5	73.5	60.7	91.0
	10	90.2	87.0	98.9
	20	95.6	96.8	99.9
500	5	74.3	72.1	87.5
	10	92.4	90.5	98.1
	20	96.7	98.1	99.7
1000	5	73.4	96.5	94.1
	10	92.3	99.6	99.3
	20	97.0	100	99.9
3000	5	72.5	99.4	93.6
	10	92.3	100	99.3
	20	97.3	100	99.9
5000	5	71.7	99.0	94.0
	10	92.0	99.9	99.5
	20	97.3	100	100

[a]The percentage of 53,312 data with a relative error not larger than a given value.

Table 6. The Relative Errors of the Approximation Given by the Basis $\{g\}$ Composed of the Normalized First- and Second-Order RS-HDMR Functions in eq. (59) Whose Expansion Coefficients α_r^i, β_{pq}^{ij}, and $\alpha_{i_1 i_2 \cdots i_l}$ Were All Obtained by Direct Monte Carlo Integration with Different Sample Sizes.

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		P	D	$P - D$
300	5	29.2	32.7	35.5
	10	55.7	60.2	65.3
	20	85.7	88.8	96.3
500	5	74.2	59.0	71.6
	10	96.3	88.6	90.9
	20	100	98.3	99.1
1000	5	87.1	89.6	83.7
	10	99.2	98.7	96.6
	20	100	100	100
3000	5	82.3	96.9	83.4
	10	98.5	100	96.7
	20	100	100	100
5000	5	87.7	98.5	85.0
	10	99.2	100	97.3
	20	100	100	100

[a]The percentage of 53,312 data with a relative error not larger than a given value. The basis functions are composed of $f_i(x_i)$ $(i = 1, 2, 3, 4)$ and $f_{12}(x_1, x_2)$, $f_{34}(x_3, x_4)$.

Conclusion

To reduce the sampling effort for the RS-HDMR expansion, a modification to the formulation was considered. The basic concept underlying this modification is to represent the high-order RS-HDMR component functions as products of low order ones, and hence, this method is called low-order term product (lp)-RS-HDMR. The illustrations for an atmospheric model shows that the orthonormal bases $\{g\}$ composed of normalized low order (first and second) RS-HDMR component functions and their products can provide satisfactory accuracy similar to that for the third-order RS-HDMR expansion with the basis $\{\varphi\}$ whose expansion coefficients are determined by the correlation method with Monte Carlo integration. Only a few hundred samples were needed in both cases.

Compared to orthonormal polynomial $\{\varphi\}$ expansion of RS-HDMR functions, the orthonormal bases $\{g\}$ composed of normalized low-order (first and second) RS-HDMR component functions and their products has two advantages. First, because $f_i(x_i)$, $f_{ij}(x_i, x_j)$ are optimal choices tailored to a given $f(\mathbf{x})$ over the entire domain, $f(\mathbf{x}) \prod g_i$ have small variances and are more stable in the iteration for the determination of the expansion coefficients $\alpha_{i_1 i_2 \cdots i_l}$ by the correlation method with Monte Carlo integration

Table 7. The Relative Errors of the Approximation Given by the Basis $\{g\}$ Composed of the Normalized First- and Second-Order RS-HDMR Functions in eq. (59) Whose Expansion Coefficients α_r^i, β_{pq}^{ij}, and $\alpha_{i_1 i_2 \ldots i_l}$ Were All Obtained by the Correlation Method with Monte Carlo Integration Using Different Sample Sizes.

Sample size (N)	Relative error (%)	Data portion (%)[a]		
		P	D	$P - D$
300	5	94.6	93.4	98.5
	10	98.1	99.1	99.8
	20	99.8	100	100
500	5	96.2	93.4	97.5
	10	99.0	97.3	99.3
	20	100	99.3	99.9
1000	5	97.1	98.3	99.2
	10	99.4	100	99.9
	20	100	100	100
3000	5	98.1	99.2	99.7
	10	99.8	100	100
	20	100	100	100
5000	5	98.2	99.3	99.8
	10	99.8	100	100
	20	100	100	100

[a]The percentage of 53,312 data with a relative error not larger than a given value. For output P the basis functions are composed of $f_i(x_i)$ $(i = 1, 2, 3, 4)$ and $f_{13}(x_1, x_3)$, $f_{24}(x_2, x_4)$; for output D the basis functions are composed of $f_i(x_i)$ $(i = 1, 2, 3, 4)$ and $f_{12}(x_1, x_2)$, $f_{34}(x_3, x_4)$; for output $P - D$ the basis functions are composed of $f_i(x_i)$ $(i = 1, 2, 3, 4)$ and $f_{14}(x_1, x_4)$, $f_{23}(x_2, x_3)$.

than $f(\mathbf{x}) \prod \varphi_i$. In the four-dimensional atmospheric model, for the orthonormal polynomial $\{\varphi\}$ expansion, only the iteration with $h(\mathbf{x})$ constructed with the second-order RS-HDMR expansion is convergent for all sample sizes. In contrast, for the orthonormal bases $\{g\}$ the iteration is always convergent for $h(\mathbf{x})$ being any

Table 8. The Number of Iterations to Obtain Converged Values of the Coefficients α_r^i and β_{pq}^{ij} for Different Sample Sizes with $h(\mathbf{x})$ Being the Second-Order RS-HDMR Expansion with $\{\varphi\}$.

Sample size (N)	Coefficient	Iteration Number		
		P	D	$P - D$
300	α_r^i	6	5	6
	β_{pq}^{ij}	7	6	6
500	α_r^i	8	10	10
	β_{pq}^{ij}	13	20	10
1000	α_r^i	3	5	2
	β_{pq}^{ij}	4	5	2
3000	α_r^i	1	2	2
	β_{pq}^{ij}	2	2	2
5000	α_r^i	1	1	1
	β_{pq}^{ij}	2	2	1

Table 9. The Number of Iterations to Obtain Converged Values of the Coefficients α_{ij} and α_{ijk} for Different Sample Sizes with $h(\mathbf{x})$ Being the Third-Order $\{g\}$ Expansion.

Sample size (N)	Coefficient	Iteration Number		
		P	D	$P - D$
300	α_{ij}	2	4	3
	α_{ijk}	2	4	3
500	α_{ij}	2	4	2
	α_{ijk}	2	4	2
1000	α_{ij}	2	3	1
	α_{ijk}	2	3	2
3000	α_{ij}	1	1	1
	α_{ijk}	1	1	1
5000	α_{ij}	1	1	1
	α_{ijk}	1	1	1

order expansion and all sample sizes. Tables 8 and 9 give two examples for the numbers of iterations to reach converged results for these two bases. For the orthonormal polynomial $\{\varphi\}$ expansion, $h(\mathbf{x})$ is the second-order RS-HDMR expansion with $\{\varphi\}$. For the orthonormal bases $\{g\}$, $h(\mathbf{x})$ is the third-order expansion with $\{g\}$. The results show that the iteration convergence for the orthonormal bases $\{g\}$ expansion is stable and fast, especially for small sample sizes.

The orthonormal bases $\{g\}$ expansion has many fewer terms compared with the $\{\varphi\}$ basis. This advantage and the fast convergence shown in Table 9 using $\{g\}$ make it possible to construct higher order expansions than orthonormal polynomials $\{\varphi\}$, especially for large dimension n. Because each term of the $\{g\}$ expansion is produced by mutually orthogonal projectors, the high-order HDMR approximation produced with this basis should yield better accuracy than the low-order ones; however, the Monte Carlo integration error for the high-order coefficients may reduce this benefit even though the correlation method is employed. Therefore, the proper order of basis to employ depends on the system and the sample size.

References

1. Rabitz, H.; Alis, O. F.; Shorter, J.; Shim, K. Comput Phys Commun 1999, 117, 11.
2. Alis, O. H.; Rabitz, H. J Math Chem 1999, 25, 197.
3. Shorter, J.; Ip, P.C.; Rabitz, H. J Phys Chem A 1999, 103, 7192.
4. Alis, O. H.; Rabitz, H. J Math Chem 2001, 29, 127.
5. Li, G.; Wang, S. W.; Rabitz, H. J Phys Chem 2002, 106, 8721.
6. Li, G.; Wang, S. W.; Rabitz, H. J Math Chem 2001, 30, 1.
7. Gordon, W. J. Proc. Symp. Approximation with Special Emphasis on Spline Functions; Schoenberg, I. J., Ed., Academic Press: New York, 1969; p 223.
8. Kalos, M. H.; Whitlock, P. A. Monte Carlo Methods; John Wiley & Sons: New York, 1986; p 31.
9. Li, G.; Wang, S. W.; Rabitz, H. J Comput Chem, to appear.
10. Wang, S. W.; Levy, II, H.; Li, G.; Rabitz, H. J Geophys Res 1999, 104, D23, 30417.