Metamodeling Strategies for High-dimensional Simulation-based Design Problems

By

Songqing Shan

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical & Manufacturing Engineering

University of Manitoba

Winnipeg

Abstract

Computational tools such as finite element analysis and simulation are commonly used for system performance analysis and validation. It is often impractical to rely exclusively on the high-fidelity simulation model for design activities because of high computational costs. Mathematical models are typically constructed to approximate the simulation model to help with the design activities. Such models are referred to as "metamodel." The process of constructing a metamodel is called "metamodeling."

Metamodeling, however, faces eminent challenges that arise from high-dimensionality of underlying problems, in addition to the high computational costs and unknown function properties (that is black-box functions) of analysis/simulation. The combination of these three challenges defines the so-called high-dimensional, computationally-expensive, and black-box (HEB) problems. Currently there is a lack of practical methods to deal with HEB problems.

This dissertation, by means of surveying existing techniques, has found that the major deficiency of the current metamodeling approaches lies in the separation of the metamodeling from the properties of underlying functions. The survey has also identified two promising approaches - mapping and decomposition - for solving HEB problems. A new analytic methodology, radial basis function–high-dimensional model representation (RBF-HDMR), has been proposed to model the HEB problems. The RBF-HDMR decomposes the effects of variables or variable sets on system outputs. The RBF-HDMR, as compared with other metamodels, has three distinct advantages: 1) fundamentally reduces the number of calls to the expensive simulation in order to build a metamodel, thus breaks/alleviates exponentially-increasing computational difficulty; 2) reveals the

functional form of the black-box function; and 3) discloses the intrinsic characteristics (for instance, linearity/nonlinearity) of the black-box function.

The RBF-HDMR has been intensively tested with mathematical and practical problems chosen from the literature. This methodology has also successfully applied to the power transfer capability analysis of Manitoba-Ontario Electrical Interconnections with 50 variables. The test results demonstrate that the RBF-HDMR is a powerful tool to model large-scale simulation-based engineering problems. The RBF-HDMR model and its constructing approach, therefore, represent a breakthrough in modeling HEB problems and make it possible to optimize high-dimensional simulation-based design problems.

Acknowledgements

Table of Contents

V

List of Tables

List of Figures

List of Copyrighted Material

[1]   Shan, S. and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-dimensional Design Problems with Computationally-Expensive Black-box Functions," *Structural and Multidisciplinary Optimization*, **41**(2(2010)), pp. 219-241.

[2]   Shan, S. and Wang, G. G., 2010, "Metamodeling for High-dimensional Simulation-Based Design Problems," *Journal of Mechanical Design*, **132**(5), pp. 051009-1-051009-11.

[3]   Shan, S. and Wang, G. G., 2010, "Turning Black-box into White Functions," *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Montreal, Quebec, Canada. August 15-18, 2010. Paper No., DETC2010-28958.

Chapter 1

Preamble

1.1. About Metamodeling

A system such as an automotive component, an entire vehicle, or a manufacturing process is typically modeled via computer simulation/analysis such as finite element analysis (FEA), and computational fluid dynamics (CFD). The computer simulation/analysis process is usually implicit and time-consuming for execution, which is therefore also called a computationally-expensive black-box function. Furthermore, outputs from these analysis/simulation processes are usually not directly useful for design, and their computation intensity makes optimization formidable. Despite the growth of computer power, the complexity of simulation models keeps increasing at a commensurate speed. To facilitate design analysis and optimization, metamodels are often constructed. Such a metamodel (also known as surrogate or response surface, or auxiliary model) is an approximation model of an underlying system implicitly defined by the given simulation model of the system. Usually it is assumed that the simulations and analysis processes are black-box functions, that is, the underlying system is completely unknown. A metamodel is constructed from a well-planned sampling scheme (that is, design of computer experiments) in the simulation input space. After validation, the metamodel serves in analyzing and/or optimizing the underlying system. The process of constructing metamodels is called metamodeling. The metamodel provides an effective mechanism for simplifying the interpretation of simulated results. The simplified interpretation helps engineers to gain insight into the underlying system. The fast

execution of the metamodel alleviates computation burden for sensitivity analysis and optimization of problems.

1.2. HEB Problems

Metamodeling techniques play an important role for analysis and optimization of computationally-expensive functions. The importance has been demonstrated by many successful applications in the literature; however, most of successful examples are low-dimensional problems. The high-dimensionality of problems is challenging because the cost of metamodeling rapidly rises with the increase of dimensionality. High-fidelity simulation models for complex design problems are often <u>h</u>igh-dimensional, computationally <u>e</u>xpensive, and <u>b</u>lack-box functions, which are called HEB problems. HEB problems exist in various engineering disciplines, but HEB problems are not well solved due to their challenges. In engineering design, HEB problems become a bottleneck for the wide application of metamodeling techniques. This dissertation aims at solving this bottleneck problem.

1.3. Threads of Dissertation

This dissertation is presented as a sandwich thesis type that contains the full text of four papers. These papers either have already been published or have been submitted for publication. This section threads them in a thesis manner. This manner in fact reflects the achievements at various stages which are logically sequenced. All four papers focus on the research theme: solving HEB problems. The contents include the survey of the state of the arts in HEB related techniques, a proposed Radial Basis Function–High-dimensional Model Representation (RBF-HDMR) model and its matching sampling

scheme, enhancement of the RBF-HDMR, and industrial application of the RBF-HDMR. They form a logically integrated work on HEB metamodeling strategy. More details are as follows.

Chapter 2 surveys strategies for tackling high-dimensionality, metamodeling approximation techniques, and optimization methodologies pertaining to black-box function problems aiming at solving HEBs. Features of various strategies, techniques, and methodologies have been discussed, which deal with a certain aspect or individual challenges respectively. These techniques and strategies cannot be easily extended or crossbred for HEB problems. Promising approaches for HEB problems are identified in the survey. The results of the survey indicate that HEB problems can be solved if a decomposable high-dimensional metamodel with low construction cost is available. This leads to the development of RBF-HDMR (Chapter 3).

Chapter 3 develops a RBF-HDMR model based on High-dimensional Model Representation (HDMR) theory. Metamodeling a high-dimensional computationally-expensive black-box function is a challenging job. Existing metamodels are designed for low-dimensional problems only. Thanks to the theory of HDMR: an integral function can be decomposed into summation of different dimensional sub-functions [1], and statistical data demonstrates that there is no high-dimensional covariance existing in most well-defined problems [2]. In this chapter, we integrate Radial Basis Function with HDMR into a new model, RBF-HDMR. An accompanying algorithm to construct the RBF-HDMR until the second-order has also been developed. This algorithm is suitable for problems with weak correlations among variables.

Chapter 4 enhances the RBF-HDMR to solve problems with high-order correlated variables and furthermore reveals the complete functional form of metamodels. The enhanced RBF-HDMR more accurately constructs black-box functions with higher-order terms. The model and the algorithm fundamentally change the exponentially growing computation difficulty to be polynomial. Moreover, with the assistance of two developed theorems and two defined matrixes, the enhanced model can explicitly express the functional form of the black-box function in a rapid manner, and thus turn the "black-box" into "white box." Testing and comparison confirm the efficiency and capability of the RBF-HDMR for HEB problems.

Chapter 5 applies the proposed RBF-HDMR to solve a practical engineering problem with 50 variables. Three application cases of power transfer capability analysis in Manitoba Hydro are studied and modeled. With a limited number of function calls, RBF-HDMR is able to model the systems and reveals interesting characteristics of the power system under different operating conditions.

Chapter 6 summarizes the work of the preceding chapters, concludes the contributions, and recommends future research directions.

1.4. A Description of Commonalities and Connecting Concepts

The four papers are all centered on solving HEB problems. Chapter 2 is the survey for high-dimensional strategies, which reveals the possible ways to solve HEB problems. Chapter 3 describes in details the proposed RBF-HDMR strategy. Chapter 4 extends the RBF-HDMR from the second-order to higher-order models. With new theorems developed in Chapter 4, the functional form of the underlying HEB problem can also be

revealed. So Chapter 4 is a natural extension and enhancement of the proposed method described in Chapter 3. Chapter 5 documents in detail the application of RBF-HDMR to a practical industry HEB problem with 50 variables. Therefore, all of the four main chapters are intrinsically and logically connected, representing the progress of the Ph.D. research. The main connecting concepts include HEB and RBF-HDMR.

1.5. Terms in Dissertation

In metamodeling community, there are some different terms for the same concept, for example, metamodel, response surface, and surrogate are often used interchangeably. This section will introduce some terms in this dissertation.

In this work, the expensive simulation/analysis process that is to be modeled or approximated is also referred to as "underlying function," "underlying system," "underlying problem," or "black-box function" in the context of metamodeling. "Sampling" and "design of computer experiments" have the same meaning. We distinguish, however, the design of computer experiments from classical design of experiments. In this dissertation, we only refer to the design of computer experiments.

Simulations appear in various types (for example, stochastic or deterministic, continuous or discrete, and static and dynamic). This dissertation takes consideration of the deterministic computer simulation, that is, negligible random errors appear. In addition, simulations are considered as an ideal underlying system, that is, no noise exists in the simulation model.

"Model" in Chapter 2 shares the same meaning of "metamodel".

1.6. Student's Contribution

After identifying the research direction with the co-author (my supervisor), I developed the RBF-HDMR model, implemented the modeling algorithm, applied the developed RBF-HDMR to industrial cases, drafted the papers, and revised them with my supervisor and one other co-author for Chapter 5.

Chapter 2

Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-box Functions[1]

Songqing Shan   G. Gary Wang

2.1. Abstract

The integration of optimization methodologies with computational analyses/simulations has a profound impact on the product design. Such integration, however, faces multiple challenges. The most eminent challenges arise from high-dimensionality of problems, computationally-expensive analysis/simulation, and unknown function properties (that is, black-box functions). The combination of these three challenges severely aggravates the difficulty and becomes a major hurdle for design optimization. This chapter provides a survey on related modeling and optimization strategies that may help to solve High-dimensional, Expensive (computationally), Black-box (HEB) problems. The survey screens out 207 references including multiple historical reviews on relevant subjects from more than 1000 papers in a variety of disciplines. This survey has been performed in three areas: (1) strategies for tackling high-dimensionality of problems, (2) model approximation techniques, and (3) direct optimization strategies for computationally-expensive black-box functions and promising ideas behind non-gradient optimization algorithms. Major contributions in each area are discussed and presented in an organized manner. The survey exposes that direct modeling and optimization strategies to address HEB problems are scarce and sporadic, partially due to the difficulty of the problem itself. Moreover, it is revealed that current modeling research tends to focus on sampling and modeling techniques themselves and neglect studying and taking the advantages of characteristics of the underlying expensive functions. Based on the survey results, two promising approaches are identified to solve HEB problems. Directions for future research are also discussed.

Keywords: high-dimensional, computationally-expensive, black-box function, approximation, design optimization, large-scale, metamodeling, surrogate

2.2. Introduction

Engineering problems often appear with various features such as being low or high dimensional, computationally cheap or expensive, and with explicit or black-box functions (a black-box function is an unknown function that given a list of inputs, corresponding outputs can be obtained without knowing its expression or internal structure). These features characterize a problem from different perspectives. Combinations of these features lead to different computational costs for problem solution. For example, the computational cost for optimizing a cheap black-box function is largely from the optimization process, while for computationally-expensive functions the computational cost is mainly from the function evaluation rather than optimization. Therefore, solution methodologies need to be custom developed for problems of different combinations of these features. This review focuses on design problems that are comprised of high-dimensional, expensive (computationally), and black-box (HEB) functions.

HEB problems widely exist in science and engineering practices (Bates *et al.* 1996; Booker *et al.* 1999; Koch *et al.* 1999; Shorter *et al.* 1999; Srivastava *et al.* 2004; Tu and Jones 2003). For example, the wing configuration design of a high speed civil transport (HSCT) aircraft (Koch *et al.* 1999) includes 26 variables, four objectives (two traditional technical and two economic objectives), and four technical constraints. The NASA synthesis tool FLOPS/ENGGEN was used to size the aircraft and propulsion system. The

NASA aircraft economic analysis code ALCCA was applied to perform economic uncertainty analysis of the system. These computer codes often are regarded as black-box functions. Each execution of FLOPS/ENGGEN and ALCCA requires approximate 5 minutes on an IBM RISC6000 7012 model 320 Planar workstation. If a two-level full-factorial analysis is taken, 67,108,864 analyses are required, which would take over 600 years to complete. In automotive industry, the crashworthiness analysis takes on average 98 hrs for one evaluation (Gu 2001). Assuming ten variables with a two-level full-factorial design, it needs 1024 analyses and takes close to 12 years to complete.

The high-dimensionality of input and output variables presents an exponential difficulty (that is, the effort grows exponentially with dimensions) for both problem modeling and optimization (Koch *et al.* 1999; Li *et al.* 2001b; Shorter *et al.* 1999). Assuming sampling *s* points in each of the *n* input variables and performing the computer simulation or experiments, this sampling calls for $\sim s^n$ experimental or computer runs to build a model, which would obviously be unrealistic for modeling of computationally-expensive functions (for instance, if *s*=10 and *n*=10, then the number of sample points is 10^{10}). Modern analysis models are often built in commercial software tools, such as Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD) tools. Besides being computationally intensive, these models (functions) are implicit and unknown to the designer, that is, black-box functions. The function implicity is a significant obstacle to design optimization (Alexandrov *et al.* 2002). As the number of variables in design problems increases, the computational demand also increases exponentially (Michelena *et al.* 1995; Michelena and Papalambros 1995b; Michelena and Papalambros 1997; Papalambros 1995; Papalambros and Michelena 1997, 2000). This kind of difficulty

10

brought by the dimensionality of problems is known as the "curse of dimensionality." Mistree's research group referred to this difficulty as the "problem size" in robust design (Chen *et al.* 1996; Koch *et al.* 1997) and multidisciplinary design optimization (Koch *et al.* 1999). The "curse of dimensionality" challenges computational analysis technologies and optimization methodologies that are used today in science and engineering disciplines.

It is observed that in the area of engineering design there are limited publications that directly address HEB problems. In general, both modeling techniques and optimization methods for computationally-expensive or black-box function are limited to problems of low dimensionality. Problems with high-dimensionality are more demanding. This chapter provides a survey of the modeling and optimization strategies that may help solving HEB problems in order to guide future research on this important topic. The survey has been performed along three routes: 1) strategies for tackling high-dimensionality in disciplines including mathematics, statistics, chemistry, physics, computer science, and various engineering disciplines, 2) model approximation techniques, which are strategies for computationally-expensive black-box functions, and 3) direct optimization strategies for computationally-expensive black-box problems, and promising ideas behind commonly used non-gradient optimization algorithms that may be helpful to solve HEB problems.

2.3. Strategies for Tackling High-Dimensionality

A spectrum of strategies for tackling high-dimensionality appears in many different disciplines since the high-dimensionality challenge is rather universal in science and

engineering fields. These strategies include parallel computing, increasing computer power, reducing design space, screening significant variables, decomposing design problems into sub-problems, mapping, and visualizing the variable/design space. These strategies tackle from different angles the difficulties caused by the high-dimensionality. Some of them may overlap and are thus not completely independent. In view of the space limit and the fact that some of strategies are studied in special areas (for example, parallel computing and increasing computer power), this section only reviews some of them that directly deal with high-dimensionality.

2.3.1. Decomposition

Decomposition is to reformulate an original problem into a set of independent or coordinated sub-problems of smaller scale. Decomposition methodology has been well studied and widely applied to complex engineering problems (Altus *et al.* 1996; Chen *et al.* 2005b; Kusiak and Wang 1993; Michelena *et al.* 1995; Michelena and Papalambros 1995b). Some reviews pertaining to the decomposition can be found in the literature (Browning 2001; Li 2008; Papalambros 1995; Papalambros and Michelena 1997, 2000). A technical map of decomposition methodology is provided in Fig. 2.1. The review is organized according to this map.

In engineering, decomposition reported in the literature can be categorized into product decomposition, process decomposition, and problem decomposition (Kusiak and Larson 1995). The product decomposition partitions a product into physical components. The application examples of product decomposition are given in (Kusiak and Larson 1995). Such decomposition allows standardization, inter-changeability, or a capture of the

12
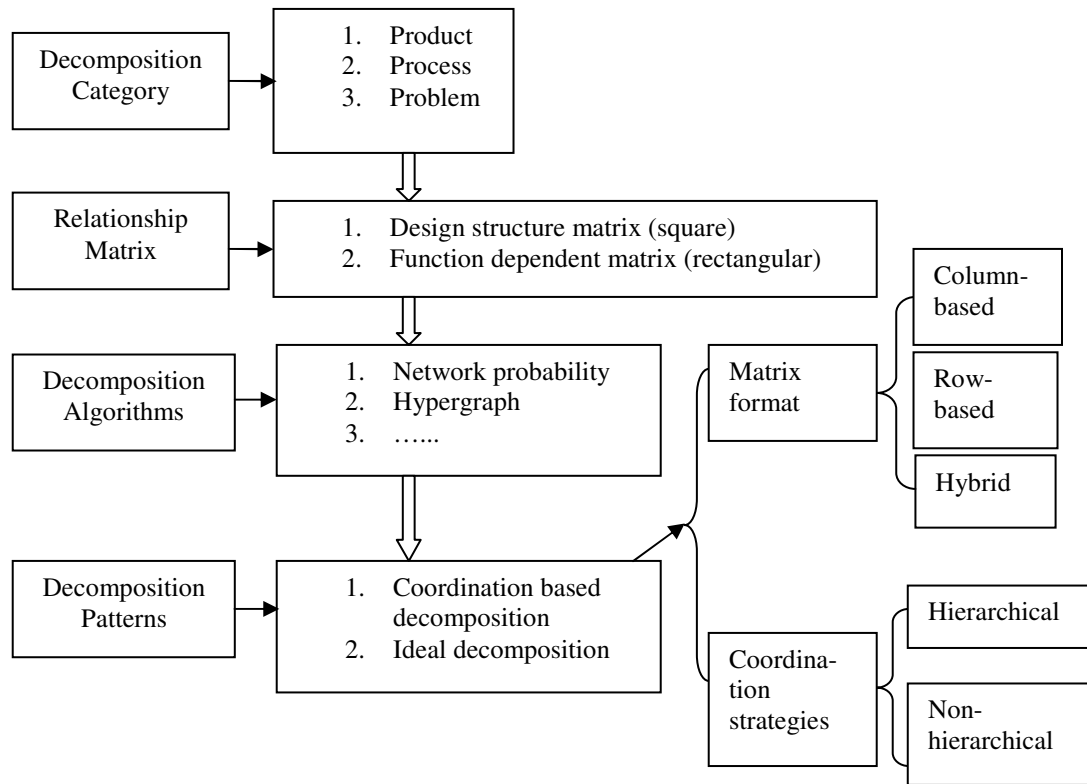
Fig. 2.1 An illustration of decomposition methodologies

product structure. Its drawback is that drawing "boundaries" around physical components is subjective. Secondly, the process decomposition applies to problems involving the flow of elements or information, such as electrical networks or the design process itself. Applications are found in (Kusiak and Wang 1993; Michelena *et al.* 1995). Thirdly, the problem decomposition divides a complex problem into different sub-problems. Such decomposition is the basis of multidisciplinary design optimization and decomposition-based design optimization. Intensive research has been done on multidisciplinary design optimization (Kodiyalam and Sobieszczanski-Sobieski 2000; Simpson *et al.* 2004) and applied in industry (Sobieszczanski-Sobieski and Haftka 1997). Decomposition-based design optimization (Michelena and Papalambros 1995b; Michelena and Papalambros

1997) advances the use of nonlinear optimization techniques in solving design problems. Such design optimization (for instance, model-based decomposition) allows the identification of weakly connected model substructures and obtains robust solutions.

Matrix is often exploited to reflect relationship in problems, which is called relationship matrix. Thus by means of partitioning the relationship matrix a problem is decomposed. Although various terms are utilized in the literature such as dependency structure matrix, interaction matrix, incidence matrix, function dependent table, and precedence matrix, there are two basic relationship matrices: design structure matrix (DSM) and function dependent matrix (FDM). DSM is a square matrix that has identical row and column listings to represent a single set of objects (Browning 2001; Li 2008). A matrix entry indicates whether (or how or to what degree that) the i-th object (row) relates to the j-th object (column). DSM captures symmetric or non-symmetric, directional or undirected relationships between any two objects of the same type. On the other hand, FDM has different row and column listings to represent two sets of objects, respectively. A matrix element indicates whether (or how or to what degree that) the i-th row object relates to the j-th column object and vice versa. FDM captures dependency relationships between two types of objects such as function dependent tables in (Krishnamachari and Papalambros 1997a, 1997b; Wagner and Papalambros 1993).

Matrix partitioning is often formed by means of mathematical tools such as graph partitioning, clustering analysis, and optimization. Thus, algorithms for matrix partitioning or decomposition are dispersed. Normally these algorithms depend on how the decomposition is modeled. They fall into three major types. The first type of algorithms models decomposition as a hyper-graph (Michelena and Papalambros 1997),

14

network reliability (Michelena and Papalambros 1995a), or an integer programming problem (Krishnamachari and Papalambros 1997b). The second type of algorithms is heuristic approaches such as (Wagner and Papalambros 1993). The third type of algorithms is clustering approaches such as (Chen *et al.* 2005a). For DSM, Browning (2001) found that mostly clustering and sequencing algorithms are used. The clustering algorithms are to reveal the architecture relationship; the sequencing algorithms are to expose the information flow relationship. For FDM, clustering algorithms are useful for design optimization and group technology. In the context of group technology, machine-part groups are formed to increase production efficiency. In the context of design optimization, function-variable groups are formed to dissolve the complexity of problems. Their common goal is to reveal independent groups (or sub-problems) in a complex problem.

Decomposition patterns exist in two types (Chen *et al.* 2005a): ideal and coordination-based decomposition. The ideal decomposition diagonalizes a relationship matrix into several completely independent blocks without any interactions between the blocks (that is, no variable belongs to two blocks). If a design strictly follows the axiomatic design theory (Suh 2001), the ideal decomposition can be obtained. The coordination-based decomposition is a more realistic decomposition pattern with interactions between the blocks. In terms of matrix format, there are column-based, row-based, and hybrid structured matrices (Chen *et al.* 2005a). Accordingly, some of column variables, row variables, or both column and row variables are taken as coordination variables. From the nature of coordination, decomposition patterns are categorized as hierarchical or non-hierarchical (Chen and Liu 1999; Krishnamachari and Papalambros

1997a; Michelena *et al.* 1999; Michelena and Papalambros 1997; Papalambros 1995; Papalambros and Michelena 1997; Wagner and Papalambros 1993). Coordination processes are to coordinate linking variables (connecting sub-problems and master problems or sub-problems and sub-problems) in order to find the optimal solution. Hierarchical decomposition is characterized by a tree structure (Renaud and Gabriele 1991) whereas non-hierarchical decomposition is characterized by a network structure (Renaud 1993; Renaud and Gabriele 1991). In hierarchical decomposition, the intrinsic hierarchical structure can be used by many optimization algorithms and thus each sub-problem can be of a smaller scale. Hierarchical decomposition schemes, however, are hard to use when lateral couplings exist between sub-problems of the hierarchy since the lateral couplings interfere with the hierarchical solution process. In non-hierarchical decomposition, likely more couplings appear because of the lack of hierarchy. Complex couplings bring a great challenge to optimization algorithms as decoupling is needed. A hybrid method combing hierarchical decomposition in the entire system and non-hierarchical decomposition in the local area (subsystems with lateral couplings) is likely useful for problems with lateral couplings.

Decomposition was recognized as a powerful tool for analysis of large and complex problems (Krishnamachari and Papalambros 1997b; Kusiak and Wang 1993). For rigorous mathematical programming, decomposing an overall model into smaller sub-models was considered as necessary by (Papalambros 1995). Complexity of design problems in the context of decomposition is analyzed in (Chen and Li 2005). The idea of decomposition penetrates in conceptual design (Kusiak and Szczerbicki 1992), optimal system design (Kim *et al.* 2003), concurrent design, complex problem modeling, etc.

16

Decomposition often accompanies with parallel approaches to enhance the efficiency. Koch *et al.* (2000) proposed an approach to build partitioned, multi-level response surfaces for modeling complex systems. This approach partitions a response surface model to two quadratic surrogates; one surrogate is constructed first and becomes a term in the other surrogate to form a two-level metamodeling process. Kokkolaras *et al.* (2006) presented a methodology for design optimization of hierarchically decomposed multilevel systems under uncertainty. Chan *et al.* (2000) designed and implemented a new class of fast and highly scalable placement algorithms that directly handled complex constraints and achieved the optimum through the use of multilevel methods for hierarchical computation. Lu and Tcheng (1991) proposed a layered-model approach. The references (Pérez *et al.* 2002a; Wang and Ersoy 2005; Ye and Kalyanaraman 2003) applied parallelization in their optimization algorithms. Eldred *et al.* (Eldred *et al.* 2004; 2000) combined a multilevel idea with parallelization to implement optimization. These methods decompose a complex optimization problem and form cascading schemes that can be implemented by multilevel or parallel approaches. Decomposition brings many advantages: improved coordination and communication between sub-problems, allowing for conceptual simplification of the problems, different solution techniques for individual sub-problems, reduced sub-problem dimensionality, reduced programming/debugging effort, modularity in parametric studies, multi-criteria analysis with single/multiple decision makers, and enhancing the reliability and robustness of optimization solutions (Michelena and Papalambros 1995b; Michelena and Papalambros 1997).

As concluding remarks, the decomposition methodology is an effective strategy for solving complex design optimization problems. Decomposition concepts are expected to advance for modeling and optimization of HEB problems.

2.3.2. Screening

Screening identifies and retains important input variables and interaction terms, whereas removes less important ones or noises in the problems of interest so that the complexity or dimensionality of the problems is reduced to save computational cost. Screening is often implemented via sampling and analysis of sampling results. Screening approaches are grouped as two categories as shown in Fig. 2.2. One category deals with a single response and the other deals with multiple responses.
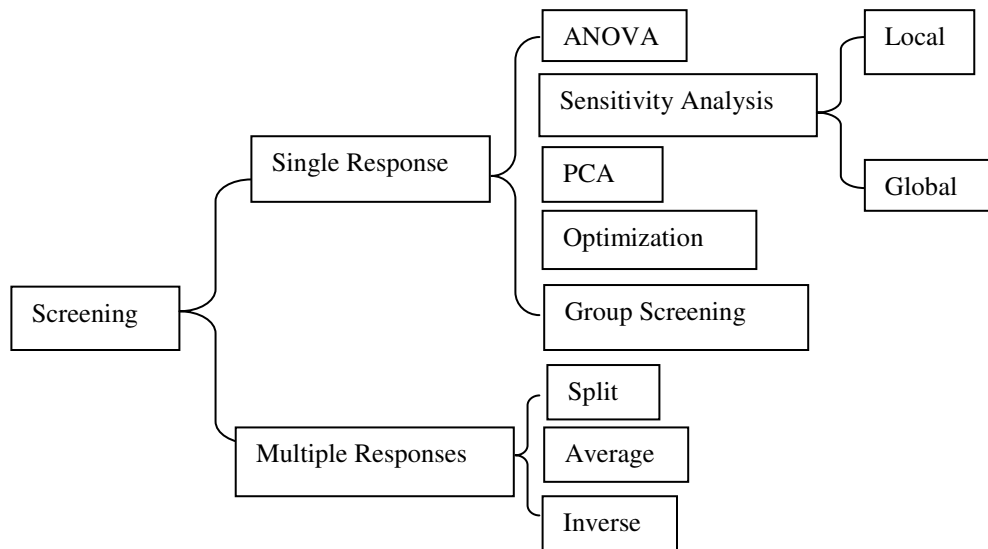
Fig. 2.2 Screening approaches

Screening for a single response is to select the most important variables or interaction terms of variables as to the response. The importance of variables or their interaction terms is judged by means of sensitivity analysis, analysis of variances

(ANOVA), principle component analysis (PCA), optimization approaches, and group screening after experiments. Some of these approaches are correlated, for example, sensitivity analysis is implemented by ANOVA. Sensitivity analysis studies how the variability of a function's output responds to changes of its inputs. It includes local and global sensitivity analyses. The local sensitivity indicates the local variability of the output with respect to input variable changes at a given point, which are partial derivatives. It restricts to infinitesimal changes in input variables. The global sensitivity, however, explains the global variability of the output over the entire ranges of the input variables, which provides an overall view of the impact of input variables on the output. It considers more substantial changes in input variables. If a probabilistic setting is considered with both inputs and outputs, sensitivity analysis is referred to as probabilistic sensitivity analysis (Oakley and O'Hagan 2004). Sensitivity analysis has been widely studied (Morris 1991; Sobol 1993; Jin *et al.* 2004; Kaya *et al.* 2004). Griensven (2006) and Queipo *et al.* (2005) introduced different techniques in sensitivity analysis. Harada *et al.* (2006) screened parameters of pulmonary and cardiovascular integrated model with sensitivity analysis. Iman and Conover (1980) utilized the sensitivity analysis approach in the modeling with application to risk assessment. Wagner (2007) applied global sensitivity analysis of predictor models in software engineering. Sobieszczanski-Sobieski (1990) discusses sensitivity analysis for aircraft design. Hamby (1994) reviewed the techniques for sensitivity analysis of environmental models. By means of analysis of variance (ANOVA) (Myers and Montgomery 1995), the main effect of a single variable or correlated effect of multiple variables can be identified. Schonlau and Welch (2006) introduced the ANOVA decomposition (functional ANOVA) theory and developed the

19

steps for identifying and visualizing the important estimated effects. Principal Component Analysis (PCA) transforms data to a new coordinate system by data projection so that variables with greatest variances in the projection come to the principal coordinates. The selection of dimensions using PCA through singular value decomposition is a popular approach for numerical variables (Ding *et al.* 2002). Welch *et al.* (1992) proposed a sequential optimization algorithm for screening. Watson (1961) proposed a group screening method. Morris (1991) designed factorial sampling plans for preliminary experiments. Tu and Jones (2003) proposed a cross-validated moving least squares (CVMLS) method, which integrated the variable screening into a metamodeling process. It screens input variables by two ways: a main effects estimate procedure using one-dimensional CVMLS analysis to eliminate insignificant inputs; and a backwards-screening procedure for calculating cross-validation error sensitivities of input variables. Shen *et al.* (2006) developed an adaptive multi-level Mahalanobis-based dimensionality reduction (MMDR) algorithm for high-dimensional indexing. The MMDR algorithm uses the Mahalanobis distance and consists of two major steps: ellipsoid generation and dimensionality optimization. Brand (2003) proposed a dimensionality reduction method by kernel eigenmaps. Ding (2002) proposed an adaptive dimension reduction approach by clustering high-dimensional data.

Screening strategies for multiple responses are different from that for a single response since the importance of variables or interaction terms varies for different responses. Strategies for a single response, however, may be used for the case of multiple responses. One method for multiple responses is to screen each response separately and select important variables or terms for each response, which is called the split method.

The split method bears two disadvantages: the screening process time increases as the number of the responses increases and the approximation response may not be consistent when some variables are fixed for another response. The average method exploits the average effects of variables across all of the responses and selects the variables or terms which have average efforts on all responses. Such a method possibly eliminates variables that are extremely important for one response. Chen *et al.* (1996) employed this approach to reduce the problem size. An inverse screening approach (Koch *et al.* 1999) identifies variables that are not important for any of the responses. This approach is accomplished by combining sets of important variables for each response and observing which variables are not included in the combined set. A two-level fractional factorial experiment is designed for screening and Pareto analysis is used to analyze the experimental results to rank the importance of variables for each response. Like screening for a single response, the problems exist on deciding a cutoff criterion and the possible loss of accuracy. Since the cutoff point of importance is subjective, it is hard to make the trade-off between the acceptable accuracy and completeness in problem formulation.

In general, screening likely pays a price of losing modeling accuracy of problems because of removed dimensionalities. As the number of variables increases, the dimensionality of the remaining problem after screening may still be high for some existing models. Screening over multiple responses inherently may not allow many variables to be removed from problems. A design with fewer runs, or with fewer levels of each input variable, may well have missed the important regions (Schonlau and Welch 2006). Advantages of screening include noises reduction, removal of unimportant variables or terms, and retaining of important variables in problems of interest, which

decreases complexities and reduces dimensionality. The use of screening depends on the purposes and experimental type. It is identified to be a good strategy for filtering noises in the physical experiments and supporting modeling. It can guide modeling and simplify the computer model. For the purpose of optimization, although it simplifies the problem, it pays the price of accuracy. The screening strategies therefore should be employed with care.

2.3.3. Mapping

Mapping has a broad sense including projection, non-linear mapping, parameter space transformation, and so on. In this section, mapping techniques are categorized into two groups: mapping aiming at dimensionality reduction and mapping aiming at optimization.

Mapping aiming at dimensionality reduction transforms a set of correlated variables into a smaller set of new uncorrelated variables that retain most of the original information. This includes non-linear mapping and projection. Projection has multiple algorithms such as projections by principal component analysis (PCA) (Dunteman 1989; Penha and Hines 2001; Shlens 2005), analysis of variance (ANOVA), and relative distance plane (RDP) mapping (Somorjai *et al.* 2004). RDP maps high-dimensional data onto a special two-dimensional coordinate system, the relative distance plane. This mapping preserves exactly the original distance between two points with respect to any two reference patterns in RDP. Besides dimensionality reduction, projection approaches are used for data classification, data clustering, and visualization of high-dimensional problems as well. Non-linear mapping is a commonly used method for easing problem

complexity. Artificial Neural Network (ANN) embodies non-linear mapping techniques. Rassokhin *et al.* (2000) employed fuzzy clustering and neural networks for nonlinear mapping of massive data sets. Sammon (1969) proposed an algorithm of nonlinear mapping for data structure analysis. This algorithm was based on point mapping of a higher-dimensional space to a lower-dimensional space such that the inherent data "structure" was approximately preserved. Saha *et al.* (1993) applied linear transformation inducing intrinsic dimension reduction. Kaski (1998) reduced dimensionality by random mapping. All above mapping techniques successfully implemented the dimensionality reduction.

Bandler *et al.* (1994) proposed a space-mapping (SM) technique aiming at optimization. This space-mapping technique made use of two models for the same system: a "coarse" model, and a "fine" model. The "coarse" model could be an empirical equation, simplified theoretical model or finite element model. These "coarse" models were less accurate and computationally inexpensive. The "fine" model could be a high precision component model or fine finite element model. These "fine" models were more accurate and computationally expensive. A mathematical mapping between the spaces of parameters of two different models was established, which maps the fine model parameter space to the coarse model parameter space such that the responses of the coarse model adjust for the responses of the fine model within some local modeling region around the optimal coarse model solution. In conjunction with the accuracy of the "fine" model and the cheap computation of the "coarse" model, an optimization algorithm was implemented. In the context of this space mapping technique, the parameter extraction (obtaining the parameters of the coarse model whose responses

match the fine model responses) was crucial since the non-uniqueness of the extracted parameters may cause the technique to diverge. Some algorithms such as Aggressive Space Mapping (ASM) (Bandler *et al.* 1995a, 1995b, Bakr *et al.* 1999a), Trust Region Aggressive Space Mapping (TRASM) (Bakr *et al.* 1998), Hybrid Aggressive Space Mapping (HASM) (Bakr *et al.* 1999b) methods were developed to obtain better parameter extraction by the same research group of the original space mapping technique. This space-mapping was then applied to optimization of microwave circuits (Bakr *et al.* 2000a) by the same researchers. Leary *et al.* (2001) developed a constraint mapping to structural optimization. Bakr *et al.* (2000b) reviewed these space mapping techniques and discussed developments in Space Mapping-based Modeling (SMM) including Space Derivative Mapping (SDM), Generalized Space Mapping (GSM), and Space Mapping-based Neuromodeling (SMN). Bandler *et al.* (2004) refreshed the state of the art of the space-mapping techniques.

The first group of mapping approaches relaxes the "curse of dimensionality" of problems for modeling, and the second eases the complexity of optimization problems. But it seems that no one has examined the possibility of mapping optimization problems from an original higher-dimensional space to a new lower-dimensional space while preserving the optimum. If this is doable, both the problem size and the optimization complexity can be reduced simultaneously. The challenge is how to ensure the optimum obtained in the lower-dimensional space is the true optimum for the higher-dimensional space.

2.3.4. Design Space Reduction

When modeling and optimizing a practical problem, ranges of design variables need to be determined. Combination of variable ranges defines the design space. In this chapter, space reduction is limited to the reduction of ranges of design variables excluding the reduction of the number of variables (discussed in screening and mapping). Space reduction means shrinking a design space so that modeling is more accurate in the modeling range or optimization effort is reduced in the optimization domain. A common space reduction approach starts with sampling a limited number of points and evaluating function values at these points. Then the design space is reduced based on feedback information from modeling on these sample points. The revised design space is again segmented using smaller increments, and the objective function is determined for new points. In this way, the focus of modeling can be in a more attractive region, which leads to more effective models. Approximated or inexpensive constraints are often employed to eliminate some portions of the design space. In the optimization formulation phase, the design space can be explored to obtain a deeper insight into the design problem, and thus the optimization focus can be made on the most interested sub-spaces that contain the optimum with high probability in the design space. Wang *et al.* developed a number of methods such as the adaptive response surface method (ARSM) (Wang *et al.* 2001), and the fuzzy clustering based approach (Wang and Simpson 2004), in which the design space is iteratively reduced. Shan and Wang then proposed a rough set based method which could systematically identify attractive regions (sub-spaces) from the original design space for both single and multiple objectives (Shan and Wang 2004, Wang and Shan 2004). Engineers could pick satisfying design solutions from these regions or

continue to search in those regions. In the optimization processes, there are some strategies to contract the design space. Shin and Grandhi (2001) reduced the space using the interval method. This method began with a box in which the global optimum was sought; it first divided the box and found the interval of the objective function and each constraint in each sub-box, and deleted the sub-boxes which could not contain the optimum. This process continued until the box size became sufficiently small.   Marin and Gonzalez (2003) solved the path synthesis optimization problems using design space reduction. The design space reduction was implemented in two ways: one eliminating redundant design points by defining some prerequisites and the other eliminating poor design points. Yoshimura and Izui (1998) implemented mechanism optimization via expansion and contraction of design spaces. Ahn and Chung (2002) utilized joint space reduction and expansion to redundant manipulator optimization. The space reduction and expansion is commonly employed as a strategy of optimization and done by moving limits of design variables. Move-limit optimization strategies (Fadel and Cimtalay 1993; Fadel *et al.* 1990; Grignon and Fadel 1994; Wujek and Renaud 1998a, 1998b) applied the conjunction of approximation with move limit concepts to optimization problems. Trust region based algorithms (Byrd *et al.* 1987; Celis *et al.* 1984; Rodríguez *et al.* 1998) made use of the idea of changing spaces. These approaches varied the bounds of design variables in optimization iterations and differed from each other in bound adjustment strategies. Space reduction strategies can be used in optimization problem formulation phases, optimization processes, and modeling processes.

2.3.5. **Visualization**

The idea of visualization is to present a problem in a visual form, allowing users to get insight into the problems, find key trends and relationships among variables in a problem, and make decisions by interacting with the data. There are various techniques for multidimensional data visualization including graph morphing, panel matrix displays, iconic displays, parallel coordinates, dense pixel displays, and stacked displays. Stump *et al.* (2002) listed advantages and disadvantages of scatter matrix/brushing and data-driven placement of Glyphs and developed an interface incorporating visualization techniques. Winer and Bloebaum (2002a; 2002b) developed a Visual Design Steering (VDS) method as an aid in multidisciplinary design optimization. VDS allows a designer to make decisions before, during, or after an analysis or optimization via a visual environment to effectively steer the solution process. Many companies are utilizing the power of visualization tools and techniques to enhance product development and support optimization (Simpson 2004). Visualization is helpful when little is known about the data and the exploration goals are implicit since users are able to directly participate in the exploration processes, shift and adjust the exploration goals if necessary. The visualization can aid in black-box function modeling. VDS for high-dimensional optimization problems, however, need to be developed.

2.3.6. **Summary Remarks**

Five main strategies for tackling high-dimensionality are reviewed. Their pros and cons are summarized in Table 2.1. Among these methods, decomposition methodology is identified as the most promising tool for high-dimensional problems, given its general

Table 2.1 Summary of strategies tacking problems of high-dimensionality

Strategy	Advantages	Disadvantages	Application
Decomposition	Reduced sub-problem dimensionality; reduced programming/debugging effort; simpler and more efficient computational procedures (such as parallel/distributed computation, concurrency, modularity); improved coordination and communication between the decomposed sub-problems; enabling different solution techniques to individual sub-problems; support of multi-criteria analysis with single/multiple decision makers; enhanced reliability and robustness of optimization solutions	Limited by decomposability	Modeling and optimization for high-dimensional or large scale problems
Screening	Removal of noises and insignificant variables and terms; distinguish the interactions in problems	May sacrifice accuracy; limited by nature of problems	Problem investigation and modeling
Mapping	Removal of correlated variables; reduced dimensionality; reduction of computational burden for optimization	Non-uniqueness of the extracted parameters; few techniques for high-dimensional problems	Modeling and optimization
Space reduction	Reduction of the effort on modeling and optimization	May miss the global optima or important sub-space	Often used at the start of optimization
Visualization	Supporting design space exploration and optimization	Difficult for high-dimensional problems	Interactive decision making; exploration

applicability. Screening and mapping approaches can be very useful in suitable context, especially when there is prior knowledge of the underlying black-box function. Mapping strategies for high-dimensional problem modeling and optimization are limited and need to be further developed. Space reduction is a common strategy used in detailed optimization algorithms. It may best suit for search strategies such as in trust region methods. Its use in the global scale, however, is to be cautioned as it is risky of missing important subspaces. Visualization techniques are very attractive for human interactive

decision making. They can be used to design an interface between the fundamental analytical approaches (such as modeling and optimization) and design engineers, in support of real design practice.

2.4. Model Approximation Techniques

Computationally-expensive problems and black-box problems are often found in science and engineering disciplines. For example, simulation and analysis processes are expensive to run and often considered black-box functions. The widely used strategies dealing with computational intensity, unknown function expressions, and both are model approximation techniques. These model approximations support engineering design optimization as well (Haftka *et al.* 1998; Wang and Shan 2007). This section first surveys the existing model approximation techniques, and then introduces a type of additive high-dimensional model representation potentially supporting the solution of HEB problems. We will then elucidate the relationship between modeling techniques and nature of underlying functions to expose the oversight/flaws in current methods and indicate the direction for new model development.

2.4.1. Existing Modeling Techniques

Model approximation techniques involve two fields: design of computer experiments and modeling. These two fields work together to serve for model approximation. In typical model approximation techniques, there are four basic tasks: (1) to decide on a sampling method (that is, experimental design); (2) to select a model to fit sampling points; (3) to choose a fitting method (for example, least square regression); and (4) to validate the fitting model. These tasks often correlate with each other. A critical

issue in model approximation is to construct a sufficiently accurate approximation model with least effort based on available information.

The research on design of computer experiments (Sacks *et al.* 1989a; 1989b; Steinberg and Hunter 1984) has been a few decades. The reviews on design of computer experiments can be found in the references (Chen *et al.* 2006; Chen *et al.* 2003; Crary 2002; John and Draper 1975; Steinberg and Hunter 1984). Multiple design of computer experiments schemes are compared by researchers. For example, McKay *et al.* (1979) compared three sampling methods (random sampling, stratified sampling and Latin hypercube sampling). Simpson *et al.* (2001a) compared and contrasted five types of experimental design and four types of approximation model. Jin *et al.* (2002) compared sequential sampling with one stage sampling. Chan (1983) analyzed the sample variance algorithms and made recommendations. Ford *et al.* (1989) summarized work in optimal experimental design in nonlinear problems. Wang and Shan (2007) listed various design of experiments approaches. Chen *et al.* (2006) summarized some of the experimental designs' pros and cons.

The design of computer experiments can be grouped into three categories: the first category of designs is constructed by combinatorial, geometrical, or algebraic methodology, such as factorial design, fractional factorial design (Myers and Montgomery 1995), orthogonal arrays (Bose and Bush 1952; Hedayat *et al.* 1999; Owen 1992a), Latin hypercube designs (Owen 1992b; Tang 1993; Ye 1998), etc. These designs have desirable structural properties, and some of them have good projective property in low-dimensional subspaces. The second category of designs is constructed by optimality

approaches, such as D, A, E, G, I_λ Optimality (Chen *et al.* 2003; John and Draper 1975; Steinberg and Hunter 1984), minimax and maximin distance designs (Johnson *et al.* 1990), and Bayesian approaches (Chaloner and Verdinelli 1995; Currin *et al.* 1988, 1991; Mitchell and Morris 1992; Morris *et al.* 1993). In Bayesian-based sampling, the mean serves as a prediction, and the standard deviation serves as a measure of uncertainty of the prediction. Measures of information based on the predictive process are used to establish design criteria, and optimization can be used to choose good designs. The second category of methods usually yield sample points of comparatively good space-filling proprieties; however, obtaining these designs can be either difficult or computationally intractable, and they may not have good projective properties in low-dimensional subspaces. The third category of methods (for instance, Jin *et al.* 2005; Morris and Mitchell 1995) combine the optimality approaches with the first category approaches (for example, Latin hypercube sampling) to improve projective property as well as space-filling property. For evaluating the experimental design, Simpson *et al.* (2001b) and Chen *et al.* (2003; 2006) discussed some metrics of merit. Those metrics of merit are summarized in Table 2.2.

For design of computer experiments, the "curse of dimensionality" presents a major hurdle as the amount of required sampling points for modeling grows with the number of design variables (Pérez *et al.* 2002b). Since a full factorial design is the most basic design, taking the full factorial design as a basis, Table 2.3 lists the cost of some experimental designs to illustrate the challenges when the number of dimension (*n*=30) is relatively high. The research on construction of designs for high-dimensional spaces has not been extensive (Currin *et al.* 1991). Another issue worthy of notice is the interactions

within experimental designs. Morris and Mitchell (1983) discussed the presence of interactions.

Table 2.2 Metrics for evaluating experimental design

Metric	Description
Orthogonality	A design is orthogonal if, for every pair of factors x_i and x_j, the sum of the cross-products of N design points $\sum_{u=1}^{N} x_{iu} x_{ju}$ is zero, which implies that the design points are uncorrelated.
Rotatability	A design is rotatable if $N \bullet Var[\hat{f}(x)]/\sigma^2$ has the same value at any two locations that are of the same distance from the design center, which maintains the same structure after rotation; where $\hat{f}(x)$ is approximation of the underlying function.
Robustness	Robustness measures how well the design performs when there are violations of the assumptions upon which the design was derived.
Minimum variance and minimum bias	Estimation having minimum variance and minimum bias

Table 2.3 Cost of some experimental designs

Experimental design	Condition (number of variables n=30)	Cost
Full factorial	Two level design	$2^{30} = 1.0737\, e9$
Fraction factorial	Half fraction	$\frac{1}{2} \times 2^{30} = 53,6870,912$
Central composite	A central composite design is a two level 2^n factorial design, augmented by n_0 center points and two 'star' points positioned at $\pm \alpha$ for each factor	527,189 for 20 factors (generated by Matlab^{TM} function "ccdesign(20)"; "ccdesign(30)" failed)

In the modeling field, approximation models can be grouped into two categories: parametric models and nonparametric models as shown in Fig. 2.3. Based on these two categories of models, semi-parametric models are developed. Parametric models have a pre-selected form of the original variables for the underlying function, and so can be

parameterized in terms of any basis functions, for example, polynomial models (linear, quadratic or higher) (Hill and Hunter 1966). Simple parametric models require a few data points to obtain a meaningful result and can be rapidly computed. However, parametric models have limited flexibility, and are likely to produce accurate approximations only when the true form of the underlying functions is close to the pre-specified parametric one (Denison 1997; Friedman 1991). They are preferred when there is prior knowledge of the underlying function.

```
┌────────────────┐      ┌──────────────────┐      ┌──────────────────────┐
│  Parametric    │      │ Full dimensional │      │ Piecewise parametric │
│  approximation │      │ models           │      │ (e.g., spline-based  │
│  (e.g., polynomial│   │ (Modeling a d-   │      │ methods)             │
│  response surface)│   │ dimensional      │      └──────────────────────┘
└────────────────┘      │ function with a  │      ┌──────────────────────┐
                        │ d-dimensional    │      │ Local parametric     │
                        │ estimate)        │      │ (e.g., Kriging)      │
┌────────────────┐      └──────────────────┘      └──────────────────────┘
│ Nonparametric  │                                ┌──────────────────────┐
│ approximation  │                                │ Roughness penalty    │
└────────────────┘                                └──────────────────────┘
                        ┌──────────────────┐      ┌──────────────────────┐
                        │ Lower-dimensional│      │ Additive models      │
                        │ models (modeling │      └──────────────────────┘
                        │ the d-dimensional│      ┌──────────────┐  ┌──────────────┐
                        │ function with    │      │ Adaptive     │  │ Projection   │
                        │ low-dimensional  │      │ computation  │  │ pursuit      │
                        │ functions)       │      └──────────────┘  └──────────────┘
                        └──────────────────┘                       ┌──────────────┐
                                                                   │ Classification│
                                                                   │ and regression│
                                                                   │ trees         │
                                                                   └──────────────┘
```

Fig. 2.3 Approximation models

In nonparametric modeling, the functional form is not known and so cannot be parameterized in terms of any basis functions, for instance, smoothing splines and kernel regression. Nonparametric approaches try to fit a function through the use of sampling data to derive the form of the model instead of "enforcing or imposing" them into a particular class of models (for instance, polynomial model). So the model can alter from the sampling data, which reflects the nature of the underlying function. Nonparametric

methods have two main classes: one models a *d*-dimensional regression function with a *d*-dimensional estimate and the other models the underlying function with lower-dimensional functions. The first class includes three types of methods: piecewise parametric, local parametric, and roughness penalty. These techniques can work well for low-dimensional problems, but become unreliable when there are many variables (Denison 1997). The second class takes the underlying function as a combination of low-dimensional functions and sums them together, which circumvents the "curse-of-dimensionality". This class includes two main strategies: additive models (Andrews and Whang 1990; Friedman and Silverman 1989; Stone 1985) and adaptive computation. Adaptive computation includes projection pursuit regression (Friedman and Stuetzle 1981), and recursive partitioning regression (Friedman 1991). Next subsection will describe one additive model. Chen (1991; 1993) proposed interaction spline models to retain the advantages of additive models with more flexibility. Some of the above modeling techniques have been extended by Bayesian approaches (Barry 1986; Denison 1997; Denison 1998; Leoni and Amon 2000; Otto *et al.* 1997). Apley *et al.* (2006) modeled approximation model uncertainty by Bayesian approach. Wang and Shan (2007) listed popular models, such as Kriging models (Joseph *et al.* 2006; Martin and Simpson 2005), radial basis functions (RBF) models (Fang and Horstemeyer 2006; Regis and Shoemaker 2007a, 2007b), response surface models (Hill and Hunter 1966; Kaufman *et al.* 1996), support vector machine (Collobert, 2001), etc. Owen and his group (An and Owen 2001; Jiang and Owen 2002, 2003) developed quasi-regression methods for model approximation. Chen *et al.* (1999) presented an OA/MARS (orthogonal array and multivariate adaptive regression splines) method. Jin *et al.* (2001) compared four models

34

(polynomial regression, multivariate adaptive regression splines, radial basis functions, and Kriging model), and Wang *et al.* (2006) compared metamodels (multivariate adaptive regression splines, radial basis functions, adaptive weighted least squares, Gaussian process and quadratic response surface regression) under practical industry settings. Simpson *et al.* (1998) compared response surface and Kriging models for multidisciplinary design optimization. Chen (2006) described the pros and cons of some models. Mechesheimer *et al.* (2002) investigated assessment methods for model validation based on leave-*k*-out cross validation. Kennedy and O'Hagan (2001) developed a Bayesian approach for calibration of computer models. Calibration is the process of fitting a model to the observed data by adjusting parameters. Some researchers studied the structures and natures of the underlying function. For example, Hooker (2004) discovered an additive structure; Chen (1991; 1993) made use of interactions; Owen (2000; 1998) discussed linearity in high dimensions. Here commonly used performance criteria for approximation models and commonly used model validation metrics are listed in Table 2.4 and Table 2.5, respectively. To the authors' knowledge, there is no specially designed validation method for HEB problems, especially when the total number of validation points is limited due to high computational cost.

Table 2.4 Commonly used performance criteria for approximation models

Criterion	Description
Accuracy	The capability of predicting underlying functions over a design space. It can be measured by RMSE, R square, RAAE, RMAE, and so on (see Table 2.5).
Interpretability or Transparency	The ability of proving the information and interactions (the underlying structure) among variables. It can be seen via function nonlinearity, interaction of the factors and factor contributions.
Flexibility or Robustness	The capability to provide accurate fits for different problems. It can be measured by variance of accuracy metrics.
Dimensionality	The amount of data required to avoid an unacceptably large variance that increases rapidly with increasing dimensionality.
Computability or Efficiency	The computational effort required for constructing the model and for predicting the response for a set of new points by the model. The computational effort required for constructing the model can be measured by the number of function evaluations and the number of iterations or time.
Simplicity	The ease of implementation
Smoothness	The derivative ability of the model function

Table 2.5 Commonly used model validation metrics

Metrics	Features
Residual	The difference between the predicted and true values at sampled points.
Mean Square Error (MSE): $$MSE = \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{m}$$	Measure the average of the "error". The "error" is the difference between the predicted and true values. MSE does not have the same unit as the output, y.
Root Mean Square Error (RMSE): $$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{m}}$$	A better measure of "error" than MSE. RMSE has the same unit as the output.
Relative Average Absolute Error: $$RAAE = \frac{\sum_{i=1}^{m}|y_i - \hat{y}_i|}{m * STD}$$	Usually correlated with MSE. A global error measurement.
R Square: $$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y})^2}$$	Usually correlated with MSE. A global error measurement.
Predicted R squares: $$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y})^2}$$	The formula is the same as the R Square. But the calculation process is similar to cross-validation. It is calculated by systematically removing each point from modeling points, constructing a new model on remaining points, and predicting function value at the removed point.
Maximum Absolute Error: $$MAX = \max |y_i - \hat{y}_i|, \ i = 1, \cdots, m$$	An absolute error measurement in a local region. Not necessarily correlated with MSE.
Relative Maximum Absolute Error : $$RMAE = \frac{MAX}{STD}$$	A relative error measurement in a local region. Not necessarily correlated with MSE.
Cross-validation	Partitioning sampled points into multiple subsets and then iteratively employing one subset as testing set and other subsets as training set (modeling) to test the accuracy of the model. It includes leave-one-out and k-fold cross-validation.

Where

m -- the number of validation points; y_i -- observed value; $\hat{y}_i$ -- predicted value; $\overline{y}$ -- the mean of the observed values; STD -- standard deviation $STD = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \overline{y}_i)^2}{m}}$

2.4.2. **High-Dimensional Model Representation**

Among the additive models, a high-dimensional model representation (HDMR), which was developed from science disciplines, has only drawn limited attention in engineering. The HDMR, given its direct relevance, potential application for high-dimensional design, and limited exposure to engineering researchers, is thus described in more detail as follows.

A HDMR represents the mapping between the input variables $\mathbf{X} = [x_1, x_2, ..., x_n]^T$ defined on the design space R^n and the output $f(\mathbf{x})$. A general form of HDMR (Li *et al.* 2001a; Rabitz and Alis 1999; Sobol 1993) is shown as

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} f_{ijk}(x_i, x_j, x_k) + \cdots +$$

$$\sum_{1 \leq i_1 < \cdots < i_l \leq n} f_{i_1 i_2 \cdots i_l}(x_{i_1}, x_{i_2}, \cdots, x_{i_l}) + \cdots + f_{12 \cdots n}(x_1, x_2, \cdots, x_n), \tag{2.1}$$

where the component f_0 is a constant representing the zero-th order effect to $f(\mathbf{x})$; the component function $f_i(x_i)$ gives the effect of the variable x_i acting independently upon the output $f(\mathbf{x})$ (the first-order effect), and can have an arbitrary dependence (linear or non-linear) on x_i. The component function $f_{ij}(x_i, x_j)$ describes the interacting contribution of the variables x_i and x_j upon the output (the second-order effect), and subsequent terms reflect the interacting effects of an increasing number of interacting variables acting together upon the output $f(\mathbf{x})$. The last term $f_{12 \cdots n}(x_1, x_2, \cdots, x_n)$ represents any residual dependence of all the variables locked together correlatively to influence the output $f(\mathbf{x})$. The HDMR expansion has a finite number of terms and is

38

always exact. The HDMR expands a d-dimensional function into summation of different functions of less than d-dimensions. The HDMR is a generalization of additive models (Andrews and Whang 1990; Chen 1991, 1993; Friedman and Silverman 1989; Stone 1985) mentioned in the Section 2.4.1. The highest dimensionality of HDMR depends on the nature of interaction variables of the function. For most well-defined systems, high-order correlated behavior of the input variables is expected to be weak and a HDMR can capture this effect (Rabitz and Alis 1999). Broad evidence supporting this statement comes from the multivariate statistical analysis of many systems where significant high correlated input variable covariance rarely appears. Owen (2000) observed that high-dimensional functions appearing in the documented success stories did not have full d-dimensional complexity.

HDMR discloses the hierarchy of correlations among input variables. Each of the component functions in HDMR reveals a unique contribution of the variables separately or correlatively to influence the output $f(\mathbf{x})$. At each new level of HDMR, higher-order correlated effects of input variables are introduced. While there is no interaction between input variables, only the constant component f_0 and the function terms $f_i(x_i)$ will exist in the HDMR model. These component functions are thus hierarchically tailored to $f(\mathbf{x})$ over the entire design R^n. A hierarchy of identified interaction functions reveals the structure of $f(\mathbf{x})$.

There is a family of HDMRs that have been developed by the use of different choices of projection operators. Rabitz and his research group (Rabitz and Alis 1999; Rabitz *et al.* 1999) illustrated ANOVA-HDMR and cut-HDMR. Wang *et al.* (2003) and

Li *et al.* (2006) presented random sampling HDMR. Mp-cut-HDMRs (Li *et al.* 2001b) (monomial-based preconditioned HDMR) were developed to improve features of Cut-HDMR. The choice of a particular HDMR is suggested by what is desired to be known about the output and is also dictated by the amount and type of available information. If the additive nature dominates in a problem, a HDMR or GHDMR (generalized HDMR) can efficiently partition the multivariate problem into low-dimensional component functions. When the multiplicative nature is predominant in a problem, a factorized high-dimensional model representation (FHDMR) (Tunga and Demiralp 2005) can be used. If the problem has a hybrid nature (neither additive nor multiplicative), HHDMR (Tunga and Demiralp 2006) (hybrid HDMR) has been developed. HDMR applications can be seen from references (Banerjee and Ierapetritou 2002; Jin *et al.* 2004; Kaya *et al.* 2004; Shorter *et al.* 1999; Taskin *et al.* 2002). Although HDMR has demonstrated good properties, the model at its current stage only offers a check-up table or need integration, lacks of a method to render a complete model, and there is no accompanying sampling method to support the development of HDMR model.

Since the purpose for introducing the HDMR is to model HEB problems, both cost and accuracy are of concern. From this perspective, a Cut-HDMR (Li *et al.* 2001a) is more attractive than other HDMR variations. Cut-HDMR expresses $f(\mathbf{x})$ by a superposition of its values on lines, planes and hyper-planes (called cuts) passing through the "cut" center $\mathbf{x}_0$ which is a point in the input variable space. The Cut-HDMR expansion is an exact representation of the output $f(\mathbf{x})$ along the cuts passing through the "cut" center. The Cut-HDMR exploration of the output surface $f(\mathbf{x})$ may be global,

and the value of $\mathbf{x}_0$ is irrelevant if the expansion is taken out to convergence. The component functions of the Cut-HDMR are listed as follows:

$$f_0 = f(\mathbf{x}_0), \tag{2.2}$$

$$f_i(x_i) = f(x_i, \mathbf{x}_0^i) - f_0, \tag{2.3}$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \mathbf{x}_0^{ij}) - f_i(x_i) - f_j(x_j) - f_0, \tag{2.4}$$

$$f_{ijk}(x_i, x_j, x_k) = f(x_i, x_j, x_k, \mathbf{x}_0^{ijk}) - f_{ij}(x_i, x_j) - f_{ik}(x_i, x_k) - f_{jk}(x_j, x_k) - f_i(x_i) - $$
$$f_j(x_j) - f_k(x_k) - f_0, \tag{2.5}$$

$$\cdots,$$

$$f_{12\cdots n}(x_1, x_2, \cdots, x_n) = f(\mathbf{x}) - f_0 - \sum_i f_i(x_i) - \sum_{ij} f_{ij}(x_i, x_j) - \cdots, \tag{2.6}$$

where $\mathbf{x}_0^i$, $\mathbf{x}_0^{ij}$ and $\mathbf{x}_0^{ijk}$ are, respectively, $\mathbf{x}_0$ without elements x_i; x_i, x_j; and x_i, x_j, x_k. $f(\mathbf{x}_0)$ is the value of $f(\mathbf{x})$ at $\mathbf{x}_0$; $f(x_i, \mathbf{x}_0^i)$ is the model output with all variables evaluated at $\mathbf{x}_0$ except for the x_i component. It is easy to prove that $f_0 = f(\mathbf{x}_0)$ is the constant term of the Taylor series (Li *et al.* 2001b); the first-order function $f_i(x_i)$ is the sum of all the Taylor series terms which only contain variables x_i, while the second-order function $f_{ij}(x_i, x_j)$ is the sum of all the Taylor series terms which only contain variables x_i and x_j, and so on. To sum up, each distinct component function of the Cut-HDMR is composed of an infinite sub-class of the full multi-dimensional Taylor series, and the sub-classes do not overlap one another.

The computational cost of generating Cut-HDMR up to the *i*-th level, when it is used for interpolation purposes, is given by (Rabitz and Alis 1999)

$$c = \sum_{i=0}^{l} \frac{n!}{(n-i)!i!}(s-1)^i, \tag{2.7}$$

41

where s is the number of sample points taken along each x axis. This computational cost can be derived from summing each term's computational cost in Eq. (2.1). If convergence of the Cut-HDMR expansion occurs at $L \leq n$, then the sum above is dominated by the L-th order term . Considering $s \geq 1$, a full space resolution is obtained at the computational cost of $\sim (ns)^L / L!$, which is approximated from Eq. (2.7). This result is in strong contrast with the conventional view of exponential scaling of $\sim s^n$. It can be seen from Eq. (2.7) that the higher-order terms in the Cut-HDMR demand a polynomially increasing number of sampling points. One approach to relieve this issue is to represent a high-order Cut-HDMR component function as a sum of preconditioned low-order Cut-HDMR component functions (Li *et al.* 2001b).

2.4.3. **Relationship among Factors for Approximation**

In the previous subsections, design of computer experiments and modeling techniques have been reviewed. These two techniques work together in metamodeling techniques. The goodness of the generated approximation models is not only related to sampling points (design of computer experiments) and the model, but also to the nature of the underlying problems. This work identifies four basic features to capture complexities of an underlying problem, that is, dimensionality, nonlinearity, interactions among variables, and importance of terms (that is, individual variables or a subset of interrelated variables). The relationship between features of the underlying problems and model approximation techniques is depicted in Fig. 2.4. In Fig. 2.4, an underlying function (high-fidelity model) is approximated by a constructed model; both the underlying function and constructed model include the same input variables; the goodness of the
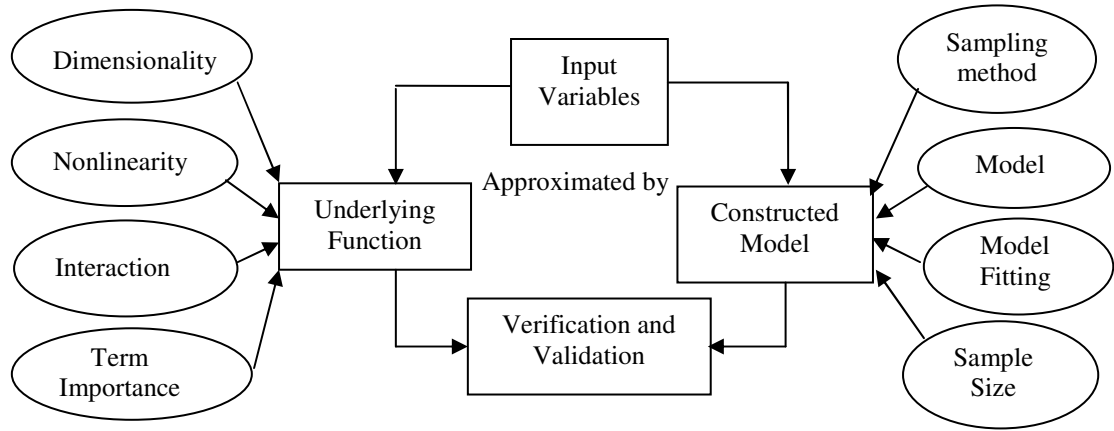
Fig. 2.4 Relationship among factors for approximation

constructed model fitting the underlying function is verified and validated by validation

criteria. The complexities of an underlying function are expressed by its dimensionality,

nonlinearity, interaction among variables, and importance of terms. Factors influencing

the model quality include modeling strategy (for instance, sampling method, model type,

model fitting method and sample size), as well as the nature (for example, the

dimensionality, nonlinearity, interaction, and term importance) of the underlying

functions. From this survey, it is observed that design of computer experiments and

modeling techniques have been widely studied at the right side of Fig. 2.4 including

sampling methods, models, model fitting, and sample size reduction. These techniques

have been successfully applied to various disciplines for low-dimensional problems. As

the dimensionality of the problems increases, it is increasingly difficult to construct most

of such models for problems of a large number of variables. Although high-

dimensionality is the major problem in metamodeling, limited publications exist in the

literature to address this issue. High-dimensional models therefore need to be developed.

43

It is observed that there are few papers that studied the entire structure of the underlying function (the left side of Fig. 2.4). We propose the use of dimensionality, nonlinearity, interaction among variables, and importance of terms, as four characteristics of an underlying/black-box function. In order to overcome the high-dimensional issue, high-dimensional models need to lighten both sides of Fig. 2.4 (that is, nature of the underlying function and approximation techniques). The models should be adaptive and can automatically explore and make use of the nature of the underlying function (dimensionality, interaction, nonlinearity, and importance of terms). These adaptive models require new methods of computer experimental designs, which should have good projective and space filling properties. Generally, there exists a tension between space filling property and small sample size. Resolution of this tension should be expected by means of exploring and using the nature of the underlying function, as well as strategies such as decomposition, additive modeling, mapping, etc. The HDMR model is designed for modeling high-dimensional problems, which bears great potential for further development.

2.5. Optimization Strategies as Related to HEB Problems

Optimization problems with computationally expensive/black-box models exist commonly in many disciplines. Optimization processes inherently require iterative evaluations of objective functions. Therefore, the cost of optimization often becomes unacceptable. Especially high-dimensional, computationally-expensive, and black-box (HEB) problems pose more demanding requirements. This section reviews current optimization strategies for computationally-expensive black-box functions, and non-

gradient optimization methods that are normally developed for cheap black-box functions. Given the broad scope of optimization, this review focuses mostly on non-gradient methods, and selects optimization methods that are considered inspiring (inevitably with bias) for the development of new optimization methods for HEB problems.

2.5.1. **Optimization Strategies for Computationally-Expensive Black-box Functions**

It can be seen from literature that implementation of optimization of computationally expensive black-box functions often uses a cheap or approximate model as a surrogate of the expensive model (for instance, Jones *et al.* 1998; Schonlau *et al.* 1998). The optimization strategies for computationally-expensive black-box functions fall into two classes as shown in Fig. 2.5: model approximation based techniques, and coarse-to-fine model based techniques.
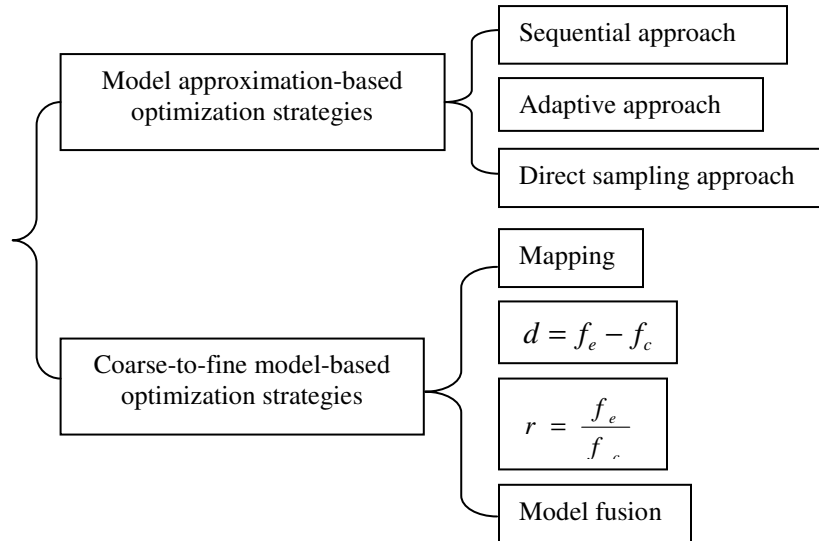
Fig. 2.5 Optimization strategies for computationally expensive problems

Model approximation-based optimization techniques utilize a cheap model to approximate an expensive model and then optimize the cheap model or use information obtained from the cheap model to guide optimization. This kind of technique is also termed metamodel-based design optimization (MBDO) strategy. There are three different types of strategies in the literature, as illustrated in Fig. 2.6 (Wang and Shan 2007). Most of the MBDO approaches fall into the first two strategies. The third strategy is rather new and demonstrates good robustness, efficiency, and effectiveness. The first strategy, though being the most straightforward one among the three, can be practical in industry when sample points are already available and budget or time does not allow for iterative sampling. When iterations of sampling are allowed, the latter two strategies in general should lead to a less total number of function evaluations. All of the MBDO methods, however, are limited by the difficulty of approximating high-dimensional problems with a small number of points.

Fig. 2.6 MBDO strategies: a) sequential approach, b) adaptive MBDO, and c) direct sampling approach

The coarse-to-fine model based techniques combine the high accuracy of a fine model (high-fidelity model) with low cost of a coarse model (or low-fidelity model). The coarse model is exploited to obtain the information of optimization functions including rapid exploration of different starting points, local minima, sensitivities and other design characteristics within a suitable time frame while the fine model is used to verify the design obtained by the coars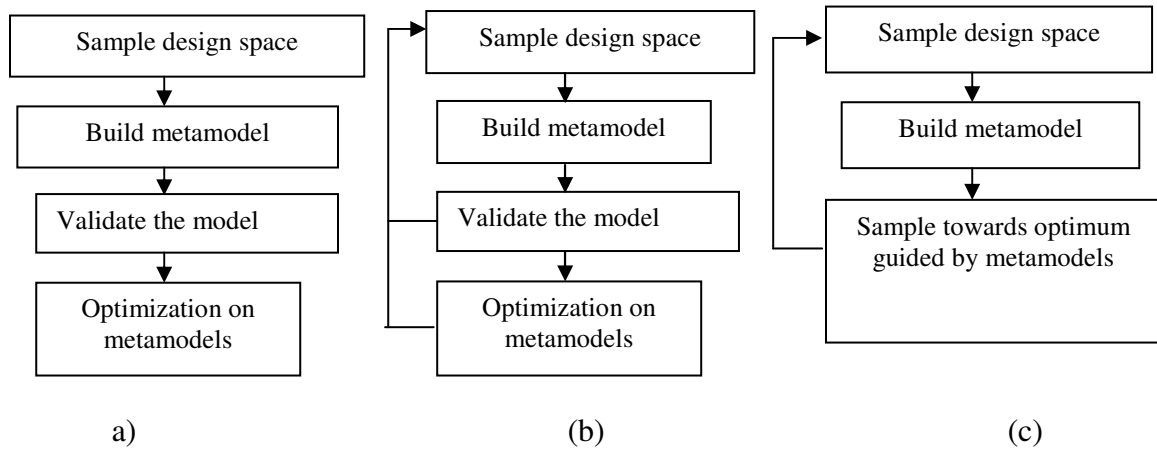e model or evaluated in important regions to improve the accuracy. There are several methods in this technique, as shown in Fig. 2.5, such as mapping, difference modeling, ratio modeling (Leary *et al.* 2003) and model fusion (Xiong *et al.* 2008). Mapping (Bakr *et al.* 1999a, 1999b, 1998; Bandler *et al.* 1995a,1995b,1994; Leary *et al.* 2001) aims to establish a relationship between the input space of the coarse model and that of the fine model such that the coarse model with the mapped parameter accurately mirrors the behavior of the fine model. This mapping approach is reviewed in Section 2.3.3. Difference modeling considers differences between two models ($d = f_e - f_c$ where f_e represents the expensive model and f_c the cheap model). Watson and Gupta (1996) modeled the differences between the two models by a neural network and applied it to the microwave circuit design. Ratio modeling is to model the ratio of fine and coarse models ($r = \dfrac{f_f}{f_c}$ where f_f is the fine model; f_c is the coarse model). Haftka (1991) calculated the ratio and derivatives at one point in order to provide a linear approximation to the ratio at other points in the design space. Nain and Deb (2002) proposed a concept of combining genetic algorithm with coarse-to-fine grain modeling. Xiong *et al.* (2008) proposed a variable fidelity optimization framework based on model fusion. The coarse-to-fine model based

techniques need a given (or easy-to-obtain) coarse model. They are suitable for problems with some prior knowledge.

2.5.2. **Non-gradient Optimization Algorithms**

There are many well known optimization algorithms such as quasi-Newton methods (Arora 1995), interior point algorithms (Rao and Mulkay 2000), generic algorithms (GA) (Holland 1975), simulated annealing (SA) (Kirkpatrick *et al.* 1983), trust region (Celis *et al.* 1984), and DIRECT (Jones *et al.* 1993). There are also various classification methods for algorithms. Multiple papers on algorithm review and comparison have been published. For example, Weise (2008) and Arora *et al.* (1995) reviewed and classified optimization algorithms. Ratschek and Rokne (1987) discussed the efficiency of a global optimization algorithm. Vanderplaats (1999) reviewed structural design optimization status. One can draw conclusions from these surveys: (1) there is no generally applicable optimization algorithm for all problems; (2) there is no analytical conclusion on which optimization algorithm is the most efficient; (3) no algorithm is found in open literature that is directly applicable to HEB problems.

In view of the enormous amount of literature on optimization algorithms, this section aims only to extract some interesting and promising ideas behind algorithms that may potentially be integrated with aforementioned various technologies (for instance, decomposition) to solve HEB problems. This review is not intended to repeat previous works on reviewing, classifying, and comparing various optimization algorithms. Considering the gradients either usually not available, or the costs needed to find gradients for black-box functions falling victim to the "curse," this chapter is limited to

non-gradient, or derivative-free, algorithms and only presents some of the often-used algorithms in engineering design.

DIRECT (dividing rectangles): this algorithm was developed by Jones's group (1993). It is a modification of the Lipschitzian approach that eliminates the need to specify a Lipschitz constant. DIRECT iteratively subdivides the design space into hyper-rectangles and selects the set of hyper-cubes that are most likely to produce the lowest objective function. Björkman and Holmström (1999) implemented the DIRECT algorithm in Matlab^{TM}. DIRECT is found to be more reliable than competing techniques for an aircraft routing problems (Bartholomew-Biggs *et al.* 2003) and have attractive results for benchmark problems (Björkman and Holmström 1999; Jones *et al.* 1993). DIRECT meets increasing difficulty with an increasing number of variables and is normally applied to low-dimensional problems. Siah *et al.* (2004) combined DIRECT with Kriging model and solved several optimization problems of 3 or 4 variables in the electromagnetic field. Their approaches fall into the ones as shown in Fig. 2.6 (a) and 2.6 (b).

Pattern Search: pattern search, originated in 1950s (Box 1957), is a direct search algorithm which searches for a set of points around the current point, looking for one at which the value of the objective function is lower than the value at the current point. The set of points is decided by a prefixed or random pattern. This approach does not require gradient information of the objective function and can solve optimization problems with discontinuous objective functions, highly nonlinear constraints, and unreliable derivative information. This algorithm is applied to unconstrained, constrained, and black-box function optimization (Audet and Dennis 2004). Its advantages are being simple, robust

and flexible. But they are easy to trap into local optima, and the number of evaluations is high. It is suitable for low-dimensional optimization problems.

Genetic algorithm (GA): Genetic algorithms (Holland 1975, Goldberg 1989) come from the idea of natural selection. Generic algorithms generate a population of points at each iteration. The population approaches an optimal solution and selects the next population by computations that involve random choices. GA is a robust stochastic global optimization algorithm. Since many evaluations are commonly required, its efficiency is generally low. In addition, parameters (population size, crossover, mutation operators, etc.) need tuning for each problem. Yoshimura and Izui (2004) successfully partitioned large-scale, yet computationally-inexpensive, problems into sub-problems and solved the sub-problems by the use of parallel GAs.

Simulated annealing (SA): simulated annealing (Kirkpatrick *et al.* 1983) was inspired by the annealing process in metallurgy. The objective function is analogous to temperature (energy). In order to get the optimal solution, the temperature changes from high to low and cooling should be sufficiently slow. SA suffers from the same drawbacks as GA in that the convergence is slow. The performance of SA depends on proper initialization of program parameters used within SA.

Trust region algorithms (Celis *et al.* 1984) dynamically control a region in the search space (so-called trust region) to pursue the optimum, which can be proved for global convergence. In Matlab^{TM} optimization toolbox, all the large-scale algorithms, except for linear programming, are based on trust-region methods.

Mode-pursuing sampling method (MPS): MPS (Sharif *et al.* 2008; Wang *et al.* 2004) is a recently developed method, which uses a variation of the objective function to act as a probability density function (PDF) so that more points are generated in areas leading to lower objective function values and fewer points in other areas. It is thus in essence a discriminative sampling method. The performance of MPS on high-dimensional problems is not yet examined.

Many other meta-heuristics non-gradient methods have been developed such as Ant Colony (Dorigo *et al.* (1996), Particle Swarm (Kennedy and Eberhart 1995), Differential Evolution (Storn and Price 1995), Fictitious Play (Lambert, *et al.* 2005), and so on. Although each algorithm brings special characteristics, there are some commonalities among the aforementioned optimization algorithms. First, most of these approaches use a set, or population, of search points such as in DIRECT, Pattern Search, GA, SA, and MPS. This will not only help explore the entire search space, it also makes the algorithm amenable to parallel computing. Second, the algorithms differentiate search regions. For example, DIRECT, Pattern Search, and Trust Region methods directly search for more attractive regions for further exploration. By using discriminative sampling, MPS inherently focuses on more attractive regions. GA and SA also indirectly move to more attractive search regions as defined by the current population. Third, most of these methods include a mechanism on where and how to sample/generate a new set of points, or a new population.

2.6. Challenges and Future Research

Challenges of HEB problems come from three aspects: 1) unknown function properties, which almost implies that sampling or stochastic methods have to be used to explore the function; 2) high computational expense for function evaluation, which means that the number of function calls should be minimized; and 3) based on these two challenges, the high-dimensional problem becomes extremely difficult and prominent due to the potentially exponentially increasing expenses. Seeing from this survey, model approximation techniques have been successfully applied to low-dimensional expensive black-box problems. In other words, progresses have been made on the first two challenges; however, further study is worthy and needed for high-dimensional problems. Currently there are only sporadic researches in dealing with aspects of HEB problems; more work therefore needs to be done. The authors believe that among current methods, two methods—mapping and decomposition—are the most promising approaches for solving HEB problems.

In specific, the mapping approach is to transform optimization problems from an original higher-dimensional space to a new lower-dimensional space while preserving the optimum of the original function. That is to say, via optimization on the new function in the lower-dimensional space, the obtained optimum may be inversely transformed to the optimum of the original problem. A few of questions regarding this transformation needs to be addressed: 1) how to preserve the original problems' optimum or how to prove the property of such preservation, and 2) how to define a reversible transformation and how to guarantee its mapping uniqueness?

The decomposition methodology has been widely used for explicit complex functions. It refers to decomposition methods, decomposed models, adaptive sampling methods, modeling validation, and optimization algorithms for these decomposed models.

Following possible research directions are suggested to stimulate more in-depth discussions.

1. New models for high-dimensional problems

Currently widely used models such as Kriging, RBF, and polynomials are not ideal for high-dimensional problems. It is felt that a different model type is needed specifically for HEB problems. Such a model type may be rooted on some sound mathematical assumptions about a high-dimensional space and exploited to explore natures of underlying problems.

2. Deeper understanding of a high-dimensional space

To develop a model for a high-dimensional space, a deeper understanding of a high-dimensional space is felt needed. It is very difficult to imagine an $n>3$ space, given our limited visualization capability. Such a limit hinders the development of intuitive sampling approaches, and also hinders our understanding of such a vast space. Although high-dimensionality of problems logically supports that the number of sampling points can grow exponentially with the number of input variables, broad evidence from statistics supports that significant high-dimensional variable covariance rarely arises (Li *et al.* 2001a, 2001b). This indicates that high-dimensional correlation relationships rapidly

disappear under more general physical conditions in high-dimensional space. In addition, some researchers believe that most engineering problems have a limited number of feasible solutions located at comparatively very small regions in a high-dimensional space. In other words, only very small regions in a vast space are of interest to us. The problem is how to validate such a proposition? If this proposition is true, how to design sampling and modeling techniques to take advantage of such a property? Besides the above mentioned evidence and propositions, are there other properties and/or knowledge about a high-dimensional space? A more in-depth theoretical study of characteristics of high-dimensional problems can help.

3. Need for new sampling schemes

The cost of modeling high-dimensional problems, in general, arises from the increase of dimensionality and the increase of the number of sample points along each dimension. Associated with a new model type for high-dimensional problems, a new sampling method may be needed. Such a sampling method should 1) support the particular model type and modeling method, 2) take advantage of problem characteristics (for instance, nonlinearity and interaction) to have some degree of "intelligence," 3) support adaptive sampling and sequential sampling, and 4) be efficient and effective in capturing the essence of the function—global trends and local details of interesting areas. Sampling methods with both good space filling properties (refining accuracy of interesting areas) and projective properties (capture the trends of the underlying functions) should work together with high-dimensional models.

4. Decomposition for Optimization Problems

Decomposition of a high-dimensional problem is deemed an important and necessary step. The issue is how to decompose a problem according to the inherent relationships among variables and functions, and yet amenable to modeling, sampling, and optimization. How to integrate the decomposition with sampling, modeling, and optimization to achieve overall efficiency and effectiveness? Decomposition-based modeling and/or decomposition-based optimization strategies with exploring capabilities need to be developed for high-dimensional problems.

2.7. Conclusion

This survey has reviewed from a variety of disciplines strategies that can potentially be used to solve high-dimensional, computationally-expensive, and black-box (HEB) problems. In closing, some comments are listed as follows:

- As the use of computer-based simulation and analysis tools becomes more popular in engineering practice, HEB problems become more common.

- There are few publications which directly address HEB problems. Optimization methods for computationally-expensive black-box functions are limited to lower dimensional problems.

- Specially designed sampling methods, model types, and modeling approaches that take advantage of the natures of underlying functions (dimensionality, linearity/nonlinearity, interaction, and importance of terms) are needed for HEB problems.

- Two promising ways — mapping and decomposition — are recommended for solving HEB problems. Decomposition-based modeling and decomposition-based optimization may be necessary.

2.8. Acknowledgments

2.9. References

Ahn K-h, Chung WK (2002) Optimization with joint space reduction and extension induced by kinematic limits for redundant manipulators. Proceedings of the 2002 IEEE International Conference on Robotics & Automation. Washington DC, May 11-15

Alexandrov N, Alter SJ, Atkins HL, Bey KS, Bibb KL, Biedron RT (2002) Opportunities for breakthroughs in large-scale computational simulation and design: NASA/TM-2002-211747

Altus SS, Kroo IM, Gage PJ (1996) A genetic algorithm for scheduling and decomposition of multidisciplinary design problems. ASME Journal of Mechanical Design 118: 486-489

An J, Owen A (2001) Quasi-regression. Journal of Complexity 17(4): 588-607

Andrews DWK, Whang Y-J (1990) Additive interactive regression models: circumvention of the curse of dimensionality. Econometric Theory (6): 466-479

Apley DW, Liu J, Chen W (2006) Understanding the effects of model uncertainty in robust design with computer experiments. ASME Journal of Mechanical Design 128: 945-958

Arora JS, Elwakeil OA, Chahande AI (1995) Global optimization methods for engineering applications: a review. Structural Optimization 9: 137-159

Audet C, Dennis JEJ (2004) A pattern search filter method for nonlinear programming without derivatives. SIAM Journal on Optimization 14(4): 980-1010

Bakr MH, Bandler JW, Biernacki RM, Chen SHS, Madsen K (1998) A trust region aggressive space mapping algorithm for EM Optimization. IEEE Transactions on Microwave Theory and Techniques 46(12): 2412-2425

Bakr MH, Bandler JW, Georgieva N (1999a) An aggressive approach to parameter extraction. IEEE Transactions on Microwave Theory and Techniques 47(12): 2428-2439

Bakr MH, Bandler JW, Georgieva N, Madsen K (1999b) A hybrid aggressive space-mapping algorithm for EM optimization. IEEE Transactions on Microwave Theory and Techniques 47(12): 2440-2449

Bakr MH, Bandler JW, Madsen K, ErnestoRayas-Sanchez J, Sondergaard J (2000a) Space-mapping optimization of microwave circuits exploiting surrogate models. IEEE Transactions on Microwave Theory and Techniques 48(12): 2297-2306

Bakr MH, Bandler JW, Madsen K, Sondergaard J (2000b) Review of the space mapping approach to engineering optimization and modeling. Journal of Optimization and Engineering 1: 241-276

57

Bandler JW, Biernacki RM, Chen SH, Grobelny PA, Hemmers RH (1994) Space mapping technique for electromagnetic optimization. IEEE Transactions on Microwave Theory and Techniques 42(12): 2536-2544

Bandler JW, Bienacki RM, Chen SH, Hemmers RH, Madsen K (1995a) Electromagnetic optimization exploiting aggressive space mapping. IEEE Transactions on Microwave Theory and Techniques 43(12): 2874-2882

Bandler JW, Biernacki RM, Chen SH, Hemmers RH, Madsen K (1995b) Aggressive space mapping for electromagnetic design. IEEE MTT-S Int. Microwave Symp. Dig., Orlando, FL, May 16-20

Bandler JW, Cheng QS, Dakroury SA, Mohamed AS, Bakr MH, Madsen K (2004) Space mapping: the state of the art. IEEE Transactions on Microwave Theory and Techniques 52 (1): 337-361

Banerjee I, Ierapetritou MG (2002) Design optimization under parameter uncertainty for general black-box models. Industrial & Engineering Chemistry Research (41): 6687-6697

Barry D (1986) Nonparametric Bayesian regression. The Annals of Statistics 14(3): 934-953

Bartholomew-Biggs MC, Parkhurst SC, Wilson SP (2003) Global optimization - stochastic or deterministic? Stochastic Algorithms: Foundations and Applications (Vol. 2827/2003: 125-137): Springer Berlin / Heidelberg

Bates RA, Buck RJ, Riccomagno E, Wynn HP (1996) Experimental design and observation for large systems. Journal of the Royal Statistical Society: B 58(1): 77-94

Björkman M, Holmström K (1999) Global optimization using the DIRECT algorithm in Matlab. Advanced Modeling and Optimization 1(2): 17-37

Booker AJ, Dennis JEJ, Frank PD, Serafini DB, Torczon V, Trosset MW (1999) A rigorous framework for optimization of expensive functions by surrogates. Structural Optimization 17(1): 1-13

Bose RC, Bush KA (1952) Orthogonal arrays of strength two and three. Annals of Mathematical Statistics 23 (4): 508-524

Brand M (2003) Continuous nonlinear dimensionality reduction by kernel eigenmaps. http://www.merl.com/papers/docs/TR2003-21.pdf. Accessed on August 8, 2008

Browning TR (2001) Applying the design structure matrix to system decomposition and integration problems: a review and new directions. IEEE Transactions on Engineering Management 48(3): 292-306

Byrd RH, Schnabel RB, Shults GA (1987) A trust region algorithm for nonlinearly constrained optimization. SIAM Journal on Numerical Analysis 24(5): 1152-1170

Box GEP (1957) Evolutionary operation: A method for increasing industrial productivity, Applied Statistics, 6:81–101

Celis MR, Dennis JEJ, Tapia RA (1984) A trust region strategy for nonlinear equality constrained optimization. In: Boggs PT, Byrd RH,Schnable RB (ed) Numerical optimization. Society for Industrial and Applied Mathematics Philadelphia, PA: 71-82

Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Statistical Science 10(3): 273-304

Chan TF, Cong J, Kong T, Shinnerl JR (2000) Multilevel optimization for large-scale circuit placement. Proceedings of the 2000 IEEE/ACM International Conference on Computer-Aided Design San Jose, California, November 05-09

Chan TF, Golub GH, LeVeque RJ (1983) Algorithms for computing the sample variance: analysis and recommendations. The American Statistician 37(3): 242-247

Chen D-Z, Liu C-P (1999) A hierarchical decomposition scheme for the topological synthesis of articulated gear mechanisms. ASME Journal of Mechanical Design 121: 256-263

Chen L, Ding Z, Li S (2005a) A formal two-phase method for decomposition of complex design problems. ASME Journal of Mechanical Design 127: 184-195

Chen L, Ding Z, Li S (2005b) Tree-based dependency analysis in decomposition and re-decomposition of complex design problems. ASME Journal of Mechanical Design 127: 12-23

Chen L, Li S (2005) Analysis of decomposability and complexity for design problems in the context of decomposition. ASME Journal of Mechanical Design 127: 545-557

Chen VCP, Ruppert D, Shoemaker CA (1999) Applying experimental design and regression splines to high-dimensional continuous state stochastic dynamic programming. Operations research 47(1): 38-53

Chen VCP, Tsui K-L, Barton RR, Allen JK (2003) A review of design and modeling in computer experiments. Handbook of Statistics, 22: 231-261

Chen VCP, Tsui K-L, Barton RR, Meckesheimer M (2006) A review on design, modeling and applications of computer experiments. IIE Transactions 38: 273-291

Chen W, Allen JK, Mavris DN, Mistree R (1996) A concept exploration method for determining robust top-level specifications. Engineering Optimization 26(2): 137-158

Chen Z (1991) Interaction spline models and their convergence rates. The Annals of Statistics 19(4): 1855-1868

Chen Z (1993) Fitting multivariate regression functions by interaction spline models. Journal of Royal Statistical Society 55(2): 473-491

Collobert, R and Bengio S (2001) SVMTorch: support vector machines for large-scale regression problems. Journal of Machine Learning Research 1: 143-160

Crary SB (2002) Design of computer experiments for metamodel generation. Analog Integrated Circuits and Signal Processing 32 (7-16)

Currin C, Mitchell T, Morris M, Ylvisaker D (1988) A Bayesian approach to the design and analysis of computer experiments. Oak Ridge National Laboratory: Technical report 6498

Currin C, Mitchell T, Morris M, Ylvisaker D (1991) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. Journal of the American Statistical Association 86(416): 953-963

Denison DGT (1997) Simulation based Bayesian nonparametric regression methods. Ph. D. Thesis, Imperial College, London University, London

Denison DGT (1998) Nonparametric Bayesian regression methods. Proceedings of the Section on Bayesian Statistical Science. American Statistics Association. http://www.ma.ic.ac.uk/statistics/links/ralinks/dgtd.link/jsmpaper.ps. Accessed Nov 6, 08

Ding C, He X, Zha H, Simon HD (2002) Adaptive dimension reduction for clustering high dimensional data. The 2002 IEEE International Conference on Data Mining (ICDM'02). IEEE: Maebashi City, Japan, December 9-12: 147-154

Dorigo M, Maniezzo V, Colorni A (1996) The ant system: optimization by a colony of cooperating agents. IEEE Trans. Sys. Man Cyber. B 26, 29-41

Dunteman GH (1989) Principal components analysis. London: Sage Publications, Inc.

Eldred MS, Giunta AA, Waanders BGvB (2004) Multilevel parallel optimization using massively parallel structural dynamics. Structural and Multidisciplinary Optimization 27(1-2): 97-109

Eldred MS, Hart WE, Schimel BD, Waanders BGvB (2000) Multilevel parallelism for optimization on MP computers: theory and experiment. Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Long Beach, CA, September, AIAA-2000-4818

Fadel GM, Cimtalay S (1993) Automatic evaluation of move-limits in structural optimization. Structural Optimization (6): 233-237

Fadel GM, Riley MF, Barthelemy JM (1990) Two points exponential approximation method for structural optimization. Structural and Multidisciplinary Optimization (2): 117-124

Fang H, Horstemeyer MF (2006) Global response approximation with radial basis functions. Journal of Engineering Optimization 38(4): 407-424

Ford I, Titterington DM, Kitsos CP (1989) Recent advances in nonlinear experimental design. Technometrics 31(1): 49-60

Friedman JH (1991) Multivariate adaptive regressive splines. The Annals of Statistics 19(1): 1-67

Friedman JH, Silverman BW (1989) Flexible parsimonious smoothing and additive modeling. Technometrics 31(1): 3-21

Friedman JH, Stuetzle W (1981) Projection pursuit regression. Journal of the American Statistical Association 76(372): 817-823

Goldberg, D E (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA

Griensven, Av, *et al.* (2006) A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology 324: 10-23

Grignon P, Fadel GM (1994) Fuzzy move limit evaluation in structural optimization. The 5th AIAA/NASA/USAF/ISSMO fifth Symposium on Multidisciplinary Analysis and Optimization Panama City, FL, September 7-9, AIAA-94-4281

Gu L (2001) A comparison of polynomial based regression models in vehicle safety analysis. Proceedings of 2001 ASME Design Engineering Technical Conferences—Design Automation Conference Pittsburgh, PA, September 9-12

Haftka RT (1991) Combining global and local approximations. AIAA Journal 29(9): 1523-1525

Haftka RT, Scott EP, Cruz JR. (1998) Optimization and experiments: a survey. Applied Mechanics Review 51(7): 435-448

Hamby, DM (1994) A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment 32: 135-154

Harada, T, *et al.* (2006) Screening parameters of pulmonary and cardiovascular integrated model with sensitivity analysis. Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug. 30 - Sept 3, 2006

Hedayat AS, Sloane NJA, Stufken J (1999) Orthogonal arrays: theory and applications. New York: Springer

Hill WJ, Hunter WG (1966) A review of response surface methodology: a literature survey. Technometrics 8(4): 571-590

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press: Ann Arbor, MI.

Hooker G (2004) Discovering additive structure in black box functions. Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Seattle, WA, USA, August 22-25

Iman, R L and Conover WJ (1980) Small sensitivity analysis techniques for computer models with an application to risk assessment. Communication Statistics - Theory and Methods A 9(17): 1749-1842

Jiang T, Owen AB (2002) Quasi-regression for visualization and interpretation of black box functions. Stanford, CA: Stanford University

Jiang T, Owen AB (2003) Quasi-regression with shrinkage. Mathematics and Computers in Simulation, 62 (3-6): 231-241

Jin R, Chen W, Simpson T W (2001) Comparative studies of metamodeling techniques under multiple modeling criteria. Structural and Multidisciplinary Optimization 23(1): 1-13

Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. The ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference September 29-October 2, Montreal, Canada

Jin R, Chen W, Sudjianto A (2004) Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design. SAE 2004 World Congress Detroit, MI, USA, March 8-11, SAE 2004-01-0429

Jin R, Chen W Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inferences 134(1): 268-287

John RCS, Draper NR (1975) D-Optimality for regression designs: a review. Technometrics 17(1): 15-23

John W Sammon J (1969) A nonlinear mapping for data structure analysis. IEEE Transactions on Computers C-18 (5): 401-409

Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. Journal of Statistical Planning and Inferences 26 (2): 131-148

Jones DR, Perttunen CD, Stuckman BE (1993) Lipschitzian optimization without the Lipschitz constant. Journal of Optimization Theory and Application 79(1): 157-181

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13: 455-492

Joseph VR, Hung Y, Sudjianto A (2006) Blind Kriging: a new method for developing metamodels. http://www2.isye.gatech.edu/statistics/papers/. Accessed on August 8, 2008

Kaski S (1998) Dimensionality reduction by random mapping: fast similarity computation for clustering. The Neural networks proceedings, 1998. IEEE World Congress on Computational Intelligence Anchorage, AK, USA, May 4-9

Kaufman M, Balabanov V, Burgee SL, Giunta AA, Grossman B, Haftka RT, *et al.* (1996) Variable-complexity response surface approximations for wing structural weight in HSCT design. Computational Mechanics(18): 112-126

Kaya H, Kaplan M, Saygin H (2004) A recursive algorithm for finding HDMR terms for sensitivity analysis. Computer Physics Communications (158): 106-112

Kennedy J, Eberhart RC (1995) Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks Perth, WA, Australia, Nov. 27, 1995: 1942-1948

Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. Journal of the Royal Statistical Society B 63(3): 425-464

Kim HM, Michelena NF, Papalambros PY, Jiang T (2003) Target cascading in optimal system design. ASME Journal of Mechanical Design 125: 474-480

Kirkpatrick S, *et al.* (1983) Optimization by simulated annealing. Science 220: 671-680

Koch PN, Allen JK, Mistree F, Mavris DN (1997) The problem of size in robust design. The Advances in design automation

Koch PN, Simpson TW, Allen JK, Mistree F (1999) Statistical approximations for multidisciplinary design optimization: the problem of size. Journal of Aircraft 36(1): 275-286

Koch PN, Mavris D, Mistree F (2000) Partitioned, multilevel response surfaces for modeling complex systems. AIAA Journal 38(5): 875-881

Kodiyalam S, Sobieszczanski-Sobieski J (2000) Bilevel Integrated system synthesis with response surfaces. AIAA Journal 38(8): 1479-1485

Kokkolaras M, Mourelatos ZP, Papalambros PY (2006) Design optimization of hierarchically decomposed multilevel systems under uncertainty. ASME Journal of Mechanical Design 128: 503-508

Krishnamachari RS, Papalambros PY (1997a) Hierarchical decomposition synthesis in optimal systems design. ASME Journal of Mechanical Design 119: 448-457

Krishnamachari RS, Papalambros PY (1997b) Optimal hierarchical decomposition synthesis using integer programming. ASME Journal of Mechanical Design 119: 440-447

Kusiak A, Larson N (1995) Decomposition and representation methods in mechanical design. ASME Journal of Mechanical Design 117 (special 50th anniversary design issue): 17-24

Kusiak A, Szczerbicki E (1992) A formal approach to specifications in conceptual design. ASME Journal of Mechanical Design 114: 659-666

Kusiak A, Wang J (1993) Decomposition of the design process. ASME Journal of Mechanical Design 115: 687-693

Lambert TJ III, Epelman MA, Smith RL (2005) A fictitious play approach to large-scale optimization. Operations Research 53(3): 477-489

Leary SJ, Bhaskar A, Keane AJ (2001) A constraint mapping approach to the structural optimization of an expensive model using surrogates. Journal of Optimization and Engineering (2): 385-398

Leary SJ, Bhaskar A, Keane AJ (2003) A knowledge-based approach to response surface modeling in multifidelity optimization. Journal of Global Optimization 26: 297-319

Leoni N, Amon CH (2000) Bayesian surrogates for integrating numerical, analytical and experimental data: application to inverse heat transfer in wearable computers. IEEE Transactions on Components and Packaging Technologies 23(1): 23-32

Li G, Hu J, Wang S-W, Georgopoulos PG, Schoendorf J, Rabitz H (2006) Random sampling-high dimensional model representation (RS-HDMR) and orthogonality of its different order component functions. Journal of Physical Chemistry A(110): 2474-2485

Li G, Rosenthal C, Rabitz H (2001a) High dimensional model representations. Journal of Physical Chemistry A 105(33): 7765-7777

Li G, Wang S-W, Rosenthal C, Rabitz H (2001b) High dimensional model representations generated from low-dimensional data samples. I. mp-Cut-HDMR. Journal of Mathematical Chemistry 30(1): 1-30

Li S (2008) Matrix-based decomposition algorithms for engineering application: survey and generic framework. International Journal of Product Development, to appear

Lu SC-Y, Tcheng DK (1991) Building layered models to support engineering decision making: a machine learning approach. ASME Journal of Mechanical Design 113: 1-9

Marin FTS, Gonzalez AP (2003) Global optimization in path synthesis based on design space reduction. Mechanism and Machine Theory (38): 579-594

Martin JD, Simpson TW (2005) Use of Kriging models to approximate deterministic computer models. AIAA Journal 43(4): 853-863

McKay MD, Bechman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2): 239-245

Meckesheimer M, Booker AJ, Barton RR, Simpson TW (2002) Computationally inexpensive metamodel assessment strategies. AIAA Journal 40(10): 2053-2060

Michelena N, Jiang T, Papalambros P (1995) Decomposition of simultaneous analysis and design models. Structural and Multidisciplinary Optimization Decomposition of Design Models: 845-850

Michelena NF, Papalambros PY (1995a) A network reliability approach to optimal decomposition of design problems. ASME Journal of Mechanical Design 117: 433-440

Michelena NF, Papalambros PY (1995b) Optimal model-based decomposition of powertrain system design. ASME Journal of Mechanical Design 117: 499-505

Michelena NF, Papalambros PY (1997) A hypergraph framework for optimal model-based decomposition of design problems. Computational Optimization and Application 8(2): 173-196

Michelena N, Papalambros P, Park HA, Kulkarni D (1999) Hierarchical overlapping coordination for large-scale optimization by decomposition. AIAA Journal 37(7): 890-896

Mitchell TJ, Morris MD (1992) Bayesian design and analysis of computer experiments: two examples. Statistica Sinica 2: 359-379

Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33(2): 161-174

Morris MD, Mitchell TJ (1983) Two-level multifactor designs for detecting the presence of interactions. Technometrics 25(4): 345-355

Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. Journal of Statistical Planning and Inferences (43): 381-402

Morris MD, Mitchell TJ, Ylvisaker D (1993) Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. Technometrics 35(3): 243-255

Myers RH, Montgomery D (1995) Response surface methodology: process and product optimization using designed experiments. Toronto: John Wiley and Sons, Inc.

Nain PKS, Deb K (2002) A computationally effective multi-objective search and optimization technique using coarse-to-fine grain modeling (KanGal Report No. 2002005). Kanpur: Indian Institute of Technology Kanpur

Oakley, JE and O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society B 66(3): 751-769

Otto J, Paraschivoiu M, Yesilyurt S, Patera AT (1997) Bayesian-validated computer-simulation surrogates for optimization and design: error estimates and applications. Mathematics and Computers in Simulation (44): 347-367

Owen AB (1992a) Orthogonal arrays for computer experiments, integration, and visualization. Statistica Sinica 2: 439-452

Owen AB (1992b) A central limit theorem for Latin hypercube sampling. Journal of Royal Statistical Society 54(2): 541-551

Owen AB (1998) Detecting near linearity in high dimensions. Stanford, CA: Stanford University

Owen AB (2000) Assessing linearity in high dimensions. The Annals of Statistics 28(1): 1-19

Papalambros PY (1995) Optimal design of mechanical engineering systems. ASME Journal of Mechanical Design 117(special 50th anniversary design issue): 55-62

Papalambros PY, Michelena NF (1997) Model-based partitioning in optimal design of large engineering systems. Multidisciplinary Design Optimization: State-of-the-Art, SIAM: 209-226

Papalambros PY, Michelena NF (2000) Trends and challenges in system design optimization. Proceedings of the International Workshop on Multidisciplinary Design Optimization Pretoria, S. Africa, August 7-10

Penha RML, Hines JW (2001) Using principal component analysis modeling to monitor temperature sensors in a nuclear research reactor. Proceedings of the Maintenance and Reliability Conference (MARCON 2001) Knoxville, TN, May 6-9

Pérez VM, Apker TB, Renaud JE (2002a) Parallel processing in sequential approximate optimization. The 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference Denver, Colorado, Apr. 22-25, AIAA-2002-1589

Pérez VM, Renaud JE, Watson LT (2002b) Reduced sampling for construction of quadratic response surface approximations using adaptive experimental design. The 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference Denver, Colorado, Apr. 22-25, AIAA-2002-1587

Queipo, NV, *et al.* (2005) Surrogate-based analysis and optimization. Progress in Aerospace Sciences (41): 1-18

Rabitz H, Alis ÖF (1999) General foundations of high-dimensional model representations. Journal of Mathematical Chemistry (25): 197-233

Rabitz H, Alis ÖF, Shorter J, Shim K (1999) Efficient input-output model representations. Computer Physics Communications (117): 11-20

Rao SS, Mulkay EL (2000) Engineering design optimization using interior-point algorithms. AIAA Journal, 38(11): 2127-2132

Rassokhin DN, Lobanov VS, Agratiotis DK (2000) Nonlinear mapping of massive data sets by fuzzy clustering and neural networks. Journal of Computational Chemistry 22(4): 373-386

Ratschek H, Rokne JG (1987) Efficiency of a global optimization algorithm. SIAM J. Numbr. Anal. 24(5): 1191-1201

Regis RG, Shoemaker CA. (2007a) Parallel radial basis function methods for the global optimization of expensive functions. European Journal of Operational Research (182): 514-535

Regis RG, Shoemaker CA (2007b) A stochastic radial basis function method for the global optimization of expensive functions. INFORMS Journal on Computing 19(4): 497-509

Renaud JE (1993) Second order based multidisciplinary design optimization algorithm development. Advances In Design Automation 65-2: 347-357

Renaud JE, Gabriele GA (1991) Sequential global approximation in non-hierarchic system decomposition and optimization. Advances in Design Automation 32-1: 191-200

Rodríguez JF, Renaud JE, Watson LT (1998) Trust region augmented Lagrangian methods for sequential response surface approximation and optimization. ASME Journal of Mechanical Design 120: 58-66

Sacks J, Schiller SB, Welch WJ (1989a) Designs for computer experiments. Technometrics 31(1): 41-47

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989b) Design and analysis of computer experiments. Statistical Science 4(4): 409-435

Saha A, Wu C-L, Tang D-S (1993) Approximation, dimension reduction, and nonconvex optimization using linear superpositions of Gaussians. IEEE Transactions on Computers 42(10): 1222-1233

Sammon, JW (1969) A nonlinear mapping for data structure analysis. IEEE on Transactions on Computers C-18(5): 401-409

Schonlau M, Welch WJ (2006) Screening the input variables to a computer model via analysis of variance and visualization. Paper presented at the Screening Methods for Experimentation in Industry, Drug Discovery, and Genetics Springer, New York.

Schonlau M, Welch WJ, Jones DR (1998) Global versus local search in constrained optimization of computer models. In: Flournoy N, Rosenberger WF, Wong WK (Ed) Lecture Notes-Monograph Series, (Vol. 34: 11-25), New Development and Applications in Experimental Design. Institute of Mathematical Statistics, Hayward, CA

Shan S, Wang GG (2004) Space exploration and global optimization for computationally intensive design problems: a rough set based approach. Structural and Multidisciplinary Optimization 28(6): 427-441

Sharif B, Wang GG, EIMekkawy T (2008) Mode pursuing sampling method for discrete variable optimization on expensive black-box functions. ASME Journal of Mechanical Design 130: 021402-1-11

Shen HT, Zhou X, Zhou A (2006) An adaptive and dynamic dimensionality reduction method for high-dimensional indexing. The VLDB Journal, http://www.itee.uq.edu.au/~zxf/_papers/VLDBJ06.pdf. Accessed on August 8, 2008

Shin Y S, Grandhi RV (2001) A global structural optimization technique using an interval method. Structural and Multidisciplinary Optimization 22: 351-363

Shlens J (2005) A tutorial on principal component analysis. http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf. Accessed on August 8, 2008

Shorter JA, Ip PC, Rabitz HA (1999) An efficient chemical kinetics solver using high dimensional model representation. Journal of Physical Chemistry A (103): 7192-7198

Siah ES, Sasena M, Volakis JL, Papalambros PY (2004) Fast parameter optimization of large-scale electromagnetic objects using DIRECT with Kriging metamodeling. IEEE Transactions on Microwave Theory and Techniques 52 (1): 276-285

Simpson TW, Mauery TM, Korte JJ, Mistree F (1998) Comparison of response surface and Kriging models for multidisciplinary design optimization. The 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization St. Louis, MI. AIAA-98-4755

Simpson TW, Lin DKJ, Chen W (2001a) Sampling strategies for computer experiments: design and analysis. International Journal of Reliability and Application 2(3): 209-240

Simpson TW, Peplinski J, Koch PN, Allen JK. (2001b) Metamodels for computer-based engineering design: survey and recommendations. Engineering with Computers 17(2): 129-150

Simpson TW (2004) Evaluation of a graphical design interface for design space visualization. Proceedings of the 45th AIAA/ASME/ASCE/AHS/ASC structures, Structural Dynamics & Materials Conference Palm Springs, California, 19 - 22 April, AIAA 2004 -1683

Simpson TW, Booker AJ, Ghosh D, Giunta AA, Koch PN, Yang RJ (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. Structural and Multidisciplinary Optimization 27: 302-313

Sobieszczanski-Sobieski J, Haftka RT (1997) Multidisciplinary aerospace design optimization: survey of recent developments. Structural Optimization 14(1): 1-23

Sobieszczanski-Sobieski, J (1990) Sensitivity analysis and multidisciplinary optimization for aircraft design: recent advances and results. Journal of Aircraft 27(12): 993-1001

Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling & Computational Experiment 1(4): 407-414

Somorjai RL, Dolenko B, Demko A, Mandelzweig M, Nikulin AE, Baumgartner R, *et al.* (2004) Mapping high-dimensional data onto a relative distance plane – an exact method for visualizing and characterizing high-dimensional patterns. Journal of Biomedical Informatics (37): 366-376

Srivastava A, Hacker K, Lewis KE, Simpson TW (2004) A method for using legacy data for metamodel-based design of large-scale systems. Structural and Multidisciplinary Optimization (28): 146-155

Steinberg DM, Hunter WG (1984) Experimental design: review and comment. Technometrics 26(2): 71-97

Stone CJ (1985) Additive regression and other nonparametric models. The Annals of Statistics 13(2): 689-705

Storn R, Price K (1995) Differential evolution – a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical Report TR-95-012, International Computer Science Institute (ICSI), Berkley, CA, March 1995

Stump G, Simpson TW, Yukish M, Bennett L. (2002) Multidimensional design and visualization and its application to a design by shopping paradigm. The 9th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Atlanta, GA, September 4-6, AIAA 2002-5622

Suh NP (2001) Axiomatic design: advances and applications. Oxford University Press. New York,

Tang B (1993) Orthogonal array-based Latin hypercubes. Journal of American Statistical Association 88(424): 1392-1397

Taskin G, Saygin H, Demiralp M, Yanalak M. (2002) Least squares curve fitting via high dimensional model representation for digital elevation model. The International symposium on GIS Istanbul-Turkey, September 23-26

Tu J, Jones DR (2003) Variable screening in metamodel design by cross-validated moving least squares method. The 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference Norfolk, Virginia, April 7-10

Tunga MA, Demiralp M (2005) A factorized high dimensional model representation on the nodes of a finite hyperprismatic regular grid. Applied Mathematics and Computation (164): 865-883

Tunga MA, Demiralp M (2006) Hybrid high dimensional model representation (HHDMR) on the partitioned data. Journal of Computational and Applied Mathematics (185): 107-132

Vanderplaats GN (1999) Structural design optimization status and direction. Journal of Aircraft 36(1): 11-20

Wagner TC, Papalambros PY (1993) A general framework for decomposition analysis in optimal design. De-Vol. 65-2 Advances in Design Automation 2: 315-325

Wagner, S (2007). Global sensitivity analysis of predictor models in software engineering. Proceedings of Third International Workshop on Predictor Models in Software Engineering (PROMISE'07) Washington, DC, USA IEEE Computer Society

Wang GG, Dong Z, Aitchison P (2001) Adaptive response surface method - a global optimization scheme for computation-intensive design problems. Journal of Engineering Optimization 33(6): 707-734

Wang GG, Shan S (2004) Design space reduction for multi-objective optimization and robust design optimization problems. SAE Transactions, Journal of Materials & Manufacturing: 101-110

Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. ASME Journal of Mechanical Design 129: 370-389

Wang GG, Simpson TW (2004) Fuzzy clustering based hierarchical metamodeling for space reduction and design optimization. Journal of Engineering Optimization 36(3): 313-335

Wang H, Ersoy OK (2005) Parallel gray code optimization for high dimensional problems. Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications Las Vegas, Nevada, August 16-18

Wang L, Shan S, Wang GG (2004) Mode-pursuing sampling method for global optimization on expensive black-box functions. Journal of Engineering Optimization 36(4): 419-438

Wang L, Beeson D, *et al.* (2006) A comparison of meta-modeling methods using practical industry requirements. The 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference May 1-4, 2006, Newport, Rhode Island, USA

Wang S-W, Georgopoulos PG, Li G, Rabits H (2003) Random sampling-high dimensional model representation (RS-HDMR) with nonuniformly distributed variables: application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene. Journal of Physical Chemistry A (107): 4707-4716

Watson GS (1961) A study of the group screening method. Technometrics 3(3): 371-388

Watson PM, Gupta KC (1996) EM-ANN models for microstrip vias and interconnects in dataset circuits. IEEE Transactions on Microwave Theory and Techniques 44(12): 2495-2503

Weise T (2008) Global optimization algorithms theory and application. http://www.it-weise.de/projects/book.pdf. Accessed Nov. 7, 2008

Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD (1992) Screening, predicting, and computer experiments. Technometrics 34(1): 15-25

Winer EH, Bloebaum CL (2002a) Development of visual design steering as an aid in large-scale multidisciplinary design optimization. Part I: method development. Structural and Multidisciplinary Optimization 23(6): 412-424

Winer EH, Bloebaum CL (2002b) Development of visual design steering as an aid in large-scale multidisciplinary design optimization. Part II: method validation. Structural and Multidisciplinary Optimization 23(6): 425 - 435

Wujek BA, Renaud JE (1998a) New adaptive move-limit management strategy for approximate optimization, Part 1. AIAA Journal 36(10): 1911-1921

Wujek BA, Renaud JE (1998b) New adaptive move-limit management strategy for approximate optimization, Part 2. AIAA Journal 36(10): 1922-1934

Xiong Y, Chen W, Tsui K-L (2008) A new variable fidelity optimization framework based on model fusion and objective-oriented sequential sampling. ASME Journal of Mechanical Design. DOI: 10.1115/1.2976449

Ye KQ (1998) Orthogonal column Latin hypercubes and their application in computer experiments. Journal of the American Statistical Association 93(444): 1430-1439

Ye T, Kalyanaraman S (2003) A unified search framework for large-scale black-box optimization.
http://www.ecse.rpi.edu/Homepages/shivkuma/research/papers/unisearch03.pdf.
Accessed on August 8, 2008

Yoshimura M, Izui K (1998) Machine system design optimization strategies based on expansion and contraction of design spaces. Proceedings of the 7th

AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization St. Louis, USA, September. AIAA-98-4749

Yoshimura M, Izui K (2004) Hierarchical parallel processes of genetic algorithms for design optimization of large-scale products. ASME Journal of Mechanical Design 126: 217-224

Chapter 3

Metamodeling for High-dimensional Simulation-based Design Problems

Songqing Shan

Dept. of Mech. and Manuf.

Engineering

University of Manitoba

Winnipeg, MB, Canada R3T 5V6

G. Gary Wang[2]

School of Engineering Science

Simon Fraser University

Surrey, BC, Canada V3T 0A3

---

[2] Corresponding author, Tel: 778 782 8495 Fax: 778 782 7514 Email: gary_wang@sfu.ca

3.1. Abstract

Computational tools such as finite element analysis and simulation are widely used in engineering. But they are mostly used for design analysis and validation. If these tools can be integrated for design optimization, it will undoubtedly enhance a manufacturer's competitiveness. Such integration, however, faces three main challenges: 1) high computational expense of simulation, 2) the simulation process being a black-box function, and 3) design problems being high-dimensional. In the past two decades, metamodeling has been intensively developed to deal with expensive black-box functions, and has achieved success for low-dimensional design problems, but when high-dimensionality is also present in design, which is often found in practice, there lacks of a practical method to deal with the so-called High-dimensional, Expensive, and Black-box (HEB) problems. This chapter proposes the first metamodel of its kind to tackle the HEB problem. The work integrates Radial Basis Function (RBF) with High-Dimensional Model Representation (HDMR) into a new model, RBF-HDMR. The developed RBF-HDMR model offers an explicit function expression, and can reveal the 1) contribution of each design variable, 2) inherent linearity/nonlinearity with respect to input variables, and 3) correlation relationships among input variables. An accompanying algorithm to construct the RBF-HDMR has also been developed. The model and the algorithm fundamentally change the exponentially growing computation cost to be polynomial. Testing and comparison confirm the efficiency and capability of RBF-HDMR for HEB problems.

3.2. Introduction

A metamodel is a "model of model," which is used to approximate a usually expensive analysis or simulation process; metamodeling refers to the techniques and procedures to construct such a metamodel. In the last two decades, research on metamodeling has been intensive and roughly along one of the four directions, including sampling and evaluation, metamodel development and evaluation, model validation, and metamodel-based optimization. Recently Wang and Shan [1] reviewed the applications of metamodeling techniques in the context of engineering design and optimization. Chen [2] summarized pros and cons of the design of computer experiments methods and approximation models. Simpson *et al.* [3] reviewed the history of metamodeling in the last two decades and presented an excellent summary on what have been achieved in the area thus far and challenges ahead.

It can be seen from the recent reviews that metamodels have been successfully applied to solve low-dimensional problems in many disciplines. One major problem associated with these models (for example, polynomial, RBF, and Kriging) and metamodeling methodologies, however, is that in order to reach acceptable accuracy the modeling effort grows exponentially with the dimensionality of the underlying problem. Therefore, the modeling cost will be prohibitive for these traditional approaches to model high-dimensional problems. In the context of design engineering, according to references [3-6], the dimensionality larger than ten ($d > 10$) is considered high if model/function

evaluation is expensive, and such problems widely exist in various disciplines [6-10]. Due to its computational challenge for modeling and optimization, the high-dimensional problem is referred to as the notorious "curse of dimensionality" in the literature. For combating the "curse of dimensionality," Friedman and Stuetzle [11] developed projection pursuit regression, which worked well with dimensionality $d \leq 50$ with large data sets. Friedman [12] proposed multivariate adaptive regression splines (MARS) model, which potentially makes improvement over existing methodology in settings involving $3 \leq d \leq 20$, with moderate sample size, $50 \leq N \leq 1000$. Sobol [13] has proved the theorem that an integrable function can be decomposed into summation of different dimensions. This theorem indicates that there exists a unique expansion of high-dimensional model representation (HDMR) for any function $f(\boldsymbol{x})$ integrable in space Ω^d. This HDMR is exact and of finite order and has a hierarchical structure. A family of HDMRs with different characters has since been developed, studied, and applied for various purposes [14-21].

In the recent review of modeling and optimization strategies of high-dimensional problems [22], it is found that the research on this topic has been scarce, especially in engineering. In engineering design, there is no metamodel developed to directly tackle HEB problems. Currently available metamodels are not only limited to low-dimensional problems, and are also derived in separation from the characteristics of the underlying problem. A different model type is therefore needed for HEB problems. This chapter proposes the RBF-HDMR model in response to such a need.

As part of the metamodeling methodology, an adaptive sampling method is also developed to support the proposed RBF-HDMR model. In the research of sampling for

metamodeling, sequential and adaptive sampling has gained popularity in recent years, mainly due to the difficulty of knowing the "appropriate" sampling size *a priori*. Lin [23] proposed a sequential exploratory experiment design (SEED) method to sequentially generate new sample points. Jin *et al.* [24] applied Enhanced Stochastic Evolution to generate optimal sampling points. Sasena *et al.* [25] used the Bayesian method to adaptively identify sample points that gave more information. Wang [26] proposed an inheritable Latin Hypercube design for adaptive metamodeling. Jin *et al.* [27] compared a few different sequential sampling schemes and found that sequential sampling allows engineers to control the sampling process and it is generally more efficient than one-stage sampling. In this work, we develop an adaptive sampling method that is rooted in the RBF-HDMR model format. Section 3.5.2 describes the method in detail.

Before we introduce the RBF-HDMR and its metamodeling method, the premise of this chapter is: 1) there exists a unique expansion of HDMR and the full expansion is exact for a high-dimensional function, and 2) for most well-defined physical systems, only relatively low-order correlations among input variables are expected to have a significant impact upon the output; and high-order correlated behavior among input variables is expected to be weak [15]. The order of correlation refers to the number of correlated variables, for instance, bivariate correlation is considered low order while multivariate (for example, five-variable) correlation is high. Premise 1 was proven in Sobol [13]. Broad evidence supporting Premise 2 comes from the multivariate statistical analysis of many systems where significant covariance of highly-correlated input variables rarely appears [6, 15]. Owen [28] observed that high-dimensional functions appearing in the documented success stories did not have full *d*-dimensional complexity.

The rapid dying-off of the order of correlations among input variables does not, however, eliminates non-linear influence of variables, or strong variable dependence, or even the possibility that all the variables are important. These premises pave the way for this work to tackle the "curse of dimensionality".

This chapter is organized as follows. Section 3.3 introduces HDMR. Section 3.4 proposes the RBF-HDMR model. Section 3.5 discusses how we address the high-dimensionality challenge and describes in detail the metamodeling approach for RBF-HDMR. A modeling example is also given for the ease of understanding of RBF-HDMR and its metamodeling approach. Section 3.6 studies the behavior of RBF-HDMR with respect to dimensionality through a study problem and testing on a suite of high-dimensional problems. The test results are also compared with those from other metamodels based on Latin Hypercube samples. Conclusions are drawn in Section 3.7.

3.3. Basic Principle of HDMR

A HDMR represents the mapping between input variables $x = [x_1, x_2, \cdots, x_d]^T$ defined in the design space and the output $f(x)$. A general form of HDMR [13,15] is shown as

$$f(x) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i < j \le d} f_{ij}(x_i, x_j) + \sum_{1 \le i < j < k \le d} f_{ijk}(x_i, x_j, x_k) + \cdots +$$

$$\sum_{1 \le i_1 < \cdots < i_l \le d} f_{i_1 i_2 \cdots i_l}(x_{i_1}, x_{i_2}, \cdots, x_{i_l}) + \cdots + f_{12 \cdots d}(x_1, x_2, \cdots, x_d), \tag{3.1}$$

where the component f_0 is a constant representing the zero-th order effect to $f(x)$; the component function $f_i(x_i)$ gives the effect of the variable x_i acting independently upon the output $f(x)$ (the first-order effect), and may have either a linear or non-linear

dependence on x_i. The component function $f_{ij}(x_i, x_j)$ describes the correlated contribution of variables x_i and x_j upon the output $f(x)$ (the second-order effect) after the individual influences of x_i and x_j are discounted, and $f_{ij}(x_i, x_j)$ could be linear or nonlinear as well. The subsequent terms reflect the effects of increasing numbers of correlated variables acting together upon the output $f(x)$. The last term $f_{12\cdots d}(x_1, x_2, \cdots x_d)$ represents any residual dependence of all the variables locked together to influence the output $f(x)$ after all the lower-order correlations and individual influence of each involved x_i (i =1,…,d) have been discounted. As the order of the component function increases, the residual impact of higher correlations decreases. If the impact of an l-th order component function is negligible, the impact of higher-order (>l-th) component functions will be even smaller and thus negligible as well. For example, if $f_{ij}(x_i, x_j)$ is negligible, then $f_{ijk}(x_i, x_j, x_k)$ will be negligible since it is the residual impact after the influences of $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are modeled. It is known that the HDMR expansion has a finite number of terms 2^d (d is the number of variables, or dimensionality) and is always exact [13].

There is a family of HDMRs with different features [14, 18-20]. Among these types, the Cut-HDMR [15, 16] involves only simple arithmetic computation and presents the least costly model with similar accuracy as other HDMR types. Therefore Cut-HDMR is chosen as our basis for the proposed RBF-HDMR. A Cut-HDMR [14-15] expresses $f(x)$ by a superposition of its values on lines, planes and hyper-planes (or cuts) passing through a "cut" center x_0 which is a point in the input variable space. The Cut-HDMR expansion is an exact representation of the output $f(x)$ along the cuts passing through

$\boldsymbol{x}_0$. The location of the center $\boldsymbol{x}_0$ becomes irrelevant if the expansion is taken out to convergence [15]. On the other hand, if HDMR expansion did not reach convergence, that is, the model omits significant high-order components in the underlying function, a poor choice of $\boldsymbol{x}_0$ may lead to large error [21]. Sobol [21] suggests using the point as $\boldsymbol{x}_0$ that has the average function value; the average is taken from function values of a certain number of randomly sampled points. The component functions of the Cut-HDMR are listed as

$$f_0 = f(\boldsymbol{x}_0), \tag{3.2}$$

$$f_i(x_i) = f(x_i, \boldsymbol{x}_0^i) - f_0, \tag{3.3}$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \boldsymbol{x}_0^{ij}) - f_i(x_i) - f_j(x_j) - f_0, \tag{3.4}$$

$$f_{ijk}(x_i, x_j, x_k) = f(x_i, x_j, x_k, \boldsymbol{x}_0^{ijk}) - f_{ij}(x_i, x_j) - f_{ik}(x_i, x_k) - f_{jk}(x_j, x_k) - f_i(x_i) -$$
$$f_j(x_j) - f_k(x_k) - f_0, \tag{3.5}$$

$$\cdots,$$

$$f_{12\cdots d}(x_1, x_2, \cdots, x_d) = f(\boldsymbol{x}) - f_0 - \sum_i f_i(x_i) - \sum_{ij} f_{ij}(x_i, x_j) - \cdots, \tag{3.6}$$

where $\boldsymbol{x}_0^i$, $\boldsymbol{x}_0^{ij}$, and $\boldsymbol{x}_0^{ijk}$ are respectively $\boldsymbol{x}_0$ without elements x_i; x_i, x_j; and x_i, x_j, x_k. For the convenience of later discussions, the points $\boldsymbol{x}_0$, $(x_i, \boldsymbol{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, $(x_i, x_j, \boldsymbol{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$, $\ldots$, are respectively called the zero-th order, first-order, second-order model-constructing point(s), $\cdots$. Accordingly, $f(\boldsymbol{x}_0)$ is the value of $f(\boldsymbol{x})$ at $\boldsymbol{x}_0$; $f(x_i, \boldsymbol{x}_0^i)$ is the value of $f(\boldsymbol{x})$ at point $(x_i, \boldsymbol{x}_0^i)$.

The HDMR discloses the hierarchy of correlations among the input variables. Each component function of the HDMR has distinct mathematical meaning. At each new order

of HDMR, a higher-order variable correlation than the previous level is introduced. While there is no correlation among input variables, only the constant component f_0 and the function terms $f_i(x_i)$ exist in the HDMR model. It can be proven that $f_0 = f(x_0)$ is the constant term of the Taylor series; the first-order function $f_i(x_i)$ is the sum of all the Taylor series terms which only contain variables x_i, while the second-order function $f_{ij}(x_i, x_j)$ is the sum of all the Taylor series terms which only contain variables x_i and x_j, and so on [14]. These component functions are optimal choices tailored to $f(x)$ over the entire d-dimensional space because these component functions are orthogonal to each other, the influence of each component term is independently captured by the model, and the component functions lead to minimum approximation error defined by $\|f(x)\text{-}f_{\text{model}}(x)\|^2$ [14, 15].

Although Cut-HDMR has demonstrated good properties, the model at its current stage only offers a check-up table, lacks of a method to render a complete model, and also lacks of accompanying sampling methods to support it. This work proposes to integrate RBF to model the component functions of HDMR.

3.4. RBF-HDMR

In order to overcome the drawbacks of HDMR, this work employs RBF to model each component function of the HDMR. Among a variety of RBF formats, this work chooses the one composed of a sum of thin plate spline plus a linear polynomial. The details of the chosen RBF format are in the 3.9 Appendix. Without losing generality, the simple linear RBF format is used for the ease of description and understanding. In RBF-HDMR, RBF models are used to approximate component functions in Eqs. (3.3-3.6),

89

$$\hat{f}_i(x_i) = \sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_i^0)|, \tag{3.7}$$

where $(x_i, x_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$,

$$\hat{f}_{ij}(x_i, x_j) = \sum_{k=1}^{m_{ij}} \alpha_{ij_k} |(x_i, x_j, x_0^{ij}) - (x_{i_k}, x_{j_k}, x_0^{ij})|, \tag{3.8}$$

where $(x_i, x_j, x_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots x_{d_0}]^T, i \neq j,$

$$\cdots,$$

$$\hat{f}_{12\cdots d}(x_1, x_2, \cdots, x_d) = \sum_{k=1}^{m_{12\cdots d}} \alpha_{12\cdots d_k} |(x - x_k)|, \tag{3.9}$$

where $(x_{i_k}, x_0^i), i = 1, \dots, d$ are points $(x_i, x_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$ evaluated at $k = 1, \dots, m_i$ along each x_i component; similarly $(x_{i_k}, x_{j_k}, x_0^{ij}), k = 1, \cdots, m_{ij}$ are points $[x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$ evaluated at $x_i, i{=}1,\dots,m_i$, and $x_j, j{=}1,\dots,m_j$, that are used to construct the first-order component functions; $x_k{=}[x_{1_k}, x_{2_k}, \cdots, x_{i_k}, \cdots, x_{j_k}, \cdots, x_{d_k}]^T$, $k{=}1,\dots, m_{12\cdots d}$, are the points built from evaluated x components for lower-order component functions.

Eqs. (3.7-3.9) are referred to as the modeling lines, planes, and hyper-planes. Substituting the above approximation expressions into the HDMR in Eq. (3.1), we have

$$f(x) \cong f_0 + \sum_{i=1}^{d} \sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)| + \sum_{1 \le i < j \le d} \sum_{k=1}^{m_{ij}} \alpha_{ij_k} |(x_i, x_j, x_0^{ij}) -$$

$$(x_{i_k}, x_{j_k}, x_0^{ij})| + \cdots + \sum_{k=1}^{m_{12\cdots d}} \alpha_{12\cdots d_k} |x - x_k|. \tag{3.10}$$

The approximation in Eq. (3.10) is called the RBF-HDMR model. Inheriting the hierarchy of HDMR, RBF-HDMR distinctly represents the correlation relationship among the input variables in the underlying function, and provides an explicit model with a finite number of terms. The component functions of multiple RBFs in the model approximate the univariates, bivariates, triple-variates, etc., respectively. The RBF-

HDMR approximation of the underlying function $f(x)$ is global. Since the HDMR component functions are orthogonal in the design space [14], approximation of HDMR component functions such as RBF-HDMR likely provides the simplest and also the most efficient model to approximate $f(x)$ over the entire d-dimensional design space.

For typical underlying functions, RBF-HDMR expands to the second-order,

$$f(x) \cong f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \leq i < j \leq d} f_{ij}(x_i, x_j)$$

$$\cong f_0 + \sum_{i=1}^{d} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, x_0^i) - (x_{i_k}, x_0^i) \right| + \sum_{1 \leq i < j \leq d} \sum_{k=1}^{m_{ij}} \alpha_{ij_k} \left| (x_i, x_j, x_0^{ij}) - \right.$$

$$\left. (x_{i_k}, x_{j_k}, x_0^{ij}) \right|. \tag{3.11}$$

The RBF-HDMR in Eq. (3.11) neglects higher-order component terms based on the assumption that the residual impact of the high-order correlation is small after the impact of individual variables and their lower-order correlations has been captured. The second model, however, does include all input variables and is capable of capturing high nonlinearity of the underlying function through nonlinear component functions.

As we know RBF is an interpolative function, each component function goes through its own model construction points; however, since RBF-HDMR is a sum of these component functions, the question is: "will the resultant RBF-HMDR go through all of the evaluated model construction points?"

Lemma:

A RBF-HDMR model passes through all the prescribed sample points used for constructing zero-th order to the current order component functions.

For clarity, the prescribed, as compared to arbitrarily selected, model-constructing points are explained as follows. For the zero-th order component, the model-constructing point is x_0; for the first-order components, the model-construction points include x_0 and (x_{i_k}, x_0^i); for the second-order components, its model-construction points are $x_0, (x_{i_k}, x_0^i), (x_{j_k}, x_0^j)$ and $(x_{i_k}, x_{j_k}, x_0^{ij}), i \neq j$.

The lemma is proved as follows. Assuming x_0 is the cut center, the RBF-HDMR at first-order is defined as $f(x) = f_0 + f_i(x_i)$. Its first-order component function $f_i(x_i)$ is approximated by one dimensional RBF function $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$ by using the function values computed from $f_i(x_{i_k}) = f(x_{i_k}, x_0^i) - f_0$, where x_{i_k} is the k-th model-constructing point along x_i, and $f(x_{i_k}, x_0^i)$ is the true function value at point (x_{i_k}, x_0^i). Since f_0 is a constant and $f_i(x_i)$ interpolates all model constructing points, the RBF-HDMR model $f(x)$ interpolates all the model constructing points x_0 and (x_{i_k}, x_0^i).

For the second-order components, the function values of these components are computed from $f_{ij}(x_i, x_j) = f(x_i, x_j, x_0^{ij}) - f_i(x_i) - f_j(x_j) - f_0$, and $f_{ij}(x_i, x_j)$ is then approximated by a two-dimensional RBF function $\sum_{k=1}^{m_{ij}} \alpha_{ij_k} |(x_i, x_j, x_0^{ij}) - (x_{i_k}, x_{j_k}, x_0^{ij})|$ with points $x_0, (x_{i_k}, x_0^i), (x_{j_k}, x_0^j)$, and $(x_{i_k}, x_{j_k}, x_0^{ij}), i \neq j$. It is easy to see $f_{ij}(x_i, x_j)$ pass through all the evaluated points since they all participated in modeling $f_{ij}(x_i, x_j)$. For first-order component functions, which are functions of only x_i and orthogonal to each other, they will have zero error at $(x_{i_k}, x_{j_k}, x_0^{ij})$ since each $f_i(x_i)$ goes through x_i. Therefore all first-order component functions, and therefore the resultant RBF-HDMR model, will pass through all model constructing points to the second-order component

function, that is, $\boldsymbol{x}_0$, $(x_{i_k}, \boldsymbol{x}_0^i)$, $(x_{j_k}, \boldsymbol{x}_0^j)$, and $(x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij})$, $i \neq j$. Similarly the RBF-HDMR model passes their model-constructing points till the d-th component. As the RBF-HDMR has a finite number of terms and each of its component function is exact on these prescribed model-constructing (or evaluated sample) points, the RBF-HDMR model will pass through all sample points. The lemma is proved.

The above lemma not only reveals an important feature of RBF-HDMR, but also is a great help to answer the following question, "if the RBF-HDMR model is built at the l-th order, how to identify if there is still $(l+1)$-th order component that need to be modeled?"

Let us start with $l=1$, which indicates that all the zero-th and first-order component functions have been modeled using points $\boldsymbol{x}_0$ and $(x_{i_k}, \boldsymbol{x}_0^i)$. If the second-order component functions are to be built, we will use the elements in these existing points to create new sample points $(x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij})$ for modeling. According to the lemma, the to-be-built second-order RBF-HDMR model is then expected to go through these sample points $(x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij})$. If the first order RBF-HDMR model cannot accurately predict the function value at the new sample point $(x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij})$, it indicates that there must exist second-order and/or higher-order correlation that has not been modeled, since the approximation error is zero for the first-order component functions at points $(x_{i_k}, \boldsymbol{x}_0^i)$ and $(x_{j_k}, \boldsymbol{x}_0^j)$.

To generalize the above discussion, we create a point $\boldsymbol{x}_k = [x_{1_k}, x_{2_k}, \cdots, x_{i_k}, \cdots, x_{j_k}, \cdots, x_{d_k}]^T, k \neq 0$ by random combining the sampled values

x_i in the first-order component construction for each input variable (that is, x_i, $i = 1,...,d$ and evaluated at $(x_i, \boldsymbol{x}_0^i)$, respectively). According to the lemma, the complete RBF-HDMR model in Eq. (3.10) should interpolate this point, $\boldsymbol{x}_k$. If an l-th order RBF-HDMR model does not interpolate this point, it indicates that there is higher-order ($>l$-th) component functions need to be modeled to decrease the prediction error, and the metamodeling should therefore continue until convergence. This fact has been incorporated in the metamodeling algorithm, which is to be detailed in Section 3.5.2.

3.5. Metamodeling for RBF-HDMR

3.5.1. Strategies for High-dimensionality

From the recent review [22], the authors find that the cost of modeling an underlying function is affected by multiple factors including the function's dimensionality, linearity/nonlinearity, ranges of input variables, and convergence criteria. Generally speaking, the cost increases as the dimensionality and nonlinearity rise, the ranges of input variables become larger, and as the convergence criteria become stricter. This section describes four strategies associated with the proposed metamodeling method for RBF-HDMR that help to circumvent/alleviate the computational difficulty brought by the increase of dimensionality without the loss of sampling resolution.

First, a RBF-HDMR model has a hierarchical structure from zero-th order to d-th order components. If this structure can be identified progressively, the cost of constructing higher-order components in HDMR can be saved. The computational cost

(that is, the number of sampling points) of generating a Cut-HDMR up to the l-th level is given by [15, 16]

$$c = \sum_{i=0}^{l} \frac{d!}{(d-i)!i!}(s-1)^i = 1 + d(s-1) + \frac{d(d-1)}{2!}(s-1)^2 + \frac{d(d-1)(d-2)}{3!}(s-1)^3 + \cdots$$

$$+ \frac{d(d-1)(d-2)\cdots(d-l+1)}{l!}(s-1)^l, \tag{3.12}$$

where s is the number of sample points taken for each x_i. The cost of Cut-HDMR is related to the highest order of the Cut-HDMR expansion where the convergence is reached. Each term in Eq. (3.12) represents the computational cost for constructing the corresponding order of component functions. The cost relates to three factors—the dimensionality d, the number of sampling points s for each variable (that is, take s levels for each variable), and the highest order of the component functions l. The highest order, l, of component functions represents the maximum number of correlated input variables. As mentioned before, only relatively low-order correlations of the input variables are expected to have an impact upon the output and high-order correlated behavior of the input variables is expected to be weak. Typically $l \leq 3$ has been found to be quite adequate [6]. Considering $l \ll d$, a full space resolution $1/s$ is obtained at the computational cost less than $[d(s-1)]^l/(l-1)!$. Thus the exponentially increasing difficulty s^d is transformed into a polynomial complexity, s^l. This strategy exploits a superposition of functions of a suitable set of low-dimensional variables to represent a high-dimensional underlying function.

Second, for components of the same order, for instance, at the second-order with bivariate correlations, not all possible bivariate correlations may be present in the

underlying function. Therefore some of the non-existing correlations among input variables can be identified and eliminated from modeling to further reduce the cost. The coefficients in Eq. (3.12), for example, $\frac{d(d-1)}{2!}$, $\frac{d(d-1)(d-2)}{3!}$, respectively denote the maximum number of probable combinations of the correlated terms at second and third order component levels. While the number of dimensionality, d, cannot be changed, the number of these coefficients can be reduced if the non-existing correlations can be identified and eliminated, and the modeling cost associated with those terms can therefore be saved. The developed metamodeling algorithm for RBF-HDMR adaptively identifies such non-existing correlations and models the underlying function accordingly, which will be described in the next section.

Third, although the number of sample points, s, for each variable cannot be reduced in order to keep a certain sampling resolution $1/s$, these sample points can be reused for modeling higher-order component functions. For example, while modeling second-order component functions, sample points on the reference axes, or hyper-planes, such as x_0, $\left(x_{i_k}, x_0^i\right)$, and $\left(x_{j_k}, x_0^j\right)$ are re-used.

Lastly, the number of sample points, s, relates to the degree of the nonlinearity of the underlying function with respect to the input variable x_i. The higher the degree of the nonlinearity is, the more sample points along x_i are needed to meet the required accuracy. For a linear component, two sample points are enough to accurately model it. The developed metamodeling algorithm for RBF-HDMR gradually explores the non-linearity of the component functions and thus conservatively allocates such cost.

In summary, the RBF-HDMR model naturally helps to transform an exponentially increasing computational difficulty into a polynomial one by neglecting higher-order component functions. The proposed metamodeling method will also adaptively explore the linearity/nonlinearity of each component function, identify non-existing variable correlations, and reuse sample points to further reduce the modeling costs.

3.5.2. Sampling and Model Construction

Based on the proposed RBF-HDMR model, a sampling and model construction algorithm is developed. The algorithm steps are described as follows:

1. Randomly choose a point $x_0 = \left[x_{1_0}, x_{2_0}, \cdots, x_{d_0}\right]^T$ in the modeling domain as the cut center. Evaluating $f(x)$ at x_0, we then have f_0.

2. Sample for the first-order component functions $f_i(x_i) = f\left([x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T\right) - f_0$ in the close neighbourhood of the two ends of x_i (lower and upper limits) while fixing the rest of x_j ($j \neq i$) components at x_0. In this work, a neighborhood is defined as one percent of the variable range which is in the design space and near a designated point. Evaluating these two end points, we got the left point value $f_{iL}(x_i) = f\left([x_{1_0}, x_{2_0}, \cdots, x_{i_L}, \cdots, x_{d_0}]^T\right) - f_0$ and the right point value $f_{iR}(x_i) = f\left([x_{1_0}, x_{2_0}, \cdots, x_{i_R}, \cdots, x_{d_0}]^T\right) - f_0$ and model the component function as $\hat{f}_i(x_i)$ by a one dimensional RBF model for each variable x_i .

3. Check the linearity of $f_i(x_i)$. If the approximation model $\hat{f}_i(x_i)$ goes through the center point, x_0, $f_i(x_i)$ is considered as linear. In this case, modeling for this component terminates; otherwise, use the center point x_0 and the two end points to re-

construct $\hat{f}_i(x_i)$. Then a random value along x_i is generated and combined with the rest of x_j ($j \neq i$) components at $\mathbf{x}_0$ to form a new point to test $\hat{f}_i(x_i)$. If $\hat{f}_i(x_i)$ is not sufficiently accurate (the relative prediction error is larger than a given criterion, for instance, 0.01%), the test point and all the evaluated points will be used to re-construct $\hat{f}_i(x_i)$. This sampling-remodeling process iterates until convergence. This process is to capture the nonlinearity of the component function with one sample point at a time. Step 3 repeats for all of the first-order component functions to construct the first-order terms of RBF-HDMR model.

4. Form a new point, $(x_i, x_j, \cdots x_d)_k = [x_{1_k}, x_{2_k}, \cdots, x_{i_k}, \cdots, x_{j_k}, \cdots, x_{d_k}]^T, k \neq 0$ by random combining the sampled value x_i in the first-order component construction for each input variable (that is, x_i, $i = 1, \ldots, d$ and evaluated at $(x_i, \mathbf{x}_0^i)$, respectively). This new point is then evaluated by expensive simulation and the first-order RBF-HDMR model. The function values from expensive simulation and model prediction are compared. If the two values are sufficiently close (the relative error is less than a small value, for example, 0.01%), it indicates that no higher-order terms exist in the underlying function, the modeling process terminates. Otherwise, go to the next Step.

5. Use the values of x_i and x_j, $i \neq j$ that exist in thus-far evaluated points $(x_i, \mathbf{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, and $(x_j, \mathbf{x}_0^j) = [x_{1_0}, x_{2_0}, \cdots, x_j, \cdots, x_{d_0}]^T$ to form new points of the form $(x_i, x_j, \mathbf{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$. Randomly select one of the points from these new points to test the first-order RBF-HDMR model. If the model passes through the new point, it indicates that x_i and x_j are not correlated and the process continues with the next pair of input variables. This is to save the cost of

modeling non-existing or insignificant correlations. Otherwise, use this new point and the evaluated points $(x_i, \pmb{x}_0^i)$ and $(x_j, \pmb{x}_0^j)$ to construct the second-order component function, $\hat{f}_{ij}(x_i, x_j)$. This sampling-remodeling process iterates for all possible two-variable correlations until convergence (for instance, the relative prediction error is less than 0.01%). Step 5 is repeated for all pairs of input variables.

Theoretically, Step 5 applies to all higher-order terms in RBF-HDMR model, Eq. (3.10), in a similar manner. In this work, the process proceeds towards the end of the second-order terms based on the Premise 2 in Section 3.2. The identification of correlations in Steps 4 and 5 is supported by the discussions in Section 3.4.

3.5.3. An Example for Metamodeling RBF-HDMR
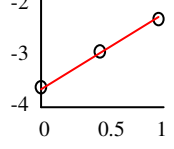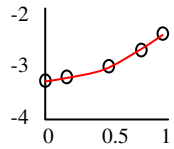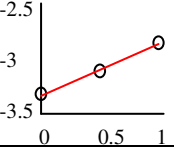
For the ease of understanding, consider the following mathematical function with d = 3,

$$f(\pmb{x}) = x_2^2 + x_1 x_3 + x_1 - 4, \quad 0 \le x_i \le 1. \tag{3.13}$$

Table 3.1 shows the modeling processes – finding f_0, modeling the first-order components $f_i(x_i), i = 1, \cdots, 3$, detecting and exploiting linearity and correlation relationship in the underlying function, and modeling higher-order components, $f_{ij}, i \ne j$, if they exist. This process continues until convergence. The first and last rows list the original function and the corresponding RBF-HDMR model, respectively. Each middle row demonstrates the modeling process in hierarchy. The detailed steps are as follows.

Step 1. Randomly sample the cut center $\boldsymbol{x}_0$ in the neighbourhood of the center of the design space, in this case, $\boldsymbol{x}_0 = [0.5, 0.5, 0.5]^T$ and then find $f_0 = f(\boldsymbol{x}_0)$.

Table 3.1 Process of modeling RBF-HDMR for the example problem

Function				$f(\boldsymbol{x}) = x_2^2 + x_1 x_3 + x_1 - 4,$ $\quad 0 \le x_i \le 1$		
f_0				$f(\boldsymbol{x}_0) = f([0.5, 0.5, 0.5]^T) = -3$		
$f_i(x_i)$	Linearity	Samples x_i	Observed $f(\boldsymbol{x})$	$f(x_i, \bar{x}^i) - f_0$	RBF coefs	Component Function Plot
	$f_1(x_1)$ (linear)	0 1 0.5	-3.75 -2.25 -3	-0.75 0.75 0	0 0 0 -0.75 1.5	
	$f_2(x_2)$ (non-linear)	0 1 0.5 0.19 0.81	-3.25 -2.25 -3 -3.2139 -2.5939	-0.25 0.75 0 -0.2139 0.4061	0.6796 0.6796 -0.1754 -0.5918 -0.5918 -0.3977 1	
	$f_3(x_3)$ (linear)	0 1 0.5	-3.25 -2.75 -3	-0.25 0.25 0	0 0 0 -0.25 0.5	
$f_{ij}(x_i, x_j)$	Correlation Relationship	Sampling $[x_i, x_j]^T$	Observed $f(\boldsymbol{x})$	$f(x_1, x_3, x_0^{ij}) - f_1(x_1)$ $- f_3(x_3) - f_0$		RBF coefs
	$f_{12}(x_1, x_2)$	null	null	null		null
	$f_{13}(x_1, x_3)$	(0.5,0.5); (0,0.5); (1,0.5); (0.5,0); (0.5, 1); (0,0); (1,1); (0,1); (1,0);	-3 -3.75 -2.25 -3.25 -2.75 -3.75 -1.75 -3.75 -2.75	0 0 0 0 0 0.25 0.25 -0.25 -0.25		0;0;0;0;0; 0.3607; 0.3607; -0.3607; -0.3607; 0;0;0
	$f_{23}(x_2, x_3)$	null	null	null		null

$\hat{f}(\boldsymbol{x})$

$\underbrace{-3}_{f_0} \underbrace{-0.75 + 1.5 x_1}_{f_1} +$

$\underbrace{0.6796 |x_2|^2 \log|x_2| + 0.6796 |x_2 - 1|^2 \log|x_2 - 1| - 0.1754 |x_2 - 0.5|^2 \log|x_2 - 0.5| - 0.5918 |x_2 - 0.19|^2 \log|x_2 - 0.19| - 0.5918 |x_2 - 0.81|^2 \log|x_2 - 0.81| - 0.3977 + x_2}_{f_2}$

$\underbrace{-0.25 + 0.5 x_3}_{f_3}$

$\underbrace{+ 0.3607 \left|\frac{x_1}{x_3}\right|^2 \log\left|\frac{x_1}{x_3}\right| + 0.3607 \left|\frac{x_1 - 1}{x_3 - 1}\right|^2 \log\left|\frac{x_1 - 1}{x_3 - 1}\right| - 0.3607 \left|\frac{x_1 - 0}{x_3 - 1}\right|^2 \log\left|\frac{x_1 - 0}{x_3 - 1}\right| - 0.3607 \left|\frac{x_1 - 1}{x_3 - 0}\right|^2 \log\left|\frac{x_1 - 1}{x_3 - 0}\right|}_{f_{13}}$

Step 2. Randomly sample x_i at its two ends, and form two new points $(x_i, \boldsymbol{x}_0^i)$ (one per end) and evaluate them. Model the component functions using the two end points. For example, for x_1, two values 0 and 1 are sampled; we can use the function values [f (0, 0.5, 0.5)-f_0] and [f (1, 0.5, 0.5)-f_0] as their function values for 0 and 1, respectively to model $f_1(x_1)$. Note that the special RBF format as in the 3.9 Appendix is used rather than the simple linear spline format to avoid matrix singularity.

Step 3. Detect the linearity of the output function with respect to each variable x_i by comparing $f(\boldsymbol{x}_0)$ and $\hat{f}(\boldsymbol{x}_0)$. If nonlinearity exists, model $f_i(x_i)$, $i = 1, \cdots, 3$ until convergence. The $f_2(x_2)$ component function, for instance, is a nonlinear function. In addition to the two end points 0 and 1, two more new sample points are generated at 0.19 and 0.81 to capture its nonlinearity. All of the component functions are plotted with a distance f_0 in the last column in Table 3.1.

Step 4. Identify if correlation exists. If no, terminate modeling. Otherwise, go to Step 5. In this case, since there exists correlation between x_1 and x_3, the modeling process continues.

Step 5. Identify the correlated terms according to Step 5 of the algorithm described in Section 3.5.2. If correlation exists in the underlying function, model the identified correlated terms. In this case, only the correlated term $x_1 x_3$ exists, which needs to be modeled as a component function. Repeat Step 5 until convergence.

To better understand Table 3.1, let us model $f_2(x_2)$ as an example. When modeling $f_2(x_2)$, five values along x_2 are generated, that is, 0, 1, 0.5, 0.19, 0.81, according to the sampling algorithm described in Section 3.5.2. By combining these

values with other x_0 component values except for x_2, we have five new points and their function values are evaluated. Deducting the f_0 value from their function values, we obtain the component function $f_2(x_2)$ values for these five points are, respectively, -0.25, 0.75, 0.0, -0.2139, 0.4061. Employing these five points and their function values to fit the RBF function as defined in the 3.9 Appendix, one can have the RBF model for $f_2(x_2)$, with five nonlinear terms and two polynomial terms, as shown in the last row of Table 3.1.

In Table 3.1, $f_3(x_3)$ is especially noteworthy. One can see from the original function expression that there is no separate x_3 term in the function, but an $x_1 x_3$ term. Why is $f_3(x_3)$ not zero? This is because HDMR first finds the first-order influence of x_3, the residual then goes to the second-order component function $f_{13}(x_1, x_3)$. Therefore, it would be wrong to mechanically match the component functions with the terms in the underlying function. As a matter of fact, the $x_1 x_3$ term in the original function has been expressed by $f_3(x_3)$ and $f_{13}(x_1, x_3)$, as well as partially by $f_1(x_1)$.

Figure 3.1 shows the distribution of all sample points in the modeling space. It can be seen that most sampled points are located on the lines and planes across the cut center x_0. Points x_{12}, x_{13}, x_{23} and x_{123} were used to identify the correlation among the variables, respectively between x_1 and x_2, x_1 and x_3, x_2 and x_3, as well as the existence of correlations among all variables x_1, x_2 and x_3 of the underlying function. It is to note that these sample points are generated as needed according to the aforementioned sampling and model construction method.
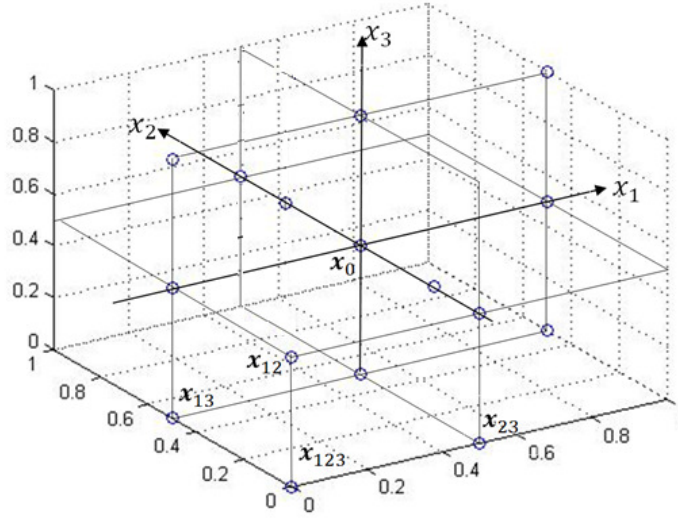
Fig. 3.1 Distribution of sample points for the example problem

Given the metamodel as expressed in the last row of Table 3.1, one can observe that all first-order functions are linear except for f_2, which indicates that x_2 has a nonlinear influence to the overall f function while others have linear effects. For the second-order components, only a nonlinear f_{13} is present, indicating other variables are not correlated.

3.6. Testing of RBF-HDMR

Several problems from the literature are used for testing the proposed RBF-HDMR and its metamodeling method. The modeling efficiency is indicated by the number of (expensive) sample points. The modeling accuracy is evaluated by three performance metrics. A comparison of the RBF-HDMR model with other metamodels is also given.

3.6.1. Performance Metrics

There are various commonly-used performance metrics for approximation models that are given in [22]. To the authors' knowledge, however, there are no specially defined performance metrics for high-dimensional approximation models in the literature. In

mathematics, where the high-dimensional problems are mostly (and yet not adequately) studied, the percentage relative error is often used as a metric for model validation. It is found, however, when the absolute errors of the metamodels are quite small, their percentage relative errors could be extremely large when the function value is close to zero. The percentage relative error measure is also dependent on the problem scale, which makes the comparison between problems disputable. In the engineering design, the cross-validation method is currently a popular method for model validation. However, Lin [23] found that the cross-validation is an insufficient measurement for metamodel accuracy. The cross-validation is actually a measurement for degrees of insensitivity of a metamodel to lost information at its data points, while an insensitive metamodel is not necessarily accurate. To be consistent with Ref. [5], which will be used for result comparison, this work employs three commonly used performance metrics —R square, relative average absolute error (RAAE) and relative maximum absolute error (RMAE) — for validating approximation models. After the RBF-HDMR modeling process is terminated, additional 10,000 new random sample points are used to evaluate the model against the three performance metrics by Monte Carlo simulation. The values of these performance metrics show the prediction capability of the RBF-HDMR on new points. It is to be noted that these three metrics all need a fairly large number of validation points to be meaningful, but for High-dimensional, Expensive, Black-box (HEB) problems such information are too costly to obtain. This is in contrast to high-dimensional problems studied in mathematics where those are inexpensive problems and a large quantity of validation points is affordable. Validation methodology for HEB problems is therefore worth further research. Since this work also chose mathematical problems for testing and

104

comparison, we can still employ the three metrics with Monte Carlo simulations for validation. These metrics are described as below:

1)   R Square

$$R^2 = 1 - \frac{\sum_{i=1}^{m}[f(x_i) - \hat{f}(x_i)]^2}{\sum_{i=1}^{m}[f(x_i) - \bar{f}(x_i)]^2},$$ (3.14)

where $\bar{f}(x_i)$ denotes the mean of function on *m* sample points. This metrics indicates the overall accuracy of an approximation model. The closer the value of R square approaches one, the more accurate is the approximation model. Note that *R* square is evaluated on the new validation points, not on the modeling points. The same is true for RAAE and RMAE.

2)   Relative Average Absolute Error (RAAE)

$$RAAE = \frac{\sum_{i=1}^{m}|f(x_i) - \hat{f}(x_i)|}{m * STD},$$ (3.15)

where *STD* stands for standard deviation. Like R square, this metric shows the overall accuracy of an approximation model. The closer the value of RAAE approaches zero, the more accurate is the approximation model.

3)   Relative Maximum Absolute Error (RMAE)

$$RMAE = \frac{\max(|f(x_1) - \hat{f}(x_1)|, |f(x_2) - \hat{f}(x_2)|, \cdots, |f(x_m) - \hat{f}(x_m)|)}{STD}.$$ (3.16)

This is a local metric. A RMAE describes error in a sub-region of the design space. Therefore, a small value of RMAE is preferred.

3.6.2. Study Problem

A problem for large-scale optimization in Matlab$^{\text{TM}}$ optimization toolbox is chosen to study the performance of RBF-HDMR and its metamodeling method as a function of the dimensionality, d.

$$f(x) = \sum_{i=1}^{d-1}[(x_i^2)^{(x_{i+1}^2+1)} + (x_{i+1}^2)^{(x_i^2+1)}], \ 0 \leq x_i \leq 1. \quad\quad (3.17)$$

This highly nonlinear problem was tested with d=30, 50, 100, 150, 200, 250 and 300. For each d, ten runs have been taken and the mean values of R square, RAAE and RMAE are charted in Fig. 3.2.

a) Mean of R^2 b) Mean of RAAE   c) Mean of RMAE
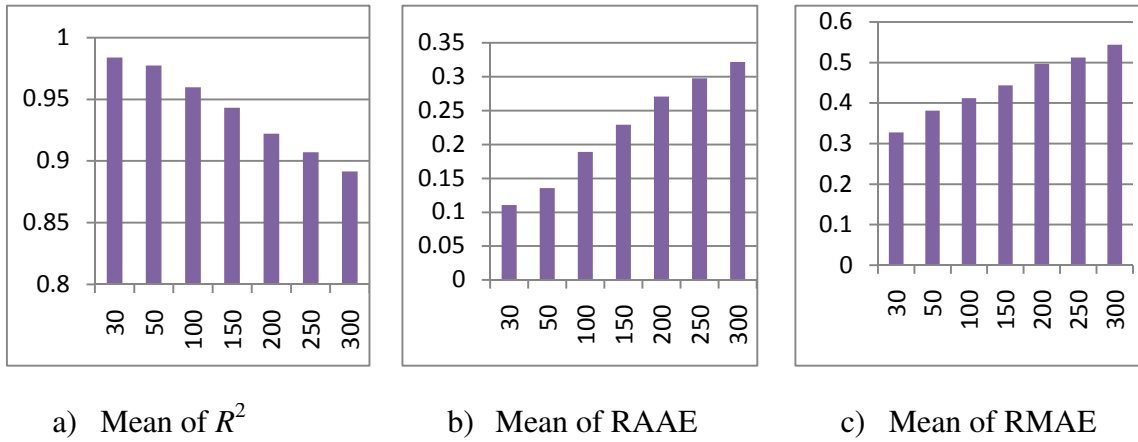
Fig. 3.2 Performance metrics mean with respect to d (x-axis) for the study problem

As seen from Fig. 3.2, although the three accuracy performance metrics deteriorate slightly as d increases, they demonstrate that the RBF-HDMR model fits well the high-dimensional underlying function. The minimum (worst) value of R square is close to 0.9, the maximum (worst) values of RAAE is about 0.32 and the maximum (worst) value of RMAE is about 0.54. The data explains the RBF-HDMR model has a good fit of the underlying function.

Regarding modeling cost, assuming five samples are taken along each axes ($s = 5$), we list the number of evaluations for RBF-HDMR model, full second-order expansion of the HDMR (polynomially increasing cost), and the full factorial design of experiments (exponentially increasing cost) for various dimensionality d in Table 3.2. The comparison clearly shows the computational advantage of the proposed RBF-HDMR. The efficiency of the proposed method will be further studied in comparison with Latin Hypercube samples in the next section.

Table 3.2 Comparison of modeling cost for the study problem

d	Cost of RBF-HDMR (second-order)	Cost of a full second-order expansion of HDMR $c = 1 + d(s-1)$ $+ \dfrac{d(d-1)}{2!}(s-1)^2$ (polynomial)	Cost of full factorial design s^d (exponential)
30	819	7081	9.31×10^{20}
50	1830	19801	8.88×10^{34}
100	6116	79601	7.88×10^{69}
150	12807	179401	7.01×10^{104}
200	22042	319201	6.22×10^{139}
250	33762	499001	5.53×10^{174}
300	47979	718801	4.91×10^{209}

3.6.3. Testing and Comparison with Other Metamodels

In order to test the effectiveness of various models (MARS, RBF, Kriging, and polynomial), Jin *et al.* [5] selected 14 problems which are classified into two categories: large scale and small scale. The large scale includes one 14-variable application, one 16-variable, and four 10-variable problems. The small scale includes five two-variable problems and three three-variable problems, among which one was repeated by adding some noise to form a new problem. Therefore, in total Ref. [5] gives 13 unique

problems, twelve are tested with RBF-HDMR except for the 14-variable application

problem, which is unavailable to authors. Since this work deals with high-dimensional

problems, only the test results for the large scale problems (Problems 1-5) are reported in

Table 3.3 with the first- and second-order RBF-HDMR models. These problems are listed

in the 3.9 Appendix. In our test, each problem runs ten times independently for

robustness testing, and then takes the average of ten runs for each problem. In Table 3.3,

NoE indicates the number of all evaluated expensive points, which include modeling

points and detecting points used for correlation identification. The NoE for the second-

order modeling includes the NoE for the first-order modeling.

Table 3.3 Modeling results for the test suite

Problem		R Square	RAAE	RMAE	NoE
1	First	0.90	0.233	1.66	95
(d=10)	Second	0.92	0.211	1.16	321
2	First	1.00	0.006	0.04	40
(d=10)	Second	1.00	0.006	0.02	232
3	First	0.99	0.049	0.59	121
(d=10)	Second	0.96	0.129	1.16	408
4	First	0.98	0.110	0.33	34
(d=10)	Second	0.98	0.107	0.28	40
5	First	0.96	0.150	0.91	59
(d=16)	Second	0.99	0.088	0.25	297
Mean	First	0.97	0.109	0.71	70
	Second	0.97	0.113	0.65	250

It is can be seen from Table 3.3 that all results of the first-order RBF-HDMR are

good enough for large scale problems, even though Problems 1 and 3 are highly

nonlinear. The results of the second-order models show slight improvement over the

first-order models for all the problems except for Problem 3, which indicates a certain

degree of over-fitting. Theoretically when convergence criteria or numeric tolerance for

nonlinearity and correlation identification is sufficiently tight, the second-order model should be more accurate than the first-order. In practice, however, the errors in nonlinearity and correlation identification and RBF model construction may be larger than the influence of higher-order components. In such circumstances, over-fitting of RBF may occur.

To understand the test results in Table 3.3, we compare the results with those from Ref. [5] in Fig. 3.3. Ref. [5] used three different Latin Hypercube Design (LHD) sample sets for the large scale problems. Their average numbers are 34, 112, and 250 for scarce, small, and large data sets, respectively. From Fig. 3.3, one can see that the first-order RBF-HDMR modeling requires a data set of a size falling in between those of the scarce and small data sets.
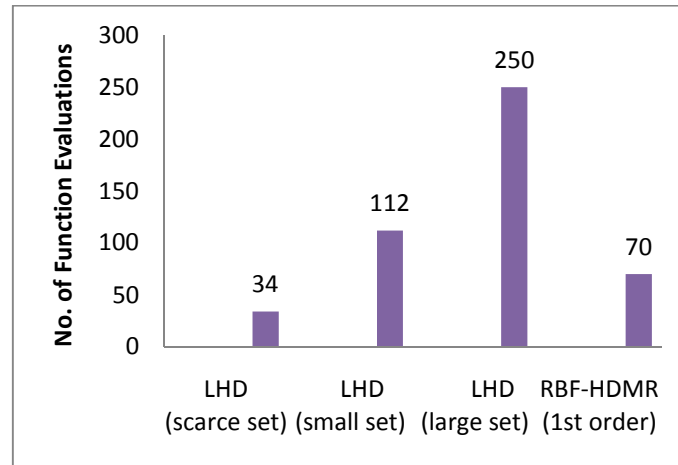

Fig. 3.3 Comparison of NoE with Latin Hypercube points from Reference [5]

Figure 3.4 shows the mean value of the same three metrics — R square, RAAE and RMAE — for the aforementioned four models applied to test problems. From Fig. 3.4, it can be seen that while the mean R square for RBF-HDMR models is 0.97, the maximum

(best) value of the four models in the reference is about 0.78. Because the exact RAAE and RMAE data for only the large-scale problems are not available in Ref. [5], we use the mean for all 14 test problems as a comparison. The RAAE and RMAE values for the 14 problems should be lower than that for the large scale problems alone. Even with these comparison values, RBF-HDMR has much smaller RAAE and RMAE values. It is also to be noted that the accuracy data from Ref. [5] is based on the average results for all data sets. The highest R square value for large sample sets (250 points) for the four models is slightly above 0.90, still significantly lower than 0.97, which is the R square value of RBF-HDMR with only 70 points.
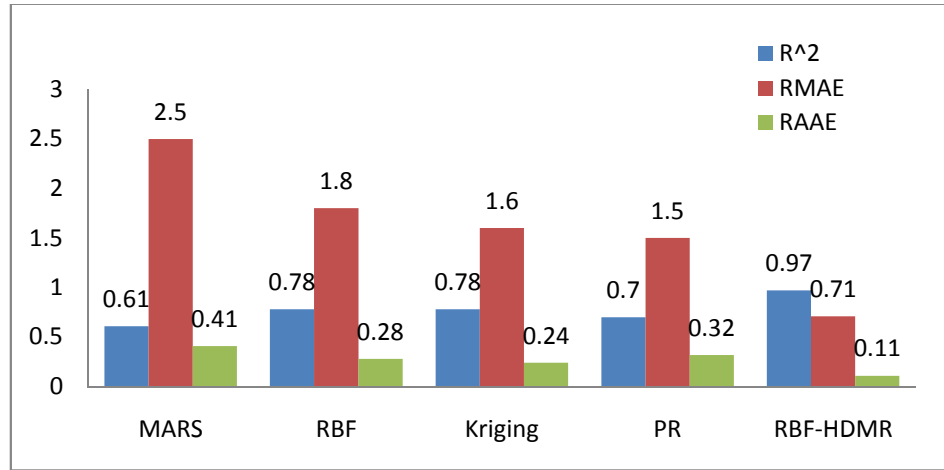

Fig. 3.4 Model accuracy comparison. Data for models other than RBF-HDMR are from Ref. [5]; R^2 values are for large-scale problems only, while RMAE and RAAE values are for all 14 test problems.

In summary, from the comparison with the reference, the proposed RBF-HDMR model and its metamodeling method generates more accurate models with fewer sample

points than conventional models such as MARS, RBF, Kriging, and polynomial functions with Latin Hypercube designs.

3.6.4. Discussion

This work employs RBF to model component functions of the HDMR, so that HDMR is no longer a check-up table but rather a complete equation. The proposed metamodeling approach takes advantages of properties of HDMR to make the sampling efficient. RBF was chosen because of 1) its simplicity and robustness in model construction 2) the ease of obtaining an explicit function expression, and 3) its ability to interpolate the sample points (this could be a problem for noisy data, which is a topic of our future research). The integration of HDMR with the interpolative feature of RBF supports the developed lemma and the sampling method, especially on identification of nonlinearity, variable correlations, and higher-order components; therefore RBF helps to reduce the sample size. The choice of a specific RBF form, as shown in the 3.9 Appendix, is deliberate as it is better than a simple linear spline for avoiding singularity. Exploration of other interpolative metamodels and selection of the best metamodel for component functions may be a topic for future research.

The proposed metamodeling approach takes advantage of the hierarchical structure of HDMR, adaptively models its component functions while exploring its inherent linearity/nonlinearity and correlation among variables. The sample points are thus limited and well controlled. The realized samples spread in the design space (refer to Fig. 3.1), but unevenly, according to complexity of regions in the space. Regions of high nonlinearity or correlation will have more sample points while linear regions have fewer

points, all according to the needs to capture the behavior of the underlying function. In contrast, the Latin Hypercube Design (LHD) has only one-dimensional uniformity and it is "blind" to the function characteristics. It is also worth noting that the metamodeling process only involves fast and simple algebraic operations, which also lends itself well for parallel computation at each order of component levels. The outputs are multitude, that is, an explicit RBF-HDMR model, function linearity/non-linearity, correlations among variables, and so on.

The RBF-HDMR at current stage, however, only models to the second-order components. If an underlying function has significant multivariate correlation, the method may be limited. New approaches are needed to extend beyond the second-order, whereas keeping the modeling cost low. Secondly, the proposed RBF-HDMR adaptively determines the location of sample points, which is attractive when there is no existing data and the goal is to reduce the number of function evaluations. In real practice, however, there are situations that some expensive data may have already been generated. Strategies needed to be developed to take advantage of the existing data when constructing RBF-HDMR. Thirdly, RBF-HDMR at its current stage only deals with deterministic problems while in practice the expensive model may be noisy. Future research is needed to deal with these issues.

3.7. Conclusion

This work proposes a methodology of metamodeling High-dimensional, Expensive, and Black-box (HEB) problems. The methodology consists of the proposed RBF-HDMR metamodel, and its accompanying metamodeling method. The RBF-HDMR model

inherits hierarchical structural properties of HDMR, provides an explicit expression with RBF components, and needs neither any knowledge about the underlying functions nor assumes *a priori* a parametric form. The modeling process automatically explores and uses the properties of the underlying functions, refines the model accuracy by iterative sampling in the subspace of nonlinearity and correlated variables, and involves only fast and simple algebraic operations that can be easily parallelized. The developed methodology alleviates or circumvents the "curse of dimensionality." Testing and comparison with other metamodels reveal that RBF-HDMR models high-dimensial problems with higher efficiency and accuracy. Future research aims to extend the modeling approach to efficiently model high-order components, to use existing data, and to deal with noisy samples.

3.8. Acknowledgment

3.9. Appendix

1. RBF model

A general radial basis functions (RBF) model [2, 5] is shown as

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \beta_i \varphi(|\boldsymbol{x} - \boldsymbol{x}_i|), \tag{A.1}$$

where β_i is the coefficient of the expression and $\boldsymbol{x}_i$ are the sampled points of input variables or the centers of RBF approximation. $\varphi(.)$ is a distance function or the radial

basis function. $|.|$ denotes a p-norm distance. A RBF is a real-valued function whose value depends only on the distance from center points x_i. It employs linear combinations of a radically symmetric function based on the distance to approximate underlying functions. Its advantages include: the number of sampled points for constructing approximation can be small and the approximations are good fits to arbitrary contours of response functions [2]. Consequently, RBF is a popular model for multivariate data interpolation and function approximations.

The key of RBF approach is to choose a p-norm and a radial basis function $\phi(.)$, both of which have multiple formats. One of the goals for choosing a format is to make the distance matrix $(A_{ij} = \varphi(|x_i - x_j|))$, for $1 \leq i, j \leq n$, n is the number of sample points) non-singular. The singularity of the distance matrix relates to the distribution of the sample points. It can be seen that there are many works on choosing a p-norm and a radial basis function $\varphi(.)$ to avoid the singularity of the distance matrix [29]. This research uses a sum of thin plate spline (the first term) plus a linear polynomial $P(x)$ (the second term),

$$\hat{f}(x) = \sum_{i=1}^{n} \beta_i \, |x - x_i|^2 log|x - x_i| + P(x),$$

$$\sum_{i=1}^{n} \beta_i \, p(x) = 0, \ P(x) = p\alpha = [p_1, p_2, \cdots, p_q][\alpha_1, \alpha_2, \cdots, \alpha_q]^T, \quad \text{(A.2)}$$

where x_i are the vectors of evaluated n sample points; the coefficients $\beta = [\beta_1, \beta_2, \cdots, \beta_n]$ and α are parameters to be found. $P(x)$ is a polynomial function, where p consists of a vector of basis of polynomials. In this work, p is chosen to be $(1, x_1, \cdots, x_d)$ including

114

only linear variable terms and therefore $q=d+1$; The side condition $\sum_{i=1}^{n} \beta_i \, \boldsymbol{p}(\boldsymbol{x}) = 0$ is imposed on the coefficients $\boldsymbol{\beta}$ to improve an under-determined system, that is, the singularity of distance matrix $\boldsymbol{A}$ [29]. To calculate the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, Eq. (A.2) may be written in the matrix form as

$$\begin{bmatrix} \boldsymbol{A} & \widetilde{\boldsymbol{P}} \\ \widetilde{\boldsymbol{P}}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{0} \end{bmatrix},$$ (A.3)

where, $A_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|^2 \log |\boldsymbol{x}_i - \boldsymbol{x}_j|, \;\; i, j = 1, \cdots, n,$

$\widetilde{P}_{ij} = p_j(\boldsymbol{x}_i), \;\; i = 1, \cdots, n; \;\; j = 1, \cdots, (d+1);$

and $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the vectors of evaluated n sample points. The theory guarantees the existence of a unique vector $\boldsymbol{\beta}$ and a unique polynomial $P(\boldsymbol{x})$ satisfying Eq.(A.2) [29].

2. Test Problems

No.	Function	Variable Ranges
1	$$f(\boldsymbol{x}) = \sum_{i=1}^{10} [[\ln(x_i - 2)]^2 + [\ln(10 - x_i)]^2] - (\prod_{i=1}^{10} x_i)^{0.2}$$	$2.1 \leq x_i \leq 9.9$, $i = 1, \cdots, 10$
2	$$f(\boldsymbol{x}) = \sum_{i=1}^{10} x_i(c_i + ln\frac{x_i}{x_1 + \cdots + x_{10}})$$	$1E - 6 \leq x_i \leq 10, i = 1, \cdots, 10$
3	$$f(\boldsymbol{x}) = \sum_{i=1}^{10} \exp(x_i)\,[c_i + x_i - \ln\,(\sum_{k=1}^{10} \exp(x_k))]$$	$-10 \leq x_i \leq 10$, $i = 1, \cdots, 10$
4	$$f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_1 x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$$	$-10 \leq x_i \leq 11$, $i = 1, \cdots, 10$
5	$$f(\boldsymbol{x}) = \sum_{i=1}^{16} \sum_{j=1}^{16} a_{ij}(x_i^2 + x_i + 1)(x_j^2 + x_j + 1)$$	$0 \leq x_i, x_j \leq 5$, $i, j = 1, \cdots, 16$

For Prob. 2 and 3

$c_{i=1,\cdots,10}$
$= -6.089, -17.164, -34.054, -5.914, -24.721, -14.986, -24.100, -10.708, -26.662, -22.179$

For Prob. 5

$$[a_{ij}]_{row1-8} = \begin{bmatrix} 1\,0\,0\,1\,0\,0\,1\,1\,0\,0\,0\,0\,0\,0\,0\,1 \\ 0\,1\,1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0\,1\,0\,1\,1\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,1\,0\,1\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,0\,0\,0\,1\,0 \end{bmatrix}$$

$$[a_{ij}]_{row9-16} = \begin{bmatrix} 0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix}$$

3.10. References

[1] Wang, G. G. and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," Transactions of the ASME, Journal of Mechanical Design, 129, pp. 370-389.

[2] Chen, V., C. P., Tsui, K.-L., Barton, R. R. and Meckesheimer, M., 2006, "A Review on Design, Modeling and Applications of Computer Experiments," IIE Transactions, 38, pp. 273-291.

[3]   Simpson, T. W., Toropov, V., Balabanov, V., Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come – or Not," 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference 10 - 12 September 2008, Victoria, British Columbia Canada, AIAA 2008-5802.

[4]   Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D., 1992, "Screening, Predicting, and Computer Experiments," Technometrics, 34(1), pp. 15-25.

[5]   Jin, R., Chen, W. and Simpson, T. W., 2001, "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria," Structural and Multidisciplinary Optimization, 23(1), pp. 1-13.

[6]   Shorter, J. A., Ip, P. C. and Rabitz, H. A., 1999, "An Efficient Chemical Kinetics Solver Using High Dimensional Model Representation," Journal of Physical Chemistry A, (103), pp. 7192-7198.

[7]   Bates, R. A., Buck, R. J., Riccomagno, E. and Wynn, H. P., 1996, "Experimental Design and Observation for Large Systems," Journal of the Royal Statistical Society: B, 58(1), pp. 77-94.

[8]   Booker, A. J., Dennis, J. E. J., Frank, P. D., Serafini, D. B., Torczon, V. and Trosset, M. W., 1999, "A Rigorous Framework for Optimization of Expensive Functions by Surrogates," Structural Optimization, 17(1), pp. 1-13.

[9]   Koch, P. N., Simpson, T. W., Allen, J. K. and Mistree, F., 1999, "Statistical Approximations for Multidisciplinary Design Optimization: The Problem of Size," Journal of Aircraft, 36(1), pp. 275-286.

[10] Tu, J. and Jones, D. R., 2003, "Variable Screening in Metamodel Design by Cross-Validated Moving Least Squares Method," 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference Norfolk, Virginia, April 7-10.

[11] Friedman, J. H. and Stuetzle, W., 1981, "Projection Pursuit Regression," Journal of the American Statistical Association, 76(372), pp. 817-823.

[12] Friedman, J. H., 1991, "Multivariate Adaptive Regressive Splines," The Annals of Statistics, 19(1), pp. 1-67.

[13] Sobol, I. M., 1993, "Sensitivity Estimates for Nonlinear Mathematical Models," Mathematical Modeling & Computational Experiment, 1(4), pp. 407-414.

[14] Li, G., Wang, S.-W., Rosenthal, C. and Rabitz, H., 2001, "High Dimensional Model Representations Generated from Low Dimensional Data Samples. I. Mp-Cut-HDMR," Journal of Mathematical Chemistry, 30(1), pp. 1-30.

[15] Rabitz, H. and Al1s, Ö. F., 1999, "General Foundations of High-Dimensional Model Representations," Journal of Mathematical Chemistry (25), pp. 197-233.

[16] Li, G., Rosenthal, C. and Rabitz, H., 2001, "High Dimensional Model Representations," Journal of Physical Chemistry A, 105(33), pp. 7765-7777.

[17] Li, G., Hu, J., Wang, S.-W., Georgopoulos, P. G., Schoendorf, J. and Rabitz, H., 2006, "Random Sampling-High Dimensional Model Representation (RS-HDMR) and Orthogonality of Its Different Order Component Functions," Journal of Physical Chemistry A, (110), pp. 2474-2485.

[18] Wang, S.-W., Georgopoulos, P. G., Li, G. and Rabits, H., 2003, "Random Sampling-High Dimensional Model Representation (RS-HDMR) with

Nonuniformly Distributed Variables: Application to an Integrated Multimedia/Multipathway Exposure and Dose Model for Trichloroethylene," Journal of Physical Chemistry A, (107), pp. 4707-4716.

[19] Tunga, M. A. and Demiralp, M., 2005, "A Factorized High Dimensional Model Representation on the Nodes of a Finite Hyperprismatic Regular Grid," Applied Mathematics and Computation (164), pp. 865-883.

[20] Tunga, M. A. and Demiralp, M., 2006, "Hybrid High Dimensional Model Representation (HDMR) on the Partitioned Data," Journal of Computational and Applied Mathematics, (185), pp. 107-132.

[21] Sobol, I.M., 2003, "Theorems and examples on high dimensional model representation," Reliability Engineering and System Safety 79, pp. 187–193.

[22] Shan, S. and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-box Functions," Structural and Multidisciplinary Optimization, 41(2), pp. 219-241 DOI: 10.1007/s00158-009-0420-2.

[23] Lin, Y., 2004, "An Efficient Robust Concept Exploration Method and Sequential Exploratory Experimental Design," Mechanical Engineering, Georgia Institute of Technology, Atlanta, Ph. D. Thesis.

[24] Jin, R., Chen, W. and Sudjianto, A., 2005, "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments," Journal of Statistical Planning and Inferences, 134(1), pp. 268-287.

[25] Sasena, M., Parkinson, M., Goovaerts, P., Papalambros, P. and Reed, M., 2002, "Adaptive Experimental Design Applied to an Ergonomics Testing Procedure,"

ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference, ASME, Montreal, Canada.

[26] Wang, G. G., 2003, "Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points," Transactions of the ASME, Journal of Mechanical Design, 125, pp. 210-220.

[27] Jin, R., Chen, W. and Sudjianto, A., 2002, "On Sequential Sampling for Global Metamodeling in Engineering Design," ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference, Montreal, Canada.

[28] Owen, A. B., 2000, "Assessing Linearity in High Dimensions," The Annals of Statistics, 28(1), pp. 1-19.

[29] Baxter, B. J. C., 1992, "The Interpolation Theory of Radial Basis Functions," Trinity College, Cambridge University, Ph. D. Thesis.

Chapter 4

Turning Black-box into White Functions[3]

Songqing Shan

Department of Mech. and Manuf.

Engineering

University of Manitoba

Winnipeg, MB, Canada R3T 5V6

G. Gary Wang[4]

School of Engineering

Science

Simon Fraser University

Surrey, BC. Canada V3T 0A3

---

[3] A revised version was submitted to Transactions of the ASME, Journal of Mechanical Design.
[4] Corresponding author, Tel: 778 782 8495, Fax: 778 782 8514, Email: gary_wang@sfu.ca

4.1. Abstract

To build a metamodel for high-dimensional expensive black-box (HEB) functions, a recently developed metamodel, radial basis function-based high-dimensional model representation (RBF-HDMR), has been found promising. This work extends the modeling capability of RBF-HDMR from the current second-order form to any higher order. More importantly, the modeling process "uncovers" black-box functions so that not only a more accurate metamodel is obtained, but also key information of the function can be gained and thus the black-box function can be turned "white." The key information that can be gained includes 1) functional form, 2) (non)linearity with respect to each variable, 3) variable correlations. The black-box "uncovering" process is based on identifying the existence of certain variable correlations through two derived theorems. The adaptive process of exploration and modeling reveals the black-box functions until all significant variable correlations are found. The black-box functional form is then represented by a structure matrix that can manifest all orders of correlated behavior of variables. The resultant metamodel and its revealed inner structure lend themselves well for applications such as sensitivity analysis, decomposition, visualization, and optimization. The proposed approach is tested with theoretical and practical examples. The test result demonstrates the effectiveness and efficiency of the proposed approach.

Key words: approximation, regression, interpolation, metamodel, response surface, prediction, RBF-HDMR, functional form, black-box function

4.2. Introduction

Metamodeling techniques find a wide range of uses in engineering. These uses include getting insight into a complex system and supporting simulation-based optimization. Metamodeling techniques involve sampling approaches, model selection, model fitting, and model validations [1-3]. From the sampling perspective, there are various sampling approaches: one-stage sampling (for example, Latin hypercube and orthogonal arrays), optimal sampling (for instance, D-optimal and G-optimal), one-stage-based optimal sampling (for example, optimal Latin hypercube, and optimal orthogonal-arrays-based Latin hypercube) and sequential sampling [1-4]. Sampling approaches evolve from static to dynamic approaches, from one-stage to optimal one-stage, and to sequential sampling. For metamodel selection and fitting, there are parametric models (for instance, polynomial) and non-parametric models (for example, radial basis function) [3]. For metamodel validation, there are various validating approaches and performance metrics such as relative error and R square [5]. There are multiple papers reviewing the advancement of the metamodeling techniques. Chen *et al.* [6] summarized the pros and cons of the sampling methods and metamodels. Queipo *et al.* [7] reflected the metamodeling techniques and optimization. Simpson *et al.* [8] reviewed the use of metamodeling techniques in multidisciplinary analysis and optimization. Wang and Shan [9] reviewed applications of metamodeling techniques in support of engineering optimization.

With the advancement of metamodeling techniques, two issues become prominent: 1) the "curse of dimensionality," and 2) how to gain more insight into a black-box through metamodeling. The two issues in fact interweave with each other. The "curse of

dimensionality" indicates the metamodeling cost (the number of function evaluations) exponentially increases as the dimensionality of black-box functions become larger. Koch *et al.* [10] presented the size problem of the black-box functions for multidisciplinary design optimization. Simpson *et al.* [11] pointed out that the high-dimensionality plagues metamodeling techniques. Shan and Wang [5] reviewed the relevant methodologies in solving HEB problems. Friedman and Stuetzle [12] developed projection pursuit regression. Friedman [13] proposed multivariate adaptive regression splines (MARS) model. A family of HDMRs with distinct characters has since been developed for various purposes [14-19]. Recently Shan and Wang [20] proposed a RBF-HDMR model and its modeling algorithm to approximate a black-box function until the second-order terms. For the second issue on gaining more insight into the black-box function, today, almost all the metamodels only provide the metamodel as a predictor and lack the capability to reveal the underlying functional form of the black-box function. This is particularly the case for commonly-used metamodels such as Kriging, support vector machine, and radial basis function (RBF). Few papers have addressed this issue, especially from the metamodeling community. For instance, Booker [21] used functional ANOVA techniques in conjunction with fitted Kriging model to disclose main effects and correlation relationships in the response. Hooker [22] from statistics developed an ANOVA approach to discover additive structure in black-box functions.

This chapter naturally advances the RBF-HDMR model to discover the intrinsic structure in black-box functions and enhance the accuracy of modeling by adaptively modeling higher-order terms beyond the current second-order form of RBF-HDMR. While the accuracy of the RBF-HDMR approximation provided by the first and second-

124

order is not sufficient, the developed approach explores high-order correlated terms, and continues to model the residual terms. Thus the accuracy of the RBF-HDMR can be further improved as far as the budget allows. Moreover, through the modeling process, key information of a black-box function such as the functional form, (non)linearity, and variable correlations can be revealed.

Section 4.3 introduces the basics of RBF-HDMR. Section 4.4 describes the modeling process of RBF-HDMR approach. Section 4.5 presents structure matrix and component correlation matrix, and theories that support identification of high-order component terms. Section 4.6 provides test results and discussion. The final remarks are in Section 4.7.

4.3. RBF-HDMR

A general form of HDMR [16] is shown as

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i < j \le d} f_{ij}(x_i, x_j) + \cdots + \sum_{1 \le i_1 < \cdots < i_l \le d} f_{i_1 i_2 \cdots i_l}(x_{i_1}, x_{i_2}, \cdots, x_{i_l})$$

$$\cdots + f_{12 \cdots d}(x_1, x_2, \cdots, x_d), \tag{4.1}$$

where the component f_0 is a constant representing the zero-th order effect to $f(\boldsymbol{x})$; the component function $f_i(x_i)$ gives the effect of the variable x_i acting independently upon the output $f(\boldsymbol{x})$ (the first-order effect), and may have an either linear or non-linear dependence on x_i. The component function $f_{ij}(x_i, x_j)$ describes the correlated contribution of the variables x_i and x_j upon the output $f(\boldsymbol{x})$ (the second-order effect)

after the individual influences of x_i and x_j are discounted, and $f_{ij}(x_i, x_j)$ could be linear or nonlinear as well. The subsequent terms reflect the effects of increasing numbers of correlated variables acting together upon the output $f(\boldsymbol{x})$. The last term $f_{12\cdots d}(x_1, x_2, \cdots x_d)$ represents any residual dependence of all the variables locked together to influence the output $f(\boldsymbol{x})$ after all of the lower-order correlation and individual influence of each involved x_i ($i = 1,\ldots,d$) have been discounted.

In order to compute component functions in Eq. (4.1), the simplest and most efficient type, Cut-HDMR [16], is explained here. For a chosen cutting center point $\boldsymbol{x}_0$, component functions of the Cut-HDMR are defined as

$$f_0 = f(\boldsymbol{x}_0), \tag{4.2}$$

$$f_i(x_i) = f(x_i, \boldsymbol{x}_0^i) - f_0, \tag{4.3}$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \boldsymbol{x}_0^{ij}) - f_i(x_i) - f_j(x_j) - f_0, \tag{4.4}$$

$$f_{ijk}(x_i, x_j, x_k) = f(x_i, x_j, x_k, \boldsymbol{x}_0^{ijk}) - f_{ij}(x_i, x_j) - f_{ik}(x_i, x_k) - f_{jk}(x_j, x_k) - f_i(x_i) -$$
$$f_j(x_j) - f_k(x_k) - f_0, \tag{4.5}$$

$$\cdots,$$

$$f_{12\cdots d}(x_1, x_2, \cdots, x_d) = f(\boldsymbol{x}) - f_0 - \textstyle\sum_i f_i(x_i) - \textstyle\sum_{ij} f_{ij}(x_i, x_j) - \cdots, \tag{4.6}$$

where $\boldsymbol{x}_0^i$, $\boldsymbol{x}_0^{ij}$, and $\boldsymbol{x}_0^{ijk}$ are respectively $\boldsymbol{x}_0$ without elements x_i; x_i, x_j; and x_i, x_j, x_k. For the convenience of later discussions, the points $\boldsymbol{x}_0$, $(x_i, \boldsymbol{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, $(x_i, x_j, \boldsymbol{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$, $\ldots$, are respectively called the zero-th

order, first-order, second-order model-constructing point(s), and so on. Accordingly, $f(\boldsymbol{x}_0)$ is the value of $f(\boldsymbol{x})$ at $\boldsymbol{x}_0$; $f(x_i, \boldsymbol{x}_0^i)$ is the model output at point $(x_i, \boldsymbol{x}_0^i)$. The Cut-HDMR, in its original form, only provides a lookup table for data interpolation; there is no explicit expression for component functions. It also does not have a sampling approach to support HDMR construction.

The recently developed RBF-HDMR [20] uses a sum of thin plate spline (the first term) plus a linear polynomial $P(\boldsymbol{x})$ (the second term) as follows to approximate each component function in Eqs. (4.3)-(4.6).

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \beta_i |\boldsymbol{x} - \boldsymbol{x}_i|^2 log|\boldsymbol{x} - \boldsymbol{x}_i| + P(\boldsymbol{x}),$$

$$\sum_{i=1}^{n} \beta_i \, \boldsymbol{p}(\boldsymbol{x}) = \boldsymbol{0},$$

$$P(\boldsymbol{x}) = \boldsymbol{p\alpha} = [p_1, p_2, \cdots, p_q][\alpha_1, \alpha_2, \cdots, \alpha_q]^T, \tag{4.7}$$

where $\boldsymbol{x}_i$ are the vectors of evaluated n sample points; the coefficients $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]$ and $\boldsymbol{\alpha}$ are parameters to be found. $P(\boldsymbol{x})$ is a polynomial function, where $\boldsymbol{p}$ consists of a vector of basis of polynomials. In this work, $\boldsymbol{p}$ is chosen to be $(1, x_1, \cdots, x_d)$ including only linear variable terms and therefore $q=d+1$; The side condition $\sum_{i=1}^{n} \beta_i \, \boldsymbol{p}(\boldsymbol{x}) = 0$ is imposed on the coefficients $\boldsymbol{\beta}$ to improve an under-determined system, that is, the singularity of distance matrix A with $A_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|^2 log|\boldsymbol{x}_i - \boldsymbol{x}_j|, \; i, j = 1, \cdots, n$. RBF is a simple interpolative function and found to provide a good approximation to arbitrary systems [6].

For the ease of description, we use linear RBF as a substitute for Eq. (4.7) without losing generality. Therefore a general RBF-HDMR model is written as

$$f(\boldsymbol{x}) \cong f_0 + \sum_{i=1}^{d} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, \boldsymbol{x}_0^i) - (x_{i_k}, \boldsymbol{x}_0^i) \right| + \sum_{1 \leq i < j \leq d} \sum_{k=1}^{m_{ij}} \alpha_{ij_k} \left| (x_i, x_j, \boldsymbol{x}_0^{ij}) - \right.$$

$$\left. (x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij}) \right| + \cdots + \sum_{k=1}^{m_{12 \cdots d}} \alpha_{12 \cdots d_k} |\boldsymbol{x} - \boldsymbol{x}_k| , \tag{4.8}$$

where $|.|$ denotes a p-norm distance; α_{i_k}, α_{ij_k}, ..., $\alpha_{12 \cdots d_k}$ are respectively the coefficient of the expression; (x_{i_k}, x_0^i), $(x_{i_k}, x_{j_k}, x_0^{ij})$, ..., $\boldsymbol{x}_k$ are the sampled points; m_i, m_{ij}, ..., $m_{12 \cdots d}$ are the number of sampled points for each term; the component $\sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, \boldsymbol{x}_0^i) - (x_{i_k}, \boldsymbol{x}_0^i) \right|$ is a function of only the i-th input variable x_i which explains the effect of the i-th input variable x_i independently acting on the output function $f(\boldsymbol{x})$; the component $\sum_{k=1}^{m_{ij}} \alpha_{ij_k} \left| (x_i, x_j, \boldsymbol{x}_0^{ij}) - (x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij}) \right|$ denotes the correlated contribution of the variables x_i and x_j upon the output $f(\boldsymbol{x})$ after the individual influences of x_i and x_j are discounted, and so on.

For a black-box function with d dimensionality, the number of all possible existing components can be expressed as

$$N = \sum_{i=0}^{d} \frac{d!}{(d-i)! i!} = 2^d. \tag{4.9}$$

It can be seen that N increases dramatically as the number of dimensionality, d, rises. This challenges both the identification of functional form and modeling accuracy if higher-order correlated terms exist in the black-box functions. RBR-HDMR has an attractive feature that it interpolates all of the prescribed points used for constructing all order component functions. The prescribed points are defined as follows. For the constant

component, the model construction point is x_0; for the first-order components, the model-construction points include x_0, and (x_{i_k}, x_0^i); for the second-order components, its model-construction points are x_0, (x_{i_k}, x_0^i), (x_{j_k}, x_0^j), and $(x_{i_k}, x_{j_k}, x_0^{ij})$.

4.4. RBF-HDMR Modeling Process

Based on the argument that most of well defined physical systems involve relatively low-order correlations of the input variables [16, 20], the RBF-HDMR modeling process up to the second-order was described as follows [20].

1. Randomly choose a point $x_0 = [x_{1_0}, x_{2_0}, \cdots, x_{d_0}]^T$ in the modeling domain. Evaluating $f(x)$ at x_0, we then have f_0.

2. Sample for the first-order component functions $f_i(x_i) = f([x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T) - f_0$ in the close neighborhood of the two ends of x_i (lower and upper limits) while fixing the rest of x_j ($j \neq i$) components at x_0. Evaluating these two end points, we got the left point value $f_{iL}(x_i) = f([x_{1_0}, x_{2_0}, \cdots, x_{i_L}, \cdots, x_{d_0}]^T) - f_0$ and similarly the right point value $f_{iR}(x_i)$. Model the component function as $\hat{f}_i(x_i)$ by a 1-D RBF model for each variable x_i using Eq. (4.7).

3. Check the linearity of $f_i(x_i)$. If the approximation model $\hat{f}_i(x_i)$ goes through the center point x_0, $f_i(x_i)$ is considered as linear. In this case, modeling for this component terminates. Otherwise, use the center point x_0 and the two end points to re-construct $\hat{f}_i(x_i)$. Then a random value along x_i is generated and combined with the rest of

x_j $(j{\neq}i)$ components at $\boldsymbol{x}_0$ to form a new point to test $\hat{f}_i(x_i)$. If $\hat{f}_i(x_i)$ is not sufficiently accurate (the relative error is larger than a given criterion, for instance, 0.01%, the test point and all the evaluated points will be used to re-construct $\hat{f}_i(x_i)$. This sampling-remodeling process iterates until convergence. This process captures the nonlinearity of the component function with one sample point at a time. Step 3 repeats the process for all of the first-order component functions to construct the first-order terms of RBF-HDMR model.

4. Form a new point, $(x_1, x_2, \cdots x_i, \cdots x_d)_k = [x_{1_k}, x_{2_k}, \cdots, x_{i_k}, \cdots, x_{j_k}, \cdots, x_{d_k}]^T, k \neq$ 0 by randomly combining the sampled value x_i in the first-order component construction for each input variable (that is, $x_i, i = 1, \cdots, d$ in evaluated $(x_i, \boldsymbol{x}_0^i)$, respectively). This new point is then evaluated by expensive simulation, as well as by the first-order RBF-HDMR model. If the two function values are sufficiently close (the relative error is less than a small value, for example, 0.01%), it indicates that no higher-order terms exist in the underlying function, and the modeling process terminates. Otherwise, go to the next step.

5. Use the values of x_i and x_j, $i \neq j$ that exist in the thus-far evaluated points $(x_i, \boldsymbol{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, and $(x_j, \boldsymbol{x}_0^j) = [x_{1_0}, x_{2_0}, \cdots, x_j, \cdots, x_{d_0}]^T$ to form new points of the form $(x_i, x_j, \boldsymbol{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$. Randomly select one of the points from these new points to test the first-order RBF-HDMR model. If the model passes through the new point, it indicates that x_i and x_j are not correlated, and the process continues with the next pair of input variables. This is to save the cost of modeling non-existing or insignificant correlations. Otherwise, use this new point and the

130

evaluated points $(x_i, \boldsymbol{x}_0^i)$ and $(x_j, \boldsymbol{x}_0^j)$ to construct the second-order component function, $\hat{f}_{ij}(x_i, x_j)$. This sampling-remodeling process iterates for all possible two-variable correlations until convergence (the relative error is less than 0.01%). Step 5 is repeated for all pairs of input variables.

Theoretically Step 5 applies to all higher-order terms in RBF-HDMR model in a similar manner. However, given the exponentially growing number of terms as shown in Eq. (4.9), even if only one extra point is needed to test if a higher-order correlation exists or not (as that in Step 5 for bivariate correlation), the number of sample points needed would increase exponentially. Therefore this work introduces some theorems first and uses the theorems to guide the modeling of higher-order component functions.

Figure 4.1 shows a simplified flow of the RBF-HDMR modeling process. The step Refine Model means increasing samples to improve the accuracy of the model without changing the functional form of the model; the step Update Functional Form adds the correlated component terms to RBF-HMDR model if the term exists. When the accuracy of modeling is reached, the modeling process terminates. The last step for higher-order components are discussed in the next section.
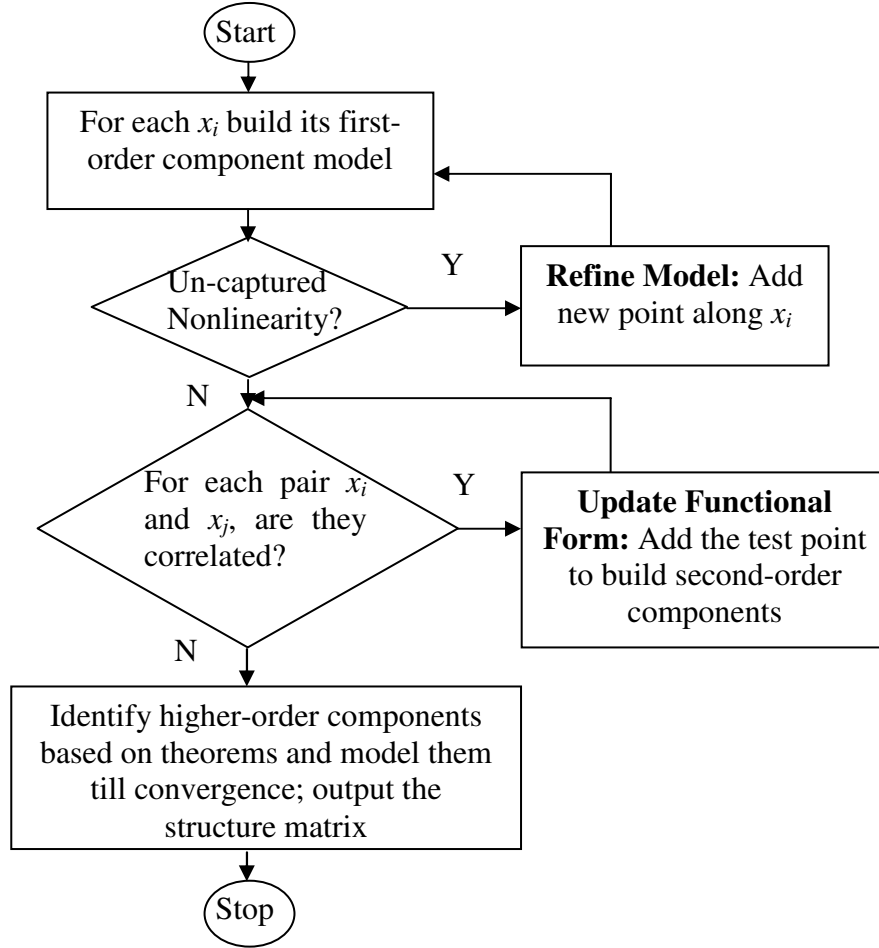
Fig. 4.1 A simplified flow of RBF-HDMR metamodeling

4.5. Principle of Functional Form Identification

This section first defines two matrices that support the identification of multivariate correlated terms and the functional form. Then theorems for identification are introduced, which form the basis for efficiently modeling higher-order components of RBF-HMDR, and also for uncovering the functional form of the black-box function.

4.5.1. Structure Matrix

A structure matrix is defined to capture the inner structure of the resultant RBF-HDMR of a black-box function as

$$SM_{d \times n} = \begin{bmatrix} 0\ 1\ 0\ 0 \dots 0\ 1 \dots 0 \dots 0 \dots 1 \\ 0\ 0\ 1\ 0 \dots 0\ 1 \dots 0 \dots 0 \dots 1 \\ 0\ 0\ 0\ 1 \dots 0\ 0 \dots 0 \dots 0 \dots 1 \\ \dots \\ 0\ 0\ 0\ 0 \dots 0\ 0 \dots 0 \dots 1 \dots 1 \\ 0\ 0\ 0\ 0 \dots 1\ 0 \dots 0 \dots 1 \dots 1 \end{bmatrix}, \qquad (4.10)$$

where d is the dimensionality of the input variable vector $\boldsymbol{x}$; n denotes the number of to-be-decided component terms. Each row corresponds to a variable, x_i. Each column corresponds to one of the component terms in the RBF-HDMR. Each element in the structure matrix is assigned as "0" or "1" ; "0" means that the variable is sampled at x_{i_0}; "1" means that the variable is sampled at non-x_{i_0} locations. For example, the column $[0_1, 0_2, \dots, 0_d]^T$ denotes the constant component term f_0; $[0_1, 0_2, \dots, 1_i, \dots, 0_d]^T$ represents the first-order component term $f_i(x_i)$; $[0_1, 0_2, \dots, 1_i, \dots, 1_j, \dots, 0_d]^T$ indicates the existence of the second-order component term $f_{ij}(x_i, x_j)$, and the last column indicates the existence of d-variate correlation component term $f_{12\dots d}(x_1, x_2, \dots, x_d)$.

The structure matrix is employed to index the corresponding component term and is created in tandem with the RBF-HDMR modeling process. Since each column in the structure matrix is associated with a unique component term and the maximum number of terms is 2^d (Eq. 4.9), a structure matrix therefore could theoretically represent a maximum of 2^d of terms [16]. Many component terms may not exist in the black-box function $f(\boldsymbol{x})$, and some others have negligible contribution to $f(\boldsymbol{x})$. These terms'

corresponding columns is eliminated from the structure matrix. For example, if the column $[0_1, 0_2, \ldots, 1_i, \ldots, 1_j, \ldots, 0_d]^T$ does not exist in the structure matrix, it means $f_{ij}(x_i, x_j)$ does not exist or is negligible. For the description convenience, a non-existing or negligible term is referred to in the rest of the paper as an insignificant term; otherwise, it is a significant term. The final output of the structure matrix thus depends on the intrinsic characteristics of the black-box function, and the structure matrix in return explicitly reveals the inner functional form of the black-box function. Each column in the structure matrix represents one term in the final RBF-HDMR and for each element x_i, a "1" in a column means that the variable exist in the corresponding component term.

4.5.2. Component Correlation Matrix

Given the fact that HDMR is built on a hierarchy of orthogonal component functions with increasing dimensionality, we can explore further the variable relationships in the context of component functions. As one understands from Eqs. 4.2-4.6, for instance, $f_{12}(x_1, x_2)$ does not simply capture the term $x_1 x_2$, but rather the residual effect $(f(x_1, x_2, x_0^{12}) - f_1(x_1) - f_2(x_2) - f_0)$ of $x_1 x_2$. In other words, the algebraic term $x_1 x_2$ is expressed by f_0, $f_1(x_1)$, $f_2(x_2)$, and $f_{12}(x_1, x_2)$ altogether. The effect of the term $x_1 x_2$ could have been well captured till the first-order components, and thus there is no need to model the second-order term $f_{12}(x_1, x_2)$; otherwise, $f_{12}(x_1, x_2)$ is necessary to be added to accurately model the term $x_1 x_2$. Whether or not $f_{12}(x_1, x_2)$ is significant helps us to define the variable correlation $x_1 x_2$ from the perspective of HDMR. This point separates our variable correlation matrix, to be defined below, from

its conventional meaning. To distinguish the difference, we define the component correlation matrix as follows.

Considering d variables (this is, $x = [x_1, x_2, \ldots, x_d]^T$), both the row list and column list in a component correlation matrix denote the same set of input variables. The matrix entry indicates whether the i-th input variable in row and the j-th input variable in column defines a component term $f_{ij}(x_i, x_j)$. A component correlation matrix (CCM) is

$$CCM_{d \times d} = [m_{ij}], \quad (i = 1,2, \ldots, d; j = 1,2, \ldots, d), \qquad (4.11)$$

where $m_{ij} = 1$ if $f_{ij}(x_i, x_j)$ exists in a HDMR formula for a particular problem; otherwise, $m_{ij} = 0$. For example, a component correlation matrix for a function of only first-order components is a diagonal matrix of 1's; the component correlation matrix of all significant bivariate component terms is a square matrix with all 1's, as shown in Fig. 4.2. Similar to a conventional correlation matrix, CCM is symmetric.

CCM can be automatically generated after completing modeling RBF-HDMR's second-order terms, because the modeling process adaptively identifies such relationships. The 0's and 1's scatter in CCM depending on characteristics of the underlying black-box function. A CCM can be reorganized by changing the order of rows and columns to exhibit patterns of correlations or extract part of rows and columns to form sub-correlation matrices. CCM captures all the bivariate component terms and implies for more-than-two-variable component terms, which will be discussed in the next subsection.

x	x_1	x_2	$\cdots$	x_i	$\cdots$	x_d
x_1	1	1	$\cdots$	1	1	1
x_2	1	1	$\cdots$	1	1	1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	1	1	$\cdots$	1	1	1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_d	1	1	$\cdots$	1	1	1

Fig. 4.2 Component correlation matrix indicating a function having all significant bi-variate terms

4.5.3. Correlation Identification for Higher-order Component Modeling

A CCM matrix shows second-order component terms between variable pairs. How can we identify higher-order component terms involving three or more variables without occurring extra sampling cost? Let us take a t-variable subset $(3 \leq t \leq d)$ from a CCM to form a new sub-matrix. Such matrices include two types, a $t \times t$ matrix with all 1's and a $t \times t$ matrix with at least one 0.

Theorem 1: The necessary condition of a t-variable $(t \geq 3)$ component term existing in a HDMR formulation for a black-box function is that the t-variable sub-matrix of CCM is a $t \times t$ matrix with all 1's and all of its component terms involving $(t-1)$ variables exist in the HDMR model.

Proof: Assuming t variables $x_{i_1}, \cdots, x_{i_t}$, if one possible component term of a subset of $(t-1)$ variables $x_{i_1}, \cdots, x_{i_{t-1}}$, $f_{i_1 \ldots i_{t-1}}$, does not exist, it means that the corporate

136

contribution of the subset $x_{i_1}, \cdots, x_{i_{t-1}}$ is not significant after all the lower-order effects being modeled. Therefore the higher-order component, $f_{i_1 \ldots i_t}$, would not be significant either. Similarly if $f_{i_1 \ldots i_t}$ exists, it means that all the lower-order components for all t-variables should exist since $f_{i_1 \ldots i_t}$ is computed from all its related lower-order components (see Eqs. 4.5 and 4.6). Therefore, all entries in the $t \times t$ matrix should be 1's. If there exists an entry of '0' between two variables x_{i_l} and x_{i_m}, it means that $f_{i_l i_m}$ does not exist or is not significant, and therefore any higher-order components involving x_{i_l}, x_{i_m} would not be significant either. Proof completed.

Theorem 1 can be used to explore higher-order component terms in RBF-HDMR. If the necessary condition is not met, then the corresponding t-variable component term does not exist, and the modeling of the term is skipped during modeling, and thus an extra sample point is saved. It is to be noted if a sub-matrix of t-variable CCM has all entry of 1's, one cannot sufficiently conclude that all of the 3rd and higher-order components exist. It is because that CCM only defines bivariate relations. Theorem 2 is therefore proposed to supplement Theorem 1.

Theorem 2: The sufficient condition of the existence of a t-variable ($t \geq 3$) component term in a RBF-HDMR formulation for a black-box function is that the value of a new point formed from existing model construction points' variable elements for up to the $(t-1)$ -th order component terms is not accurately predicted by the RBF-HDMR model of $(t-1)$ -th order.

Proof: Assume a RBF-HDMR model of $(t-1)$-th order is built, that is,

$$\tilde{f}(x) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i < j \le d} f_{ij}(x_i, x_j) + \cdots$$

$$+ \sum_{1 \le i_1 < \cdots < i_{t-1} \le d} f_{i_1 i_2 \cdots i_{(t-1)}} \left(x_{i_1}, x_{i_2}, \cdots, x_{i_{(t-1)}} \right). \tag{4.12}$$

All its model construction points up to the $(t-1)$-th order include x_0, (x_i, x_0^i), (x_i, x_j, x_0^{ij}), $\cdots$, and $(x_{i_1}, x_{i_2}, \cdots, x_{i_{(t-1)}}, x_0^{i_1 i_2 \cdots i_{(t-1)}})$. One picks variable elements from these points to form a new point $(x_{i_1}, x_{i_2}, \cdots, x_{i_t}, x_0^{i_1 i_2 \cdots i_t})$. If the current $(t-1)$-th order RBF-HDMR cannot accurately predict the function value of the new point, it means that the interaction of all t-variables has not been captured by the $(t-1)$-th order model and therefore the t-variable $(t \ge 3)$ correlated component term should exist. Proof completed.

Theorem 2, the sufficient condition, can be used for confirming the existence of the correlated component terms. The difficulty is what if the $(t-1)$-th order model accurately predicts the function values of all possible new points? Strictly speaking, one cannot sufficiently conclude from Theorem 2 that there does not exist t-th order or higher components. Assuming the misjudgment under such situation is a rare occurrence, for practical algorithm development, the sufficient condition is loosened as follows.

Loosened sufficient condition: If a RBF-HDMR of $(t-1)$-th order can exactly predict the function value at the test point constructed from existing model points' variable elements for up to the $(t-1)$-th order component terms, then it is deemed that there is no t-variable or higher-order component terms in the black-box function.

4.5.4. Modeling of High-order Component Functions in RBF-HDMR

The theorems developed in Section 4.5.3 are employed to identify and model high-order component functions in RBF-HDMR.

The general process follows that for the modeling of second-order component, as described in Section 4.4. The main difference is on the identification of high-order component functions, in order to avoid exponentially increasing sampling cost, and to reveal the functional form of the black-box function. The logic for component identification is illustrated in Fig. 4.3.
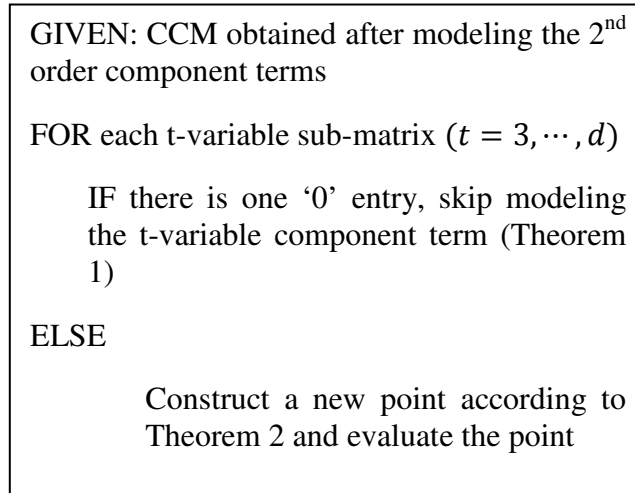
GIVEN: CCM obtained after modeling the 2nd order component terms

FOR each t-variable sub-matrix $(t = 3, \cdots, d)$

    IF there is one '0' entry, skip modeling the t-variable component term (Theorem 1)

ELSE

        Construct a new point according to Theorem 2 and evaluate the point

Fig. 4.3 Process for high-order component identification

The structure matrix of the black-box function, as shown in Eq. 4.10, is constructed and updated with the modeling process. Once the modeling is completed, its functional form is captured in the structure matrix. The following example explains the process in detail.

4.5.5. An Example

One example is used to illustrate the modeling and functional form identification process. This problem is modified from [22] and expressed as

$$f(x) = \pi^{x_1 x_2}\sqrt{2x_3} - \sin^{-1} x_4 + \log(x_3 + x_5) - \frac{x_9}{x_{10}}\sqrt{\frac{x_7}{x_8}} - x_2 x_7 + x_6^2. \qquad (13)$$
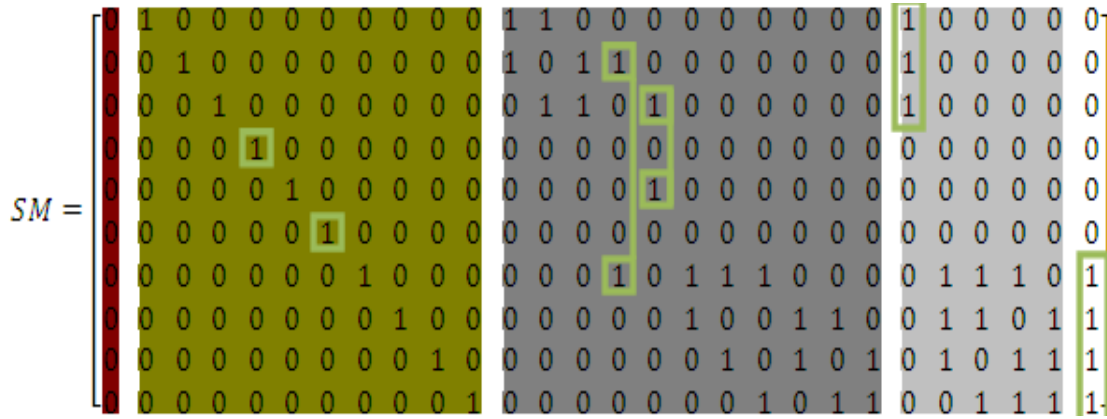

Fig. 4.4 The structure matrix of the example

The resultant structure matrix of this example is shown in Fig. 4.4. In the structure matrix, we shadow component terms from zero-th order to the highest order with gradually lighter colors. Also we mark the independent component terms by rectangular boxes. Matching the structural matrix with steps in Section 4.4, the first colored block corresponds to Step 1; the second colored block is implemented by Step 2 and 3; Step 4 happens between the second colored block and the third colored block; Step 5 fills the third colored block and generates a component correlation matrix; the last two colored blocks are implemented via Step 5 by means of the two derived theorems, and the algorithm shown in Fig. 4.3. From the fourth and sixth rows, one can see that input variables x_4 and x_6 are independent variables, that is, having only first-order component

terms. Observing from the second rightmost block, one can see that variables $x_1, x_2,$ and x_3 form a three-variable component term; from the last block, variables x_7, x_8, x_9, x_{10} form a four-variable component term. The middle color block shows that two-variable correlation exists between variable pairs x_2 and x_7, as well as x_3 and x_5. It is to be noted that multiple variable component terms can be ignored if the terms are trivial to the output. From the final structure matrix, one can extract following mathematical expression

$$f(x) = f_0 + g_4(x_4) + g_6(x_6) + g_{2,7}(x_2, x_7) + g_{3,5}(x_3, x_5) + g_{1,2,3}(x_1, x_2, x_3) +$$

$$g_{7,8,9,10}(x_7, x_8, x_9, x_{10}), \tag{4.14}$$

where, $g_i(x_i) = f_i(x_i), i = 4,6,$

$$g_{2,7}(x_2, x_7) = f_2(x_2) + f_7(x_7) + f_{2,7}(x_2, x_7),$$

$$g_{3,5}(x_3, x_5) = f_3(x_3) + f_5(x_5) + f_{3,5}(x_3, x_5),$$

$$g_{1,2,3}(x_1, x_2, x_3) = \sum_{i=1}^{3} f_i(x_i) + \sum_{i=1, i<j}^{3} f_{ij}(x_i, x_j) + f_{1,2,3}(x_1, x_2, x_3),$$

$$g_{7,8,9,10}(x_7, x_8, x_9, x_{10}) =$$

$$\sum_{i=7}^{10} f_i(x_i) + \sum_{i=7, i<j}^{10} f_{ij}(x_i, x_j) + \sum_{i=7, i<j<k}^{10} f_{ijk}(x_i, x_j, x_k) +$$

$$f_{7,8,9,10}(x_7, x_8, x_9, x_{10}).$$

Eq. (4.14) corresponds to the structure depicted by Fig. 4.4. f_0 corresponds to the first column. The numerical models of all component functions have been obtained using the modeling process described in Section 4.4 and stored in the final model. The final model manifests high-dimensional correlated behavior of variables. The linearity/nonlinearity information regarding to each input variable is also saved in the final model and can be readily output.

The CCM corresponding to the SM in Fig. 4.4 is

$$
CCM = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}.
\tag{4.15}
$$

From this CCM, one can roughly see the correlation among the variables in Eq. (4.14). The top left corner 3×3 sub-matrix indicates that x_1, x_2 and x_3 may be correlated; however, it needs to be judged by the structure matrix. Similarly the bottom right corner 4×4 sub-matrix indicates that x_7, x_8 x_9 and x_{10} may be correlated. Both the 4th row and 4th column have only one '1' element at the diagonal position, which shows that x_4 is only in first-order component term $f_4(x_4)$ and not in higher-order terms. Similarly, it is true for x_6. Variables x_2 and x_7 correlate strongly and $f_{2,7}(x_2,x_7)$ must be modeled; this is also true for x_3 and x_5.

The modeling result is given in Table 4.1, where NoE accumulates the number of function evaluations from lower to higher order, for example, it requires in total of 441 points to model the function up to the fourth order, 393 points to the third order, and so on. "id" means accumulated NoE spent on identification of functional form, which is used for modeling if the term exists, and 35 sampling points in total are generated in the second order for this purpose. The column "model" indicates the NoE used for modeling. R Square, RAAE and RMAE are model performance metrics which will be introduced in the next section. Table 4.1 shows that the second RBF-HDMR models the underlying function well, and usually no more modeling effort is needed. However, one can see that

the performance metrics like R Square and RAAE became worse from the second to the third order. This phenomenon indicates "over fitting", this is, the gain from modeling higher-order terms is less than the error brought from the modeling process. Over fitting is one of common issues in metamodeling techniques. Wang *et al.* [23] discussed over fitting about RBF metamodel. Tecko *et al.* [24] presented comparison of over fitting and over training in artificial neural network. In artificial neural network community, some additional techniques such as early stopping and cross validation are used to avoid over fitting. Due to limited space, discussion on over fitting is not extended here.

Table 4.1 Modeling results of the example

Order	R Square	RAAE	RMAE	NoE		
				id	model	total
First	0.3809	0.3331	10.399	0	137	137
Second	0.9386	0.1150	5.1044	35	238	273
Third	0.9182	0.1372	2.4582	35	358	393
Fourth	0.9187	0.1361	2.4580	35	406	441

4.6. Test Examples

4.6.1. Problem Description

To test the effectiveness and efficiency of the proposed approach, fifteen test problems are selected based on the criteria: 1) the number of variables ≥ 10; 2) high nonlinearity of the performance behavior; and 3) multiple variables are correlated. The

criteria are chosen to expose challenges for metamodeling HEB problems. Scalable problems with different dimensionality are treated as one problem. In total fifteen problems that satisfy the criteria are found from the book of Schittkowski [25], which offers 188 problems for testing nonlinear optimization algorithms and a few of them for testing data fitting algorithms. Most of these problems have some application background. Fifteen problems that satisfy our criteria are listed in 4.10 Appendix. Among these problems, the first ten are classified by Schittkowski as "theoretical" problems denoted by "T", and remaining five problems as "practical" problems represented by "P". Detailed backgrounds of these practical problems are omitted due to space limit. "Order" stands for highest order which is analyzed by RBF-HDMR. The modeling accuracy is evaluated by four performance metrics which are introduced in the next section.

4.6.2. Performance Metrics

1)   R Square

$$R^2 = 1 - \frac{\sum_{i=1}^{m}[f(x_i) - \hat{f}(x_i)]^2}{\sum_{i=1}^{m}[f(x_i) - \bar{f}(x_i)]^2},$$
(4.16)

where $\bar{f}(x_i)$ denotes the mean of function on the m sampling points. This metrics indicates the overall accuracy of an approximation model. The closer the value of R square approaches one, the more accurate is the approximation model. Note R square in this work is computed on 10,000 new test sample points for each problem, rather than on the modeling points. The same is true for the next two metrics.

2)   Relative Average Absolute Error (RAAE)

$$RAAE = \frac{\sum_{i=1}^{m}|f(x_i)-\hat{f}(x_i)|}{m*STD},$$ (4.17)

where *STD* stands for standard deviation. Like R square, this metric shows the overall accuracy of an approximation model. The closer the value of RAAE approaches zero, the more accurate is the approximation model.

3)   Relative Maximum Absolute Error (RMAE)

$$RMAE = \frac{\max\left(|f(x_1)-\hat{f}(x_1)|,|f(x_2)-\hat{f}(x_2)|,\cdots,|f(x_m)-\hat{f}(x_m)|\right)}{STD}.$$ (4.18)

This is a local metric. A RMAE describes error in a sub-region of the design space. Therefore, a small value of RMAE is preferred.

4.6.3. Test Results

Expressions for the 15 problems are listed in the 4.10 Appendix. Table 4.2(a) and 4.2(b) show the results of 14 test examples except for Problem 12. The problem 12 is discussed in Section 4.6.4. The results represent the average of 10 independent runs. It can be seen that RBF-HDMR models well 14 problems out of 15.

In Table 4.2(a), problems 1, 2, 11, and 15 are chosen for detailed report; other results are in Table 4.2(b) for brevity. For problems in Table 4.2(a), the two matrices for each problem are shown, along with its mathematical function description as those in Eq. 4.14. For Problem 1, bivariate correlations exist, which are clearly shown with brighter colors from the two matrices. The RBF-HDMR model also reached high accuracy with up to the second-order components. Problem 2 has high-order multivariate correlations, but these terms have small influence and can be neglected. Problem 11 has up to 6-th

145

order correlations. The fourth-order RBF-HDMR model reaches R Square value of 0.938 and the cost grows significantly due to strong variable correlations as the model moves one order higher. Problem 15 has 50 variables, but its internal structure is very simple and there is no strong correlation between variables. The modeling cost is thus low. Comparing Problems 1 and 2, both have d=10, the NoE is 196 and 584, respectively for the second-order model. As one can see that the cost for Problem 2 is significantly higher. Problem 2 consists of multiple correlated second-order and third-order terms, which are evident from its SM and CCM. Its CCM has all 1's elements. SM also appears more complicated than that of Problem 1. Then for Problem 11 with 11 variables, the cost for the second-order is similar to that of Problem 2 but doubles each time the order increases by 1.

Problem 11 has even more complex structure than Problem 2 with multiple correlated high-order terms until the sixth order, which explains its high modeling cost. For Problem 15, since it is a first-order problem, the NoE is only 251 even with d=50. Table 4.2(b) shows the test results for other functions. These results show the effectiveness and efficiency for a wide range of problems. In addition, this work did not choose to compare the RBF-HDMR with other metamodeling techniques. Interested readers can refer to Ref. [20] for such comparison.
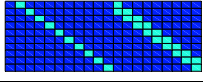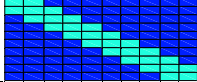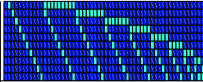
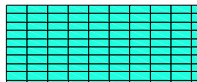**Table 4.2(a) Test results of examples**
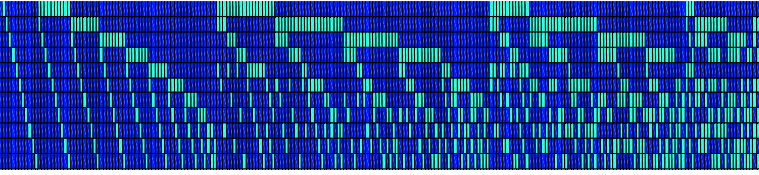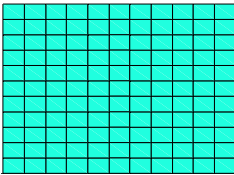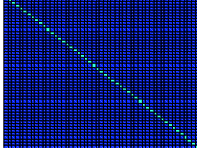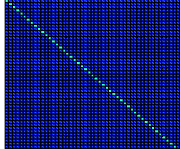
Function		R Square	RAAE	RMAE	NoE		
					id	model	total
1	First	0.8683	0.2770	1.8064	0	51	51
d=10	Second	0.9900	0.0804	0.3457	37	159	196
	SM		**CCM**		$f(x) = c + \sum_{i=1}^{9} g_{i,i+1}(x_i, x_{i+1})$ No higher than 2nd order		
2	First	0.2756	0.4027	11.0485	0	51	51
d=10	Second	0.9713	0.1025	2.7502	1	583	584
	SM		**CCM**		$f(x) = c + \sum_{1 \le i < j \le 10} g_{i,j}(x_i, x_j)$ Higher-order negligible		
11	First	-0.3517	0.9050	5.6850	0	56	56
d=11	Second	0.5412	0.4508	4.7432	1	579	580
	Third	0.8946	0.2368	2.0660	82	1251	1333
	Fourth	0.9368	0.1835	1.6048	82	2211	2293
	Only output first four orders						
	SM				**CCM**		
	Only write out 3 highest correlated terms $$f(x) = c + \cdots + g_{2,3,6,7,9,10}(x_2, x_3, x_6, x_7, x_9, x_{10})$$ $$+ g_{2,4,6,8,9,11}(x_2, x_4, x_6, x_8, x_9, x_{11})$$ $$+ f_{3,4,7,8,10,11}(x_3, x_4, x_7, x_8, x_{10}, x_{11});$$ No higher than 6th order						
15	First	0.9024	0.2986	0.7368	0	251	251
d=50	SM		CCM		$f(x) = c + \sum_{i=1}^{50} g_i(x_i)$ No higher than 1st order		

Table 4.2(b) Test results of examples

Function		R Square	RAAE	RMAE	NoE id	NoE model	NoE total
3	First	0.8622	0.2888	1.6642	0	101	101
$d=10$	Second	0.9438	0.2027	0.8231	136	209	345
4	First	0.8918	0.2555	1.3816	0	101	101
$d=20$	Second	0.9471	0.2008	0.6896	166	265	431
5	First	-0.1862	0.8074	4.9560	0	101	101
$d=20$	Second	0.8778	0.2663	1.5423	161	341	502
6	First	0.4029	0.5293	5.3301	0	151	151
$d=30$	Second	0.8715	0.2161	4.6852	1	3631	3632
7	First	-2.6392	1.7225	6.869286	0	151	151
$d=30$	Second	0.7746	0.4213	1.6282	1	5367	5368
8	First	0.5126	0.5494	3.1200	0	151	151
$d=31$	Second	0.9506	0.1789	0.7747	407	499	906
9	First	0.4968	0.5585	4.2935	0	101	101
$d=20$	Second	0.9906	0.0852	0.4121	172	265	437
10	First	0.9206	0.1936	3.1821	0	99	99
$d=20$	Second	0.9971	0.0373	1.0299	1	1642	1643
13 $d=14$	First	0.9997	0.0135	0.0716	0	71	71
14 $d=30$	First	0.9999	0.0075	0.0168	0	151	151

4.6.4. Discussion

The test problem 12 is expressed as

$$f(x) = 100000 \prod_{i=1}^{11} x_i^{a_i}, \quad 0.1 \leq x_i \leq 100, \ i = 1, \cdots, 11, \tag{4.19}$$

$$a = \begin{pmatrix} -0.00133172, & -0.002270927, -0.00248546, -4.67 \\ -4.671973, -0.00814, -0.008092, -0.005, -0.000909, \\ -0.00088, -0.00119 \end{pmatrix}.$$

In this problem, α_i is employed for specifying the role of the corresponding variable x_i. α_i significantly affects the final output and modeling results. This effect can be seen from Fig. 4.5. Fig. 4.5 plots the $f(x)$ (vertical axis) with regard to x_i (all x_i are held equal to each other). As α_4 and α_5 getting smaller, the output curve becomes extremely steep.

148

$$\alpha_4 = \alpha_5 = 0 \qquad\qquad \alpha_4 = \alpha_5 = -0.5 \qquad\qquad \alpha_4 = \alpha_5 = -4.67$$
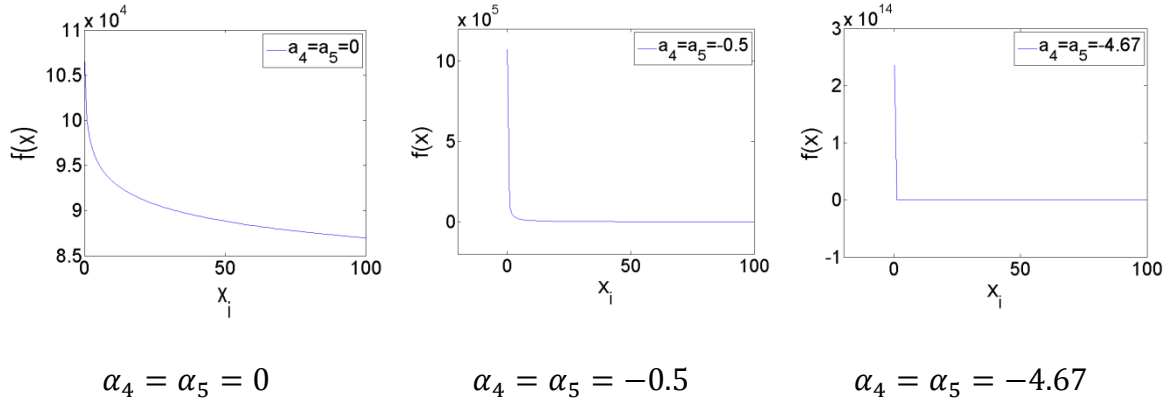
Fig. 4.5 Deterioration of $f(x)$ when decreasing coefficients α_4 and α_5

When they reach the value above at (-4.67 -4.67), the output forms a right angle with the x-axis and the modeling error is prohibitive. Table 4.3 shows the modeling results. The structure matrices and correlation matrices are given in Fig. 4.6. It is very interesting to see as the coefficients decrease from 0 to -0.5, the SM and CCM change from simple structure to complex ones with multivariate correlations. It means as α_i decreases, high-order variable correlations becomes stronger and finally dominates $f(x)$.

Table 4.3 The results of the example 12

α_4	α_5	R Square	RAAE	RMAE	NoE
0	0	0. 9914	0.0406	0.9234	126
-0.5	-0.5	0.6791	0.1438	20.1244	144
-4.67	-4.67	0.0000	0.0114	99.7205	146

149
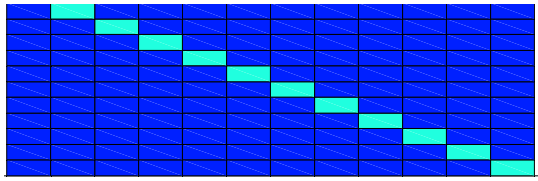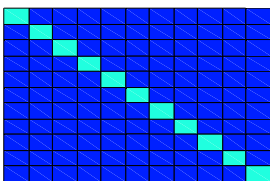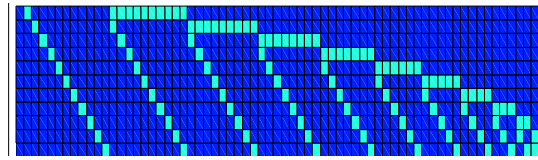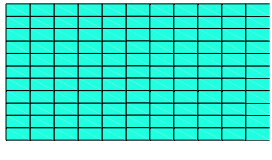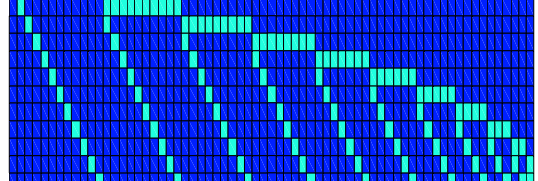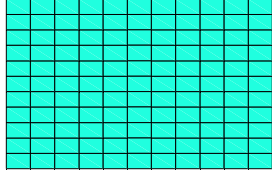
	SM	CCM
$\alpha_4 = \alpha_5 = 0$		
$\alpha_4 = \alpha_5 = -0.5$		
$\alpha_4 = \alpha_5 = -4.67$		

Fig. 4.6 Structure matrices and correlation matrices of Problem 12

4.7. Final Remarks

This work extends the recently developed RBF-HDMR method to model higher than second-order component functions, based on which a black-box function can be "uncovered." Key information about a black-box function such as functional form, variable (non)linearity, and variable correlations can be obtained through the modeling process. A structure matrix (SM) is developed to present the functional form of the black-box problems. A component correlation matrix (CCM) is defined to describe correlation relationship among the variables. Note that SM and CMM depend on the characteristics of HEB problems and do not dictate the exclusive association with RBF-HDMR. First, if other metamodels are used to model the component terms of HDMR,

SM and CCM remain exactly the same as in the context of RBF-HDMR. Second, even if HDMR is not used, for instance, for a second-order polynomial response surface, its constant, first-order and second-order terms correspond to the first a few columns of SM as defined in Eq. 4.10; CCM can represent the correlation among variables according to the coefficients. However, the full second-order response surface is a parametric model whose functional form is postulated. Therefore the high-order terms may be significantly but ignored once the second-order polynomial model is chosen. Although SM and CCM are applicable for other metamodel techniques, the challenge lies on how to uncover the functional form and fill the matrices. In this work, two theorems are developed to support the efficient identification of high-order correlation terms in the context of HDMR. Multiple test examples show the effectiveness and efficiency of the proposed approach. Future work will extend the methodology to support ANOVA analysis, direct problem decomposition, and design optimization.

4.8. Acknowledgement

4.9. References

[1]   Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Statistical Science, **4**(4), pp. 409-435.

[2]   Myers, R. H. and Montgomery, D., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* John Wiley and Sons, Inc., Toronto.

[3]   Simpson, T. W., Lin, D. K. J. and Chen, W., 2001, "Sampling Strategies for Computer Experiments: Design and Analysis," International Journal of Reliability and Application, **2**(3), pp. 209-240.

[4]   Jin, R., Chen, W. and Sudjianto, A., 2005, "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments," Journal of Statistical Planning and Inferences, **134**(1), pp. 268-287.

[5]   Shan, S. and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-box Functions," Structural and Multidisciplinary Optimization, (2010) 41, pp. 219-241. DOI 10.1007/s00158-009-0420-2.

[6]   Chen, V., C. P., Tsui, K.-L., Barton, R. R. and Meckesheimer, M., 2006, "A Review on Design, Modeling and Applications of Computer Experiments," IIE Transactions, 38, pp. 273-291.

[7]   Queipo N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R. and Tucker, P. K., 2005, "Surrogate-Based Analysis and Optimization," Progress in Aerospace Sciences, (41), pp. 1-18.

[8]   Simpson, T. W., Toropov, V. and Balabanov, V., 2008, "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - or Not," 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Victoria, British Columbia, Canada. 10-12 September, 2008. Paper No. 2208-5802.

[9]   Wang, G. G. and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," ASME Journal of Mechanical Design, 129, pp. AIAA 370-389.

[10]   Koch, P. N., Simpson, T. W., Allen, J. K. and Mistree, F., 1999, "Statistical Approximations for Multidisciplinary Design Optimization: The Problem of Size," Journal of Aircraft, 36(1), pp. 275-286.

[11]   Simpson, T. W., Peplinski, J., Koch, P. N. and Allen, J. K., 2001, "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," Engineering with Computers, 17(2), pp. 129-150.

[12]   Friedman, J. H. and Stuetzle, W., 1981, "Projection Pursuit Regression," Journal of the American Statistical Association, 76(372), pp. 817-823.

[13]   Friedman, J. H., 1991, "Multivariate Adaptive Regressive Splines," The Annals of Statistics, 19(1), pp. 1-67.

[14] Tunga, M. A. and Demiralp, M., 2006, "Hybrid High Dimensional Model Representation (Hhdmr) on the Partitioned Data," Journal of Computational and Applied Mathematics (185), pp. 107-132.

[15] Rabitz, H., Alis, O. F., Shorter, J. and Shim, K., 1999, "Efficient Input-Output Model Representations," Computer Physics Communications, (117), pp. 11-20.

[16] Rabitz, H. and Al1s, Ö. F., 1999, "General Foundations of High-Dimensional Model Representations," Journal of Mathematical Chemistry, (25), pp. 197-233.

[17] Shorter, J. A., Ip, P. C. and Rabitz, H. A., 1999, "An Efficient Chemical Kinetics Solver Using High Dimensional Model Representation," Journal of Physical Chemistry A, (103), pp. 7192-7198.

[18] Li, G., Wang, S.-W., Rosenthal, C. and Rabitz, H., 2001, "High Dimensional Model Representations Generated from Low Dimensional Data Samples. I. Mp-Cut-Hdmr," Journal of Mathematical Chemistry, 30(1), pp. 1-30.

[19] Tunga, M. A. and Demiralp, M., 2005, "A Factorized High Dimensional Model Representation on the Nodes of a Finite Hyperprismatic Regular Grid," Applied Mathematics and Computation (164), pp. 865-883.

[20] Shan, S. and Wang, G. G., 2009, "Metamodeling for High Dimensional Simulation-Based Design Problems," ASME Journal of Mechanical Design, Vol. 132, pp. 5:051009-1-051009-11.

[21] Booker, A. J., 1998, "Design and Analysis of Computer Experiments," 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization, AIAA, St. Louis, MO, pp. 118-128.

[22] Hooker, G., 2004, "Discovering Additive Structure in Black Box Functions," Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, August 22-25.

[23] Wang, L., Beeson, D., Wiggs, G. and Rayasam, M., 2006, "A Comparison of Meta-Modeling Methods Using Practical Industry Requirements," 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 1 - 4 May 2006, Newport, Rhode Island, AIAA 2006-1811.

[24] Tecko, I. V., Livingstone, D. J. and Luik, A. I., 1995, "Neural Network Studies. 1. Comparison of Overfitting and Overtraining," J. Chem. Inf. Comput. Sci., **35**, pp. 826-833

[25] Schittkowski, K., 1987, *More Test Examples for Nonlinear Programming Codes,* Springer-Verlag New York, Inc., New York, NY, USA.

4.10.   Appendix

No.	Function	Variable Ranges	class	order
1	$f(\mathbf{x}) = (x_1 - 1)^2 + (x_{10} - 1)^2 + 10\sum_{i=1}^{9}(10 - i)(x_i^2 - x_{i+1})^2$	$-3 \le x_i \le 2,$ $i = 1, \cdots, 10$	T	2
2	$f(\mathbf{x}) = \left[\sum_{i=1}^{10} i^3(x_i - 1)^2\right]^3$	$-3 \le x_i \le 3,$ $i = 1, \cdots, 10$	T	3
3	$f(\mathbf{x}) = \sum_{i=1}^{10}[100(x_i - x_{i+10})^2 + (x_i - 1)^2]$	$-3 \le x_i \le 5,$ $i = 1, \cdots, 10$	T	2
4	$f(\mathbf{x}) = \sum_{1}^{5}[100(x_i^2 + x_{i+5})^2 + (x_i - 1)^2 + 90(x_{i+10}^2 + x_{i+15})^2 + (x_{i+10} - 1)^2$ $+ 10.1[(x_{i+5} - 1)^2 + (x_{i+15} - 1)^2] + 19.8(x_{i+5} - 1) \times (x_{i+15} - 1)]$	$-3 \le x_i \le 5,$ $i = 1, \cdots, 5$	T	2
5	$f(\mathbf{x}) = \sum_{1}^{5}[(x_i + 10x_{i+5})^2 + 5(x_{i+10} - x_{i+15})^2 + (x_{i+5} - 2x_{i+10})^4 + 10(x_i - x_{i+15})^4]$	$-2 \le x_i \le 5,$ $i = 1, \cdots, 5$	T	2
6	$f(\mathbf{x}) = 1 - exp\left[-\frac{1}{60}\sum_{1}^{30}x_i^2\right]$	$0 \le x_i \le 3.5,$ $i = 1, \cdots, 30$	T	30
7	$f(\mathbf{x}) = (\mathbf{x}^T A\mathbf{x})^2, A = diag(1,2,3, \dots, 30)$	$-2 \le x_i \le 3,$ $i = 1, \cdots, 30$	T	2
8	$f(\mathbf{x}) = \sum_{i=1}^{29}[100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$	$-2 \le x_i \le 2,$ $i = 1, \cdots, 30$	T	2
9	$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} - 2x_1 , A = \begin{bmatrix} 1 & -1 & & & & & 0 \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \cdots & \cdots & \cdots & 0 \\ & & & \cdots & \cdots & \cdots \\ & & & -1 & 2 & -1 \\ 0 & & & & -1 & 2 \end{bmatrix}$	$0 \le x_i \le 25,$ $i = 1, \cdots, d,$ $d = 20, 50, 100$	T	2
10	$f(\mathbf{x}) = \sum_{1}^{20}x_i^2 + \left[\sum_{1}^{20}\frac{1}{2}i x_i\right]^2 + \left[\sum_{1}^{20}\frac{1}{2}i x_i\right]^4$	$0 \le x_i \le 5,$ $i = 1, \cdots, 20,$	T	4
11	$f(\mathbf{x}) = \sum_{1}^{65}\left[y_i - [x_1 exp[-x_5 t_i] + x_2 exp[-x_6(t_i - x_9)^2] + x_3 exp[-x_7(t_i - x_{10})^2] + x_4 exp[-x_9(t_i - x_{11})^2]]\right]^2$, $t_i = 0.1(i - 1), (i = 1, \dots, 65),$ $y = \begin{pmatrix} 1.366\ 1.191\ 1.112\ 1.013\ 0.991\ 0.885\ 0.831\ 0.847\ 0.786\ 0.725\ 0.746 \\ 0.679\ 0.608\ 0.655\ 0.616\ 0.606\ 0.602\ 0.626\ 0.651\ 0.724\ 0.649\ 0.649 \\ 0.694\ 0.644\ 0.624\ 0.661\ 0.612\ 0.558\ 0.533\ 0.495\ 0.500\ 0.423\ 0.395 \\ 0.375\ 0.372\ 0.391\ 0.396\ 0.405\ 0.428\ 0.429\ 0.523\ 0.562\ 0.607\ 0.653 \\ 0.672\ 0.708\ 0.633\ 0.668\ 0.645\ 0.632\ 0.591\ 0.559\ 0.597\ 0.625\ 0.739 \\ 0.710\ 0.729\ 0.720\ 0.636\ 0.581\ 0.428\ 0.292\ 0.162\ 0.098\ 0.054 \end{pmatrix}$	$0 \le x_1 \le 1.6;$ $0 \le x_i \le 2,$ $i = 2, \cdots, 5;$ $2 \le x_i \le 8, i = 6, \cdots, 8;$ $1 \le x_9 \le 6,$ $4.5 \le x_i \le 6,$ $i = 10, 11;$	P	6
12	$f(\mathbf{x}) = 100000 \prod_{i=1}^{11} x_i^{a_i}$, $a = (-0.00133172\ -0.002270927\ -0.00248546\ -4.67\ -4.671973 -0.00814\ -0.008092\ -0.005\ -0.000909\ -0.00088 -0.00119)$	$0.1 \le x_i \le 100,$ $i = 1, \cdots, 11$	P	11
13	$f(\mathbf{x}) = \sum_{1}^{14} a_i/x_i$ $a = \begin{pmatrix} 12842.275\ 634.25\ 634.25\ 634.125\ 1268\ 633.875\ 633.75 \\ 1267\ 760.05\ 33.25\ 1266.25\ 632.875\ 394.46\ 940.838 \end{pmatrix}$	$0.0001 \le x_i,$ $i = 1, \cdots, 11;$ $x_i \le 0.04, i = 1, \cdots, 5;$ $x_i \le 0.03, i = 6, \cdots, 14;$	P	1
14	$f(\mathbf{x}) = \sum_{1}^{30}\alpha_i(\mathbf{x})$ $\alpha_i(\mathbf{x}) = 420x_i + (i - 15)^3 + \sum_{j=1}^{30} v_{ij}\left[(\sin(\log(v_{ij})))^5 + (\cos(\log(v_{ij})))^5\right]$ $v_{ij} = \sqrt{x_j^2 + i/j}$	$-2 \le x_i \le 2,$ $i = 1, \cdots, 30$	T	1
15	$f(\mathbf{x}) = \sum_{1}^{d} i (x_i^2 + x_i^4)$	$-2 \le x_i \le 4,$ $i = 1, \cdots, d,$ $d = 20\ or\ 50$	P	1

Chapter 5

Large-scale Metamodeling of Power Transfer Capability Using RBF-HDMR[5]

S. Shan

Dept. of Mechanical and Manufacturing Engineering

University of Manitoba

Winnipeg, MB R3T 5V6

W. Zhang

Manitoba Hydro

System Performance Dept.

453 Dovercourt Drive

Winnipeg, MB, R3C 2P4

G. G. Wang[6]

Simon Fraser University

Engineering Science

250-13450 102 Ave

Surrey, BC V3T 0A3

5.1. Abstract

Metamodeling a large-scale simulation system is challenging due to "curse of dimensionality." This chapter applies the recently developed Radial Basis Function-High Dimensional Model Representation (RBF-HDMR) to model the transfer capability of a large-scale electric power grid subject to contingency disturbances, a problem with 50 variables. The simulation-based transfer capability analysis has a combination of high-dimensional, computationally-expensive, and black-box features, which is called a HEB problem. Two complementary sets of performance metrics are proposed to validate the resultant RBF-HDMR, which is used for power grid characteristic analysis and operation planning. The accuracy and efficiency of the proposed methodology is demonstrated through three practical cases under different operating scenarios. The results show that RBF-HDMR is effective in modeling this HEB problem.

Keywords: power transfer capability, large-scale, high-dimensional, computationally expensive, black-box function, metamodel, RBF-HDMR, engineering optimization, response surface

5.2. Introduction

The power grid analysis is crucial for reliable and efficient system operations. Such analysis is commonly implemented by mean of a simulation process. The power transfer capability on the Manitoba-Ontario interconnection is limited by respecting the operating reliability criteria for the Winnipeg River 115 kV systems in Manitoba, Canada, subject to contingency disturbances. The system operating guides and instructions are derived from extensive simulation studies with a number of generation patterns for the six

hydraulic generating plants on the Winnipeg River and a thermal generating plant at Selkirk, a suburb of Winnipeg. Any deviations from the pre-determined generation patterns would require additional simulation studies or conservative operating limits must be applied. This renders the current study approach with two major drawbacks: (1) time-consuming for expensive simulations and (2) lack of understanding to the underlying power system. Therefore, it is important to derive a methodology that can provide insight into the power grid characteristics and maximize the power transfer capability of the interconnection under forecast system operating conditions without extensive simulation effort.

Metamodeling (also called response surface methodology) [1] provides an effective mechanism for facilitating simulation processes and simplifying the interpretation of simulation results. These techniques construct a mathematical model of simulation analysis by a well-planned sampling scheme to relate the simulation outputs as a function of relevant input factors. The mathematical model of this kind is called a metamodel since it is a model of the simulation model. The planned sampling is also known as the design of computer experiments. After validation, the metamodel often surrogates the expensive simulation processes for system analysis and optimization.

There are various metamodels developed and studied, for example, polynomial response surface [1,2], Kriging model [3,4], radial basis function (RBF) [5,6], and support vector regression (SVR) [7]. These models are often selected based on practitioners' experience or knowledge. The choice of a model type affects metamodeling results. The works [5, 7, 8] compared the performance of some of these metamodels. With advances in metamodeling techniques and increase in demands for application,

some researchers attempt to exploit the best one of multiple fitting metamodels based on certain criteria or constructing a weighed metamodel consisting of multiple individual models to increase metamodels' accuracy and capabilities (see [9-11]).

Metamodeling techniques have been widely used in various disciplines. Kuczera and Mourelatos [12] solved the reliability analysis of multiple failure region problems by the use of metamodels as indicators to determine the failure and safe regions. Yang *et al.* [13] used five metamodeling techniques for vehicle frontal impact simulation. Apley *et al.* [14] applied metamodeling techniques to an engine piston design problem and studied the effort of modeling uncertainty in robust design. Georgopoulou and Giannakoglou [15] proposed a metamodel-assisted evolutionary algorithm to solve power generating unit commitment with probabilistic outrages. Shan *et al.* [16] successfully applied the metamodeling methodologies to the transfer capability of a power grid with five input variables.

From both foundational development and applications of metamodeling techniques, a general impression is that most of these successful cases are low-dimensional problems. HEB problems are challenging for metamodels and sampling methods. Computational complexity for HEB problems arises with the increase in dimensionality, known as "curse of dimensionality." For example, for a problem with 50 variables, if a two-level-per-factor design of computer experiments is taken, 2^{50} simulations are needed. If each simulation needs 13 minutes to run (as is the case for the Manitoba power system in this work), it needs 2.8234×10^{10} years to implement in total. Obviously, the computation expense is formidable for high-dimensional cases. Shan and Wang [17] reviewed the strategies to solve HEB problems. The existing techniques are often limited in their

respective ways. For example, dimensionality reduction techniques are not applicable for problems in which each variable has similar importance. Recently a novel RBF-HDMR [18,19] has been developed based on HDMR theories [20,21]. This work uses the RBF-HDMR to model the transfer capability analysis of a large-scale power grid.

The presentations of this work are organized as follows. Section 5.3 describes the framework of the methodology. Section 5.4 introduces the high-dimensional metamodel, RBF-HDMR, and its sampling scheme. Section 5.5 proposes two complementary sets of performance metrics to validate the resultant metamodel. Section 5.6 presents case studies. Summary is drawn in Section 5.7.

5.3. Framework of Methodology

As introduced, this study employs RBF-HDMR to build a mathematical model of the maximized transfer capability of interconnected power grid on the Manitoba-Ontario interconnection. In specific, the process starts with systematically planned sampling of generation patterns. The generation patterns are inputs to the power grid simulation model, on which various analyses are performed. The output will be the feasibility of the sample point (generation pattern) and the maximum value of the power transfer capability for a given input. Once the samples have been evaluated, a metamodel can be built for the maximized transfer capability as a function of the generation patterns. After validation, the metamodel can be used to plan the interchange across the interconnections under forecast system operating conditions. Figure 5.1 delineates the framework of the methodology.
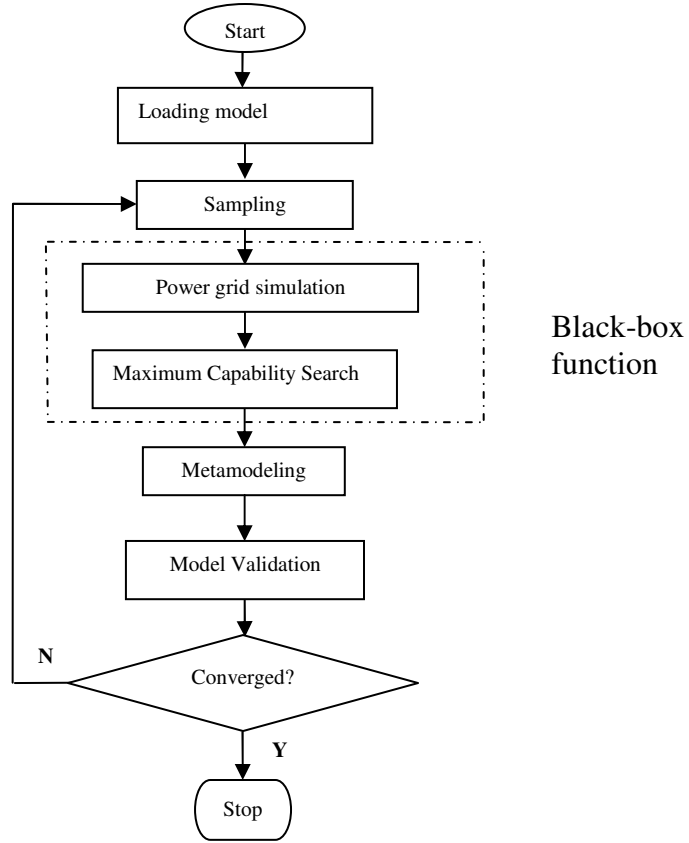
Fig. 5.1 Framework of the methodology

In the procedures of Fig. 5.1, loading model represents starting the simulation model under a specified operating condition; sampling is to simulate the inputs, that is, the generation patterns; contingency analysis for the given input is performed by the power system simulation tool PSS/E^{TM}; an in-house IPLAN [22] program is run to search for the corresponding maximum transfer capability; the maximum transfer capability dependant on the input patterns is modeled by a RBF-HDMR. The RBF-HDMR treats power grid simulation and the maximum transfer capability search in the dash box as a black-box function, that is, it uses the input/output data without the knowledge about how the simulation model processes these inputs to get outputs; after validation, the RBF-HDMR modeling process is completed and the resultant RBF-

HDMR model can be used in applications for system operations planning. This approach provides a solution to effectively maximize the transfer capability of the interconnection under forecast system operating conditions.

5.4. RBF-HDMR techniques

5.4.1. RBF-HDMR model

A general RBF-HDMR model is written as

$$f(\mathbf{x}) \cong f_0 + \sum_{i=1}^{d} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, \mathbf{x}_0^i) - (x_{i_k}, \mathbf{x}_0^i) \right| + \sum_{1 \leq i < j \leq d} \sum_{k=1}^{m_{ij}} \alpha_{ij_k} \left| (x_i, x_j, \mathbf{x}_0^{ij}) - \right.$$

$$\left. (x_{i_k}, x_{j_k}, \mathbf{x}_0^{ij}) \right| + \cdots + \sum_{k=1}^{m_{12\cdots d}} \alpha_{12\cdots d_k} |\mathbf{x} - \mathbf{x}_k|, \tag{5.1}$$

where f_0 denotes the constant term evaluated at $\mathbf{x}_0$ (any chosen reference point in modeling domain); $\mathbf{x}_0^i$, $\mathbf{x}_0^{ij}$, ..., are respectively $\mathbf{x}_0$ without element/s x_i, x_i and x_j, ... (that is, $(x_i, \mathbf{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, $(x_i, x_j, \mathbf{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$, ... respectively represents a point in the modeling domain); d is the number of dimensionality for the input variable vector $\mathbf{x}$; $\left| \cdot \right|$ denotes a p-norm distance; α_{i_k}, α_{ij_k}, ..., $\alpha_{12\cdots d_k}$ are respectively the coefficient of the expression and $(x_{i_k}, \mathbf{x}_0^i)$, $(x_{i_k}, x_{j_k}, \mathbf{x}_0^{ij})$, ..., $\mathbf{x}_k$ are the sampled points of input variables or the centers of radial basis function (RBF) approximation; m_i, m_{ij}, ..., $m_{12\cdots d}$ are the number of sampled points for each term; the component $\sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, \mathbf{x}_0^i) - (x_{i_k}, \mathbf{x}_0^i) \right|$ represents the ith input variable x_i term which explains the effect of the ith input variable x_i independently acting on the output function $f(\mathbf{x})$; the component $\sum_{k=1}^{m_{ij}} \alpha_{ij_k} \left| (x_i, x_j, \mathbf{x}_0^{ij}) - (x_{i_k}, x_{j_k}, \mathbf{x}_0^{ij}) \right|$ denotes the correlated contribution of the variables x_i and x_j upon the output $f(\mathbf{x})$ after the individual

163

influences of x_i and x_j are discounted; the subsequent similar components reflect the effects of increasing numbers of correlated variables acting together upon the output $f(x)$. The last component $\sum_{k=1}^{m_{12\cdots d}} \alpha_{12\cdots d_k}|x - x_k|$ models any residual dependence of all the variables locked together to influence the output $f(x)$ after all the lower-order correlations and individual influence of each involved x_i ($i = 1,\ldots,d$) have been discounted. Without losing generality, note that Eq. (5.1) uses a simple linear spline function as the basis function for the ease of description and understanding; in implementation, a sum of thin plate spline plus a linear polynomial is used to avoid singular matrices. Details of the basis function can be found in Ref. [18].

The functional form of the RBF-HDMR is represented by its structure matrix which is defined as

$$
SM_{d\times n} = \begin{bmatrix}
0\ 1\ 0\ 0\ \ldots\ 0\ 1\ \ldots\ 0\ \ldots\ 0\ \ldots\ 1 \\
0\ 0\ 1\ 0\ \ldots\ 0\ 1\ \ldots\ 0\ \ldots\ 0\ \ldots\ 1 \\
0\ 0\ 0\ 1\ \ldots\ 0\ 0\ \ldots\ 0\ \ldots\ 0\ \ldots\ 1 \\
\ldots \\
0\ 0\ 0\ 0\ \ldots\ 0\ 0\ \ldots\ 0\ \ldots\ 1\ \ldots\ 1 \\
0\ 0\ 0\ 0\ \ldots\ 1\ 0\ \ldots\ 0\ \ldots\ 1\ \ldots\ 1
\end{bmatrix},
\tag{5.2}
$$

where d is the dimensionality of the input variable vector x; n denotes the number of to-be-decided component terms of Eq. (5.1). Each row corresponds to a variable x_i. Each column corresponds to one of the component terms in Eq. (5.1). Each element in the structure matrix is assigned as "0" or "1" ; "0" means that the variable does not appear in the component term; "1" means that the variable appears in the component term. For example, the column $[0_1, 0_2, \ldots, 0_d]^T$ denotes the constant component term f_0; $[0_1, 0_2, \ldots, 1_i, \ldots, 0_d]^T$ represents the first-order component term $f_i(x_i)$;

$[0_1, 0_2, \dots, 1_i, \dots, 1_j, \dots, 0_d]^T$ indicates the existence of the second-order component term $f_{ij}(x_i, x_j)$, and the last column indicates the existence of d-variate correlation component term $f_{12\dots d}(x_1, x_2, \dots, x_d)$. In total, there are 2^d number of component terms; however, many components may not exist and some others disappear due to their negligible contribution to $f(x)$. The structure matrix is gradually generated with the RBF-HDMR modeling process. It indexes uniquely the component terms and explicitly expresses the functional form of black-box functions [19].

The RBF-HDMR has many attractive features: 1) has an explicit functional form to "reveal" correlation among the variables; 2) provides with (non)linearity information with regard to inputs; 3) discloses relative importance of variable terms, and 4) efficiently and effectively models large scale complex systems (for example, high-dimensional nonlinear problems) [18,19].

5.4.2. Sampling and Metamodeling Scheme

The sampling scheme of RBF-HDMR is illustrated by Fig. 5.2. The sampling is carried out sequentially from the lower to higher orders (0D, 1D, 2D, 3D, …,) and the set of samples is gradually augmented by new sampling points. In Fig. 5.2, the dots denote new sampling points. The circles denote the sampling points inherited from previous order/s. The inherited sampling points in Fig. 5.2d are ignored for clarity. Those circles are distributed on orthogonal axes or planes passing the reference points (the center point), as shown in both Figs. 5.2b and 5.2c. Starting from the second-order (2D), new sampling points are not on those axes or planes. As seen from Fig. 5.2c and 5.2d, the sampling points are augmented with corner points first, then points close to center points

of each quadrant, if needed, and at last other areas that need more exploration according to the characteristics of the underlying function. More detailed sampling steps are described as follows.
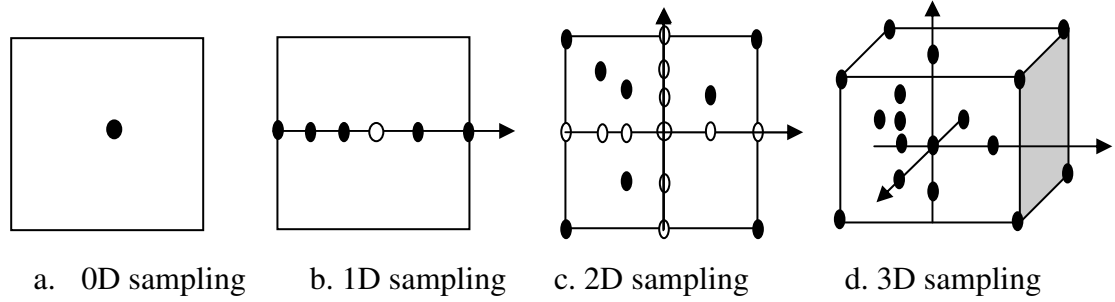

| a. 0D sampling | b. 1D sampling | c. 2D sampling | d. 3D sampling |

Fig. 5.2 An illustration of RBF-HDMR sampling scheme up to the 3rd order

1. Choose a reference point $\boldsymbol{x_0} = [x_{1_0}, x_{2_0}, \cdots, x_{d_0}]^T$ in the modeling domain (see Fig. 5.2a). It is recommended that taking the point $\boldsymbol{x_0}$ makes $f_0 = \bar{f}$ (that is, the mean of the function $f(\boldsymbol{x})$ in the modeling domain). Due to unknown mean of the function, this work takes one point in the neighborhood of the center of the modeling domain. Evaluating $f(\boldsymbol{x})$ at $\boldsymbol{x_0}$, we then have f_0. According to the theory of HDMR and the premise, the reference $\boldsymbol{x_0}$ is irrelevant if the model converges.

2. Sample for the first-order component functions $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, \boldsymbol{x_0^i}) - (x_{i_k}, \boldsymbol{x_0^i})| = f([x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T) - f_0$ in the close neighborhood of the two ends of x_i (lower and upper limits) while fixing the rest of x_j ($j \neq i$) components at $\boldsymbol{x_0}$. In this work, a neighborhood is defined as one percent of the variable range which is in the design space and near a designated point. Evaluating these two end points, we got the left point value $\alpha_{i_L} |(x_i, \boldsymbol{x_0^i}) - (x_{i_L}, \boldsymbol{x_0^i})| = f([x_{1_0}, x_{2_0}, \cdots, x_{i_L}, \cdots, x_{d_0}]^T) - f_0$ and the right point value $\alpha_{i_R} |(x_i, \boldsymbol{x_0^i}) - (x_{i_R}, \boldsymbol{x_0^i})| = f([x_{1_0}, x_{2_0}, \cdots, x_{i_R}, \cdots, x_{d_0}]^T) - f_0$ and define the component

166

function as $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$ by finding coefficients α_{i_L}, α_{i_R}, ... for each variable x_i (see Fig. 5.2b).

3. Check the linearity of $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$. If the approximation model $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$ goes through the center point, $x_0, \sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$ is considered as linear. In this case, modeling for this component terminates; otherwise, use the center point x_0 to update $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$. Then a random value along x_i is generated and combined with the rest of x_j ($j \neq i$) components at x_0 to form a new point to test $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$. If $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$ is not sufficiently accurate (the relative prediction error is larger than a given criterion, for instance, 0.01%), the test point and all the evaluated points will be used to re-construct $\sum_{k=1}^{m_i} \alpha_{i_k} |(x_i, x_0^i) - (x_{i_k}, x_0^i)|$. This sampling-remodeling process iterates until convergence. This process is to capture the nonlinearity of the component function with one sample point at a time. Step 3 repeats for all of the first-order component functions to construct the first-order terms of RBF-HDMR model individually (see Fig. 5.2b).

4. Form a new point, $(x_i, x_j, \cdots x_d)_k = [x_{1_k}, x_{2_k}, \cdots, x_{i_k}, \cdots, x_{j_k}, \cdots, x_{d_k}]^T, k \neq 0$ by randomly combining the sampled value x_i in the first-order component construction for each input variable (that is, x_i, $i = 1, ..., d$ and evaluated at (x_i, x_0^i), respectively). This new point is then evaluated by expensive simulation and the first-order RBF-HDMR model. The values of expensive simulation and model prediction are compared. If the two values are sufficiently close (the relative error is less than a small value, for example,

0.01%), it indicates that no higher order terms exist in the underlying function, the modeling process terminates. Otherwise, go to the next step. This new point does not appear in Fig. 5.2 since it has high dimensionality.

5. Use the values of x_i and x_j, $i \neq j$ that exist in thus-far evaluated points $(x_i, \boldsymbol{x}_0^i) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_{d_0}]^T$, and $(x_j, \boldsymbol{x}_0^j) = [x_{1_0}, x_{2_0}, \cdots, x_j, \cdots, x_{d_0}]^T$ to form new points of the form $(x_i, x_j, \boldsymbol{x}_0^{ij}) = [x_{1_0}, x_{2_0}, \cdots, x_i, \cdots, x_j, \cdots, x_{d_0}]^T$. Randomly select one of the points from these new points (that is, one of four corners) to test the first-order RBF-HDMR model. The corner points on x_i and x_j planes (see Fig. 5.2c) have high priority. If the model passes through the new point, it indicates that x_i and x_j are not correlated and the process continues with the next pair of input variables. This is to save the cost of modeling non-existing or insignificant correlations; otherwise, use this new point and the evaluated points $(x_i, \boldsymbol{x}_0^i)$ and $(x_j, \boldsymbol{x}_0^j)$ to construct the second-order component function, $\sum_{k=1}^{m_{ij}} \alpha_{ij_k} |(x_i, x_j, \boldsymbol{x}_0^{ij}) - (x_{i_k}, x_{j_k}, \boldsymbol{x}_0^{ij})|$. This sampling-remodeling process iterates for all possible two-variable correlations until convergence (the relative prediction error is less than 0.01%). Step 5 is repeated for all pairs of input variables.

Step 5 applies to all higher-order terms in RBF-HDMR model, Eq. (5.1), in a similar manner, which is shown in Fig. 5.2d. Ref. [19] described in detail how further computation savings are achieved by employing derived theorems for higher-order terms.

5.5. Model Validation

Metamodels need to be validated prior to use. There are a collection of papers addressing metamodel validation (for instance, Refs. [23, 24]). The validation depends on the purpose of metamodeling [25]. Different purposes lead to various validation performance metrics and approaches. For example, Kleijnen and Sargent [25] prefer the absolute error than the absolute relative error while the latter is deficient; if a single large error is catastrophic (for example, nuclear simulations), the maximum error will be an important performance metric. Different validation performance metrics have their pros and cons [26]. Single performance metric reflects one aspect of a metamodel and none of a single performance is the best criterion for all circumstances.

The goodness of a metamodel approximating an unknown function is commonly measured by some distance. The distance measure includes L_1 distance and L_2 distance. L_1 distance is the absolute value of the piecewise difference between the actual value $f(x_i)$ and predicated value $\hat{f}(x_i)$; while L_2 distance provides the square root of the piecewise square of the difference between the actual value $f(x_i)$ and predicted value $\hat{f}(x_i)$. Whether the distance is scaled or mathematically changed further leads to different expressions to form various performance metrics. The scaled performance metrics are often used to eliminate the influence of variations in function units and scales. Due to diversity of underlying systems' features and lack of *a priori* knowledge about these diversities, one performance metric or even one class of similar performance metrics may not be able to reflect the goodness of metamodels. This work introduces two complementary sets of performance metrics: function-value-scaled and function-

variability-scaled groups. The former is relatively scaled to the values of functions (based on L_1 distance). The latter is relatively scaled to the variability of functions (for instance deviations or variations of functions, based on L_2 distance).

5.5.1. Function-value-scaled Set

This group of performance metrics addresses the goodness of approximation models with regard to the values of underlying systems.

1) Mean Absolute Relative Error (MARE)

A mean absolute percentage error is defined as

$$MARE = \frac{100}{m} \sum_{i=1}^{m} e_i ,$$

$$e_i = \begin{cases} \left|\frac{f(x_i)-\hat{f}(x_i)}{f(x_i)}\right|, & |f(x_i)| \geq 1 \\ \min\left(\left|\frac{f(x_i)-\hat{f}(x_i)}{f(x_i)}\right|, \frac{|f(x_i)-\hat{f}(x_i)|}{(|f(x_i)|+|\hat{f}(x_i)|)/2}\right), & |f(x_i)| < 1 \end{cases}, \quad i = 1, \cdots, m. \qquad (5.3)$$

The relative error e_i piecewise measures the error of the model. Note that relative errors $e_i = \left|\frac{f(x_i)-\hat{f}(x_i)}{f(x_i)}\right|$ is inflated when function values approach or equal to zero. Assuming a small number of inflating points in the validation point set, this inflation drawback is alleviated by using $e_i = \frac{|f(x_i)-\hat{f}(x_i)|}{(|f(x_i)|+|\hat{f}(x_i)|)/2}$ just in case.

2) Maximum Absolute Relative Error (MxARE)

A maximum absolute relative error is

$$MxARE = 100 * max\left(\left|\frac{f(x_1)-\hat{f}(x_1)}{f(x_1)}\right|, \left|\frac{f(x_2)-\hat{f}(x_2)}{f(x_2)}\right|, \cdots, \left|\frac{f(x_m)-\hat{f}(x_m)}{f(x_m)}\right|\right). \qquad (5.4)$$

Note that the absolute relative error definition is not modified to address the deficiency in the vicinity whereas $f(x_m)$ approaches zero. This is to expose the real behavior of the underlying system in extreme cases. MARE and MxARE can complement each other within this set of performance metrics, as one describes the average and the other the maximum relative error, respectively.

3)   Median Absolute Relative Error (MdARE)

"Median" in median absolute relative error (MdARE) is defined as-

$$MdARE = 100 * median \left(\left| \frac{f(x_1)-\hat{f}(x_1)}{f(x_1)} \right|, \left| \frac{f(x_2)-\hat{f}(x_2)}{f(x_2)} \right|, \cdots, \left| \frac{f(x_m)-\hat{f}(x_m)}{f(x_m)} \right| \right). \quad (5.5)$$

In this set of performance metrics, MARE is a global performance metric. The smaller the value of MARE is, the better the metamodel is in general; $MxARE$ is a local performance metric. It not only reflects the status of one sample point, but also signals whether inflation appears in MARE. If an inflation appears in $MxARE$, MARE must be modified. The deficiency of inflation thus can be dealt with. MdARE assists MARE and $MxARE$ to measure a metamodel.

5.5.2. Function-variability-scaled Set

This group of performance metrics deals with the goodness of approximation models with regard to the variability of underlying systems.

1) R Square

The error of a function $f(x)$ being approximated by $\hat{f}(x)$ can be estimated by normalized Euclidean distance as

$$\varepsilon = \frac{1}{V}\int[f(x)-\hat{f}(x)]^2 dx. \tag{5.6}$$

If the function is square integrable,

$$\varepsilon = \frac{1}{VM}\sum_{i=1}^{M}[f(x_i)-\hat{f}(x_i)]^2 \tag{5.7}$$

while sampling is considered,

where V denotes the variance of $f(x)$. The smaller is the value of ε, the better is the approximation. However, this metric depends on the variance V. The variance V reflects the intrinsic variability feature of $f(x)$. According to the definition of variance, we rewrite

$$\varepsilon = \frac{\sum_{i=1}^{M}[f(x_i)-\hat{f}(x_i)]^2}{\sum_{i=1}^{M}[f(x_i)-\bar{f}(x)]^2} , \tag{5.8}$$

where $\bar{f}(x)$ is the mean of the function $f(x)$ in the defined domain. Likewise, the error ε becomes inflated while the function output is or close to be a constant.

The goodness of modeling is often estimated by R Square

$$R^2 = 1 - \frac{\sum_{i=1}^{m}[f(x_i)-\hat{f}(x_i)]^2}{\sum_{i=1}^{m}[f(x_i)-\bar{f}(x_i)]^2} . \tag{5.9}$$

Obviously,

$$R^2 = 1 - \varepsilon . \tag{5.9'}$$

Normally, if an approximation has $\varepsilon \ll 1$ or R square approaches one, then this approximation has a good accuracy with regard to variability. Similar to (MARE), R square is deficient when V is close to zero. Let us assume that $f(x)$ in the defined domain is or close to a constant, that is, the mean of $f(x)$, and when $\hat{f}(x_i)$ closely approaches $f(x_i)$ (therefore, a good approximation), we have

$$\varepsilon = \lim_{\substack{f(x) \to \bar{f}(x) \\ f(x_i) \to f(x_i)}} \frac{\sum_{i=1}^{M}[f(x_i) - \hat{f}(x_i)]^2}{\sum_{i=1}^{M}[f(x_i) - \bar{f}(x)]^2} = 1 \ , \tag{5.10}$$

In RBF-HDMR (Eq. 5.1), the roughest approximation is the 0th order approximation $f(x) \cong f_0$. Then, the best 0th order approximation is $f_0 = \bar{f}$. Assuming the output $f(x)$ is a constant, $f(x) \equiv f_0 \equiv \bar{f}$ indicates that the $\hat{f} = f_0$ accurately models $f(x)$ and there is no need for higher-order terms. This corresponds to the scenario of $\varepsilon = 1$ or R square is 0. Hence if $\varepsilon \cong 1$ or R Square approaches zero, one cannot conclude the approximation is poor.

We caution that our R Square differs from the classical R Square (coefficient of determination [27,28]) although both R Square have an identical formula. In our case, R Square is measured on validation points, not on the modeling points. The evaluation of R Square is also not dependent on the types of metamodels.

2) Relative Average Absolute Error (RAAE)

$$RAAE = \frac{\sum_{i=1}^{m}|f(x_i) - \hat{f}(x_i)|}{m * STD} , \tag{5.11}$$

where *STD* stands for standard deviation,

$STD = \sqrt{\frac{\sum_{i=1}^{m}[f(x_i)-\bar{f}(x)]^2}{m}}$. Like R square, this metric shows the overall accuracy of an approximation model. The closer the value of RAAE approaches zero, the more accurate is the approximation model. Similar to the discussion for R square, one can use this metric to find if a model is accurate regarding variability. The caveat is when the function is almost a constant,

$$RAAE = \frac{\sum_{i=1}^{m}|f(x_i)-\hat{f}(x_i)|}{m*STD} = \frac{\sum_{i=1}^{m}|f(x_i)-\hat{f}(x_i)|}{\sqrt{m*\sum_{i=1}^{m}[f(x_i)-\bar{f}(x)]^2}} = \frac{1}{\sqrt{m}}\lim_{\substack{f(x)\to\bar{f}(x) \\ f(x_i)\to f(x_i)}} \frac{\sum_{i=1}^{m}|f(x_i)-\hat{f}(x_i)|}{\sqrt{\sum_{i=1}^{m}[f(x_i)-\bar{f}(x)]^2}} = \frac{1}{\sqrt{m}}.$$

(5.12)

In this case, RAAE value is not zero but approaches to $\frac{1}{\sqrt{m}}$. Therefore RAAE bears similar disadvantage as R Square when the output is or close to be a constant.

3) Relative Maximum Absolute Error (RMAE)

$$RMAE = \frac{\max(|f(x_1)-\hat{f}(x_1)|,|f(x_2)-\hat{f}(x_2)|,\cdots,|f(x_m)-\hat{f}(x_m)|)}{STD}.$$

(5.13)

This is a local metric. A RMAE describes error in a sub-region of the design space. Therefore, a small value of RMAE is preferred. Similarly the variation of STD affects RMAE.

Equations 5.3-5.13, defines two sets of performance metrics. Each set of performance metrics explains the accuracy of models from different perspectives. In each set, each performance metrics reflects the accuracy of metamodels from one standpoint, and they work together to reflect the overall goodness of metamodels. The judgment of

goodness of a metamodel is related to the nature of the underlying system and application requirements of the metamodel. For example, some application requires that maximum error not exceed a limit, while others need the global metamodel has minimum total errors. This work employs global performance metrics (mean absolute relative error (MARE) and relative average absolute error (RAAE)) from each group as convergence criteria for metamodeling. We set MARE as 0.1 and RAAE as 0.3; either one criterion is met, the modeling process stops. Other performance metrics are employed as a supplement to explain the local performance of models.

5.6. Case Studies

This section first introduces the power grid system of interest, and then describes three different practical operating scenarios to be studied, and at last the gained understandings of power grid and transfer capability are summarized.

5.6.1. Power Grid Description

Figure 5.3 shows a single line diagram of the Winnipeg River area system and Manitoba-Ontario interconnections. The 115kV transmission system interconnects the Winnipeg River generating plants, Selkirk generating station, Ontario-Manitoba Tie (OMT), and the major 230kV transmission grid surrounding the City of Winnipeg. There are totally six hydraulic generating plants on the Winnipeg River. Four plants, Seven Sisters (6 generators), Great Falls (6 generators), McArthur Falls (8 generators), and Pine Falls (6 generators) are connected through the 115kV transmission system. The other two plants, Pointe du Bois (PDB) (16 generators) and Slave Falls (8 generators), feed radically into the City of Winnipeg 66kV system. Selkirk generating station is a gas fired

thermal plant located near the City of Winnipeg and connected to the 115kV transmission system. The Ontario-Manitoba interconnection consists of two 230kV tie lines from Manitoba to northwestern Ontario. The interface is controlled by the 115kV phase shifting and 115/230kV voltage regulating transformers at Whiteshell station near the Manitoba-Ontario border. Generation levels of the hydraulic plants on the Winnipeg River are a function of river system management and economic operation of the plants. The total generation levels can vary significantly from a maximum of 593 MW to minimum of 298 MW depending on the river flow. The Selkirk generating plant has two units of 65 MW each and is operated when required to regulate the system reliability. It is from the operating experience that the Selkirk generation has a unique impact on the system performance; so, it will be interesting to study this impact on the transfer capability of the interconnection. This information will be useful in coordinating generation operations in the concerned areas.

The case studies are for the winter load with all the transmission lines in service. The hydraulic generation levels on Winnipeg River are shown in 5.9 Appendix. They are organized by generating plants and generators. Temperature is assumed at a 0°C representing winter operations, therefore, the metamodel will not show the sensitivity to the temperature change for the case studies. Committing to the generation units, this simulation proceeds with 50 inputs (number of hydraulic generators) and 1 parameter (thermal generators at Selkirk for system reliability). The transfer capability of this simulation-based power grid can be abstracted out as $f = f(x, \rho)$. The input x representing the input generators and the parameter ρ denoting the status of the generators at Selkirk determine a specific grid operating scenario. It takes about thirteen

176

minutes on a desktop computer (Pentium 4 CPU 2.53 GHz) to evaluate one simulation. Even a huge number of simulation runs cannot possibly cover the entire 50-dimensional input variable space. Scattered outcomes of sporadic simulations do not provide a global understanding of the functional relationship between the transfer capability (f) and generation patterns (x, ρ). Three case studies using RBF-HDMR are presented in Section 5.6.2.
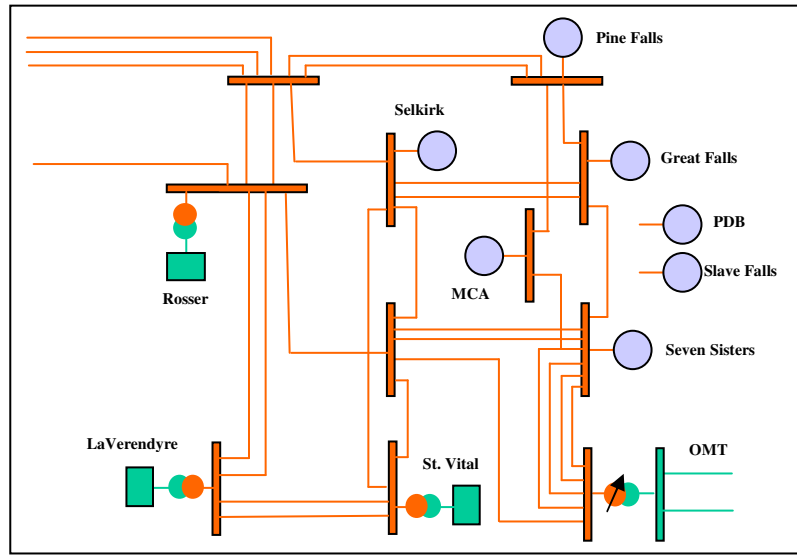
Fig. 5.3 Power system in Winnipeg River area

5.6.2. Three Model Cases

Three cases are modeled, which include the cases that no Selkirk generators are online ($\rho = 0$), one Selkirk generator with 65 MW is online ($\rho = 1$), and two Selkirk generators with 130 MW are online ($\rho = 2$), respectively. The two sets of performance metrics proposed in Section 5.5 are used to validate metamodels. In order to observe the changes of these performance metrics, five batches of 100 random test points (non-

modeling points) are sequentially and accumulatively used to evaluate the metamodels against the performance metrics until all total 500 test points are used. The values of these performance metrics show the prediction capability of the RBF-HDMR on new input points.

Case 1 No Selkirk Generators on Line ($\rho = 0$)

This metamodel has been constructed by 315 expensive simulation points. Table 5.1 lists the performance metric values of accumulated samples of five batches. It shows

Table 5.1 Performance metric values of Case 1

Performance metric		Validation Sample Size				
		100	200	300	400	500
Function-value-scaled Set	MARE (%)	6.27	6.72	6.56	6.61	6.57
	MxARE (%)	20.23	20.23	20.23	20.23	20.23
	MdARE (%)	4.89	5.63	5.53	5.64	5.64
Function-variability-scaled Set	RSquare	0.02	0.00	-0.01	-0.02	0.01
	RAAE	0.79	0.83	0.83	0.83	0.83
	RMAE	2.50	2.48	2.52	2.75	2.75

that the mean absolute relative error (MARE) is less than 7%. The maximum absolute relative error (MxARE) in 500 validation points is 20.23%. The medium absolute error (MdARE) is less that 6%. The values of function-value-scaled performance metrics indicate that resultant RBF-HDMR metamodel has small absolute relative error. R Square swings around zero, which indicates the metamodel may not be a good approximation or

system output approaches a constant. Such situation needs other performance metrics to judge. Relative Average Absolute Error (RAAE) stays at 0.83, which explains the output approach to a constant. Relative Maximum Absolute Error (RMAE) vibrates at 2.5. If one judges this resultant metamodel only by the values of the function-variability-scaled performance metrics, the misleading (that is, bad metamodeling) may occur.
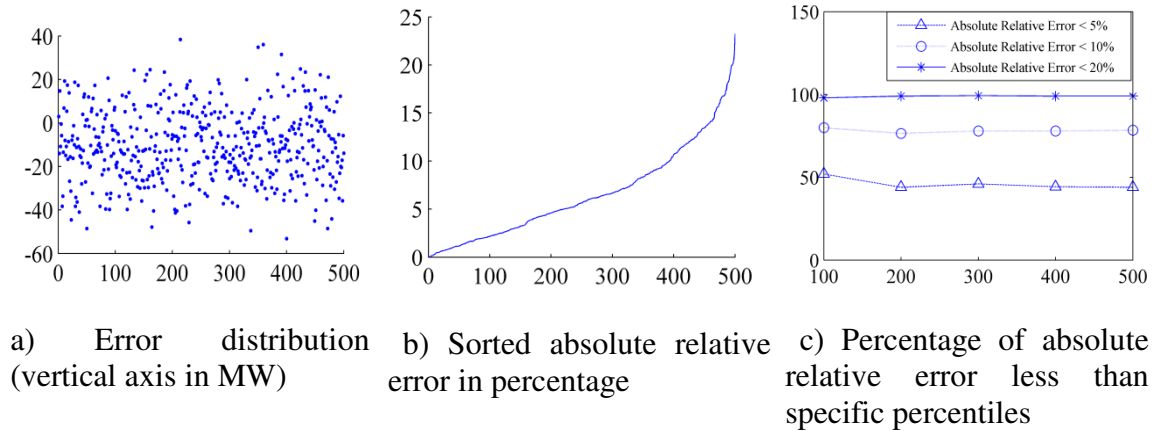

a) Error distribution (vertical axis in MW)

b) Sorted absolute relative error in percentage

c) Percentage of absolute relative error less than specific percentiles

Fig. 5.4 Error plots for Case 1

Figure 5.4a shows the error distribution of 500 randomly sampled points. In Fig. 5.4a, the horizontal coordinate provides the order of the sampled points, and the vertical coordinate represents the errors. The mean of the errors is approximate 10 mega watts (MW), which equals to the predetermined increments of power transfer for OMT. This error means that the power transfer capability as the function output is discretely simulated at 10 MW increments. It is believed that the mean error may mostly be due to this discreteness. Figure 5.4b sorts absolute relative errors of 500 test samples. It demonstrates that the 80% test samples have relative errors of less than 10%. Figure 5.4c displays the percentage changes of the absolute relative error less than the specific percentiles 5%, 10% and 20%, respectively. In Fig. 5.4c, the horizontal coordinates

denotes the number of accumulated samples of five batches and the vertical coordinate represents the percentages. It can be seen that almost 100% sample points have less than 20% relative errors, and about 42% sample points have less than 5%. The percentage of all specified percentiles becomes stable as the test sample size increases.
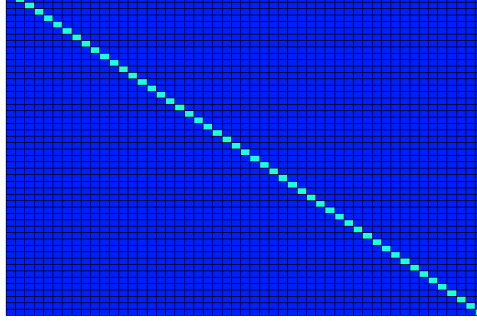

Fig. 5.5 Structure matrix for Case 1

Figure 5.5 shows the structure matrix of this case, which uncovers the functional form of the transfer capability with respect to the output of generators. The functional form consists of the constant term and the first-order terms only, which shows no significant correlations between generators. In all first-order terms, some terms are linear, and others are nonlinear. The linearity/non-linearity of each component is stored in the final model. In more detail, the functional form is written as

$$f(x) \cong$$

$$f_0 + \sum_{i=1\cdots8,20,24,27\cdots50} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, x_0^i) - (x_{i_k}, x_0^i) \right| +$$

$$\sum_{i=9,\cdots,19,21,22,23,25,26} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, x_0^i) - (x_{i_k}, x_0^i) \right|, \tag{5.14}$$

where $f_0 = 264MW$, the coefficients $\boldsymbol{\alpha}_i$ for each input variable x_i in the second term of Eq. (5.14) is zero, which shows both linearity and no contribution to the variation of

transfer capability $f(x)$. The variation of transfer capability $f(x)$ is brought about by the last term. The magnitude of coefficients embodies the sensitivity of corresponding generators to the transfer capability. Due to high-dimensionality and limited space, we choose to plot the impact curves of three generators in Fig. 5.6. In Fig. 5.6, the horizontal axis denotes the outputs of a generator and vertical axis represents the transfer capabilities. The dots display the modeling points. Figure 5.6a is the plot of the first generator at Great Falls, Fig. 5.6b is the plot of the first generator at Seven Sisters and Fig. 5.6c is the plot of the first generator at Pine Falls.

a) generator at Great Falls b) generator at Seven Sisters c) generator at Pine Falls
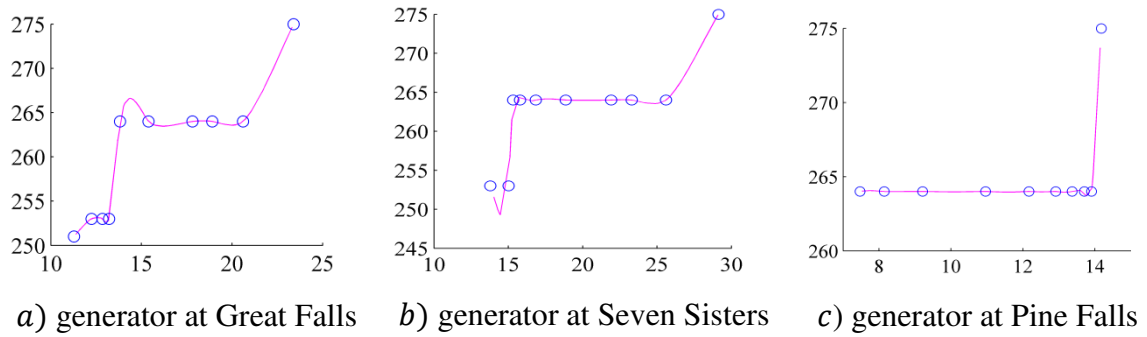
Fig. 5.6 Transfer capability impact curves of three generators; x axis shows the output from each generator and y axis is the power transfer at OMT. Unit: MW

Case 2 One Selkirk Generator on Line ($\rho = 1$)

For this case, one Selkirk generator with 65 MW is on line ($\rho = 1$). The resultant metamodel is built by 251 expensive simulation points. Table 5.2 displays the values of two sets of performance metrics with various size test samples. The mean absolute relative error (MARE) reaches 0.03%. The maximum absolute relative error ($MxARE$) is 4.9%. The medium absolute relative error (MdARE) is zero. The values of the function-

181

value-scaled performance metrics indicate that the resultant RBF-HDMR accurately predicts the transfer capability. All R Square values are negative and close to zero. This phenomenon indicates that the current metamodel is either not accurate enough to approximate the simulation system, or the simulation function values are close to a constant. The final judgment needs to be supported by other performance metrics. Relative average absolute error (RAAE) has small values as the test sample size increases. Relative Maximum Absolute Error (RMAE) increases as the standard deviation decreases. The relative error of all test samples is less than 5%. All these performance metrics support that the transfer capability of the power system is a constant of 300 MW and the final RBF-HDMR model is a good approximation of the transfer capability under this operating scenario.

Table 5.2 Performance metric values of case 2

Performance metric		Validation Sample Size				
		100	200	300	400	500
Function-value-scaled Set	MARE (%)	0.08	0.05	0.04	0.03	0.03
	MxARE (%)	4.90	4.90	4.90	4.90	4.90
	MdARE (%)	0	0	0	0	0
Function-variability-scaled Set	R Square	-0.02	-0.02	-0.02	-0.01	-0.01
	RAAE	0.16	0.14	0.14	0.13	0.13
	RMAE	9.7	13.3	15.9	18.1	19.8

Among additional 500 random samples, at 482 points the predicated value equals to the actual value; 4 have minus one MW errors; 13 have minus two MW errors; and only

one has minus fourteen MW error. For this case, since the transfer capability is a constant 300 MW, the influence of the 10 MW increment does not exist. The 13 smaller errors may be due to the control dead-band (acceptable deviation) of the phase shifting transformers. Other is believed to be an outlier. For clarity, Fig. 5.7 shows the error distribution. It shows that the errors distribute close to zero.
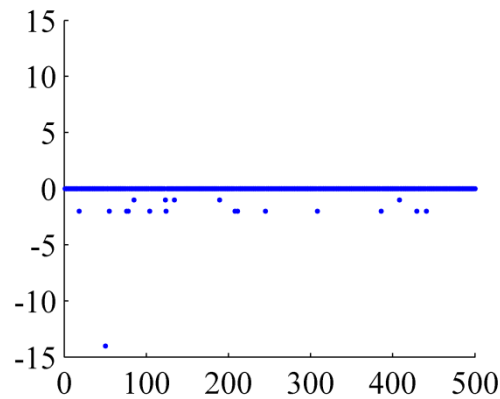
Fig. 5.7 Error distribution of Case 2 (vertical axis shows errors)

Figure 5.8 plots the statistical data of 500 random samples for this case. The top plot displays the change of the mean of different test sample sizes. It can be seen that the mean of this approaches the constant 300 MW. The middle plot shows the standard deviation of the samples. The deviation becomes smaller as the sample size increases. The bottom plot explains the variance of this case. The variance becomes smaller as the sample size increases as well.
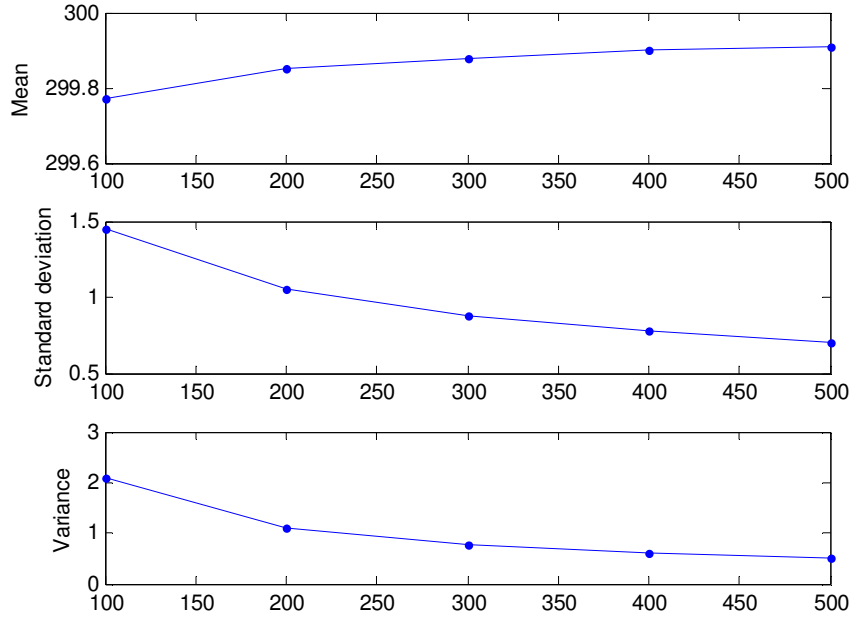
Fig. 5.8 Statistics of transfer capability for Case 2 (vertical axis unit: MW)

The structure matrix of Case 2 shows a similar structure as that of Case 1, as shown in Fig. 5.4, which indicates there is no significant correlation between generators. However, the functional form can be written as

$$f(x) \cong f_0 + \sum_{i=1}^{50} \sum_{k=1}^{m_i} \alpha_{i_k} \left| \left(x_i, x_0^i \right) - \left(x_{i_k}, x_0^i \right) \right|, \tag{5.15}$$

where $f_0 = 300MW$, all the coefficients $\boldsymbol{\alpha}_i$ for each input variable x_i are close to be zeroes. All of first-order terms are linear. The linearity/non-linearity of each component can be found in the final model. No generator causes significant changes of the transfer capability. In other words, this indicates that the transfer capability reaches a stable state of 300 MW.

Case 3 Two Selkirk Generators on Line ($\rho = 2$)

In Case 3, two Selkirk generators with 130 MW are on line ($\rho = 2$). The cost of the final metamodel consists of 284 expensive evaluation samples. Table 5.3 gives the performance metric values of various test samples. The mean absolute relative error (MARE) is approximate 0.4%. The maximum absolute relative error (MxARE) has 9.04%. The medium absolute relative error (MdARE) is 0.05%. The values of function-value-scaled performance metrics display the accuracy of the final RBF-HDMR prediction. R Square values swings around zero which indicates that the transfer capability is around some certain fixed value. Relative Average Absolute Error (RAAE) is under 0.4.

Table 5.3 Performance metric values of case 3

Performance metric		Validation Sample Size				
		100	200	300	400	500
Function-value-scaled Set	MARE (%)	0.42	0.39	0.36	0. 35	0. 36
	MxARE (%)	9.04	9.04	9.04	9.04	9.04
	MdARE (%)	0.04	0.05	0.05	0.05	0.05
Function-variability-scaled Set	RSquare	0.09	0.01	-0.03	-0.04	0.01
	RAAE	0.30	0.35	0.38	0.39	0.36
	RMAE	6.06	7.65	8.91	9.53	8.37

Figure 5.9a displays the predicted and actual values of 500 random test samples. For clarity, Fig. 5.9b plots the trends of the first 100 of 500 test samples, which shows the final RBF-HDMR captures the overall trends of the model behavior. Figure 5.10a shows

the error distribution. Those errors distribute close to zero except for thirteen points. Some of these errors may be due to the 10 MW simulation increment. In other words, the transfer capability may not be at 300 MW for some generation patterns. Some others

a) Actual and predicted values of Case 3

b) Trend of the first 100 random test samples

Fig. 5.9 Values of case 3

a) Error distribution

b) Absolute relative errors
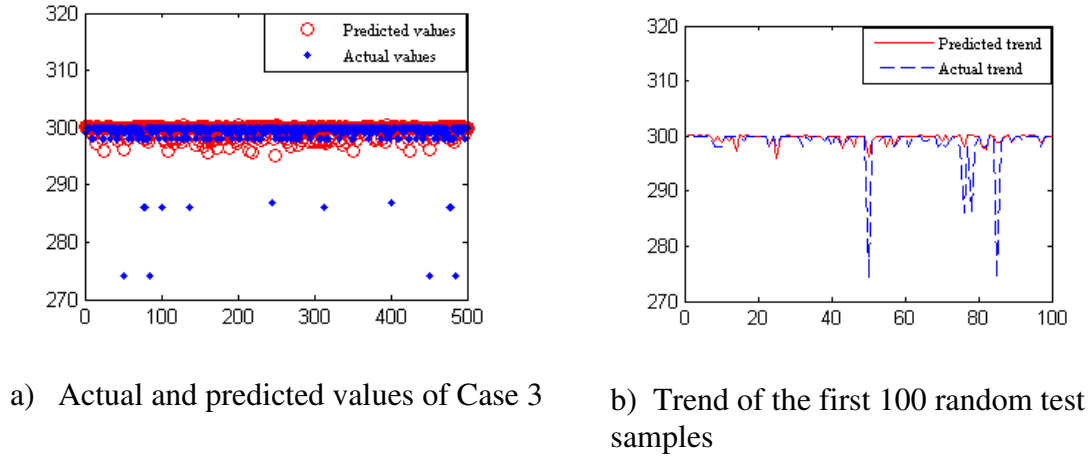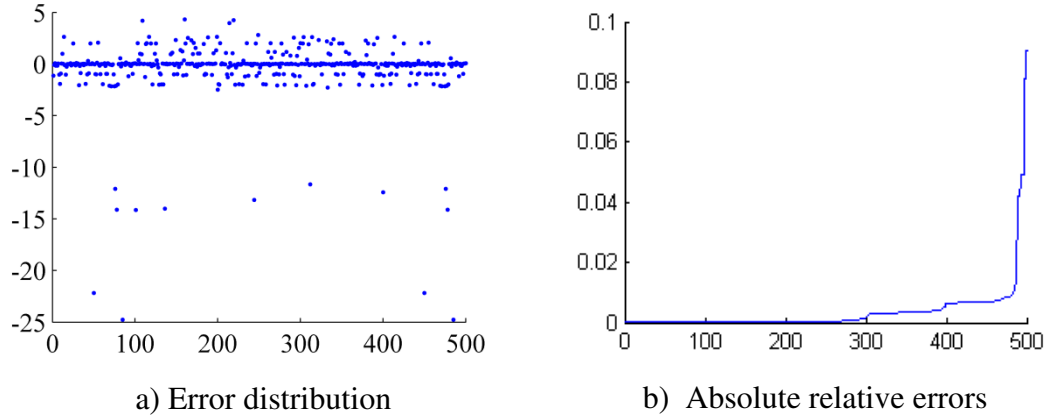
Fig. 5.10 Errors of case 3

may be outliers. Figure 5.9b shows the absolute relative errors. These relative errors are small except for thirteen points. The outputs are close to a constant, that is, its variance is small. The mean of various sizes of samples is approximate 299 MW. It can be seen that

final RBF-HDMR predicts the outputs of the simulation function except for thirteen points.

The structure matrix is similar to that of Case 1, which still shows no correlation between input generators. The functional form is

$$f(x) \cong$$

$$f_0 + \sum_{i=1\cdots8,10,12,13,19,\cdots,50} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, x_0^i) - (x_{i_k}, x_0^i) \right| +$$

$$\sum_{i=9,11,14,\cdots,18} \sum_{k=1}^{m_i} \alpha_{i_k} \left| (x_i, x_0^i) - (x_{i_k}, x_0^i) \right|, \tag{5.16}$$

where $f_0 = 300MW$, the coefficient α_i in the second term is a $\mathbf{0}$ vector, and the linearity/nonlinearity of these components can be obtained in the final metamodel. The variation of transfer capability $f(x)$ is brought by the last term. The nonlinearity information of the components in the last term can be found in the model. The magnitude of coefficients reflects the sensitivities of generators. Three generators in Great Falls and two generators in Seven Sisters contributed to variation of transfer capability $f(x)$.

5.6.3. Understanding of Power Grid and Transfer Capability

In Section 5.6.2, three model cases have been developed by RBF-HDMR with satisfactory results. This conclusion can be drawn from the values of performance metrics and statistical analysis of test data. The resultant RBF-HDMRs capture the characteristics of power grid under different state parameter ρ and achieve good accuracy of the transfer capability prediction for Ontario-Manitoba interconnection. Each model finds the

functional form of underlying power grid which discloses the relative importance of generators to the power transfer capabilities. As the nonlinearity of underlying power grid changes, the functional forms of the resultant models change, which can be seen from comparison of Eqs. 5.14, 5.15 and 5.16. The modeling cost varies. As the nonlinear terms increase, the modeling cost increase. The underlying system model in Case 1 has the most nonlinear terms in Eq. (5.14) and costs 315 simulation points; the model in Case 3 has less nonlinear terms in Eq. (5.16) and cost 284 simulation points; and the linear underlying system model in Case 2 has no nonlinear terms and costs only 251 simulation points. The resultant models show that the thermal generating plant at Selkirk plays a critical role in determining the transfer capability over OMT interconnection. For Case 1 when both generators in Selkirk are off line, the power transfer capability is strongly impacted by the hydraulic generation patterns. This impact is highly nonlinear and exerted mainly by generators in three hydraulic generating plants (Great Falls, Seven Sisters and Pine Falls). With one generator at Selkirk (Case 2) on line, the transfer capability is beyond 300 MW. Any individual generator will not impact the transfer capability at this transfer level. When two generators at Selkirk are on line (Case 3), the transfer capability of the power grid manifests low-order nonlinearity, which means the stability of transfer capability becomes worse than in Case 2. The transfer capability is mildly impacted by generators from two generating plants (Great Falls and Seven sisters). These observations are supported by the system operating experience. In summary, both RBF-HDMR models and operational experience indicate that the Selkirk generation has a unique or non-linear impact on the power transfer capability on the Manitoba-Ontario interconnection. It is anticipated that Case 1 with no Selkirk generation and Case 3 with

both units on line would display more nonlinear characteristics while Case 2 with one unit on line shows fairly linear characteristics and is the most favorable operating condition.

Based on the observations of the three case studies from modeling standpoint, we can conclude that RBF-HDME models all three cases well. Among them, Case 2 is modeled the best, Case 3 is ranked second, and Case 1 is third. This conclusion is supported by the performance metrics in Tables 5.1-5.3. The mean absolute relative error (MARE) values are ranked by order of Case 2, Case 3 and Case 1. The relative average absolute error (RAAE) values are ranked in the same order as MARE, which indicates the consistency. All R Square values of three cases are close to 0, which is due to the fact that their outputs all approach to a constant. Both Cases 2 and 3 approach 300 MW and the error distribution of Case 3 is more scattered than that of Case 2. The mean of Case 1 approaches a constant of 264 MW and shows high nonlinearity. In Case 1 error distribution is much more scattered than those for both Case 2 and Case 3. These results also show that the proposed performance metric sets overcome the deficiency of a single metric to avoid potential misjudgment.

Compared with traditional metamodels, our experience is that RBF-HDMR model has obvious advantages in efficiency and accuracy for HEB problems. For example, a five variable similar power network is modeled by various polynomials [16]. The cost of the five variable cases is, on average, about two times higher than that for 50 variables.

5.7. Summary

This work applies the RBF-HDMR metamodeling technique to investigate the transfer capability of large-scale power grid on the Manitoba-Ontario Interconnection under various operating conditions. The result of the research work enhances the understanding of the performance of the Manitoba-Ontario interconnected power grid. Three cases are studied, and the results demonstrate that RBF-HDMR model can capture characteristics of the studied power grid and achieve a good prediction of transfer capability. The developed metamodels provide useful information in operational planning of the interconnected power grid.

In addition, the study finds the insufficiency of conventional performance metrics. Two sets of performance metrics are proposed for validation of metamodels.

5.8. Acknowledgement

5.9. Appendix Winnipeg River generation output ranges

Plant	Bus No	ID	InID	Low	High	Plant	Bus No	ID	InID	Low	High
		Generator		Generation Level				Generator		Generation Level	
McArthur Falls	67992	1	1	3.2	7.0	Pointe du Bois	68025	1	27	4.8	9.1
	67992	2	2	3.2	6.9		68025	2	28	1.96	3.72
	67993	3	3	3.2	7.0		68025	3	29	1.96	3.72
	67993	4	4	3.2	7.0		68025	4	30	1.96	3.72
	67994	5	5	3.2	7.25		68026	5	31	2.43	4.6
	67994	6	6	3.2	7.0		68026	6	32	2.43	4.6
	67995	7	7	3.2	7.14		68026	7	33	1.82	3.46
	67995	8	8	3.2	7.21		68027	8	34	2.44	4.63
Great Falls	68000	1	9	11.27	23.4		68027	9	35	2.17	4.12
	68001	2	10	11.27	20.9		68027	10	36	2.44	4.63
	68002	3	11	11.27	25.3		68028	11	37	2.57	4.88
	68003	4	12	11.27	20.2		68028	12	38	2.76	5.23
	68004	5	13	11.27	22.0		68029	13	39	2.77	5.26
	68005	6	14	11.27	23.4		68029	14	40	2.77	5.26
Seven Sisters	68013	1	15	13.78	29.13		68030	15	41	3.02	5.72
	68014	2	16	13.78	28.71		68030	16	42	3.02	5.72
	68015	3	17	13.78	27.46	Slave Falls	68041	1	43	4.51	8.54
	68016	4	18	13.78	28.5		68041	2	44	4.21	7.99
	68017	5	19	13.78	26.01		68042	3	45	4.49	8.51
	68018	6	20	13.78	25.59		68042	4	46	4.49	8.51
Pine Falls	68019	1	21	7.47	14.18		68060	5	47	4.51	8.54
	68020	2	22	7.47	14.18		68060	6	48	4.51	8.54
	68021	3	23	7.47	14.08		68061	7	49	4.48	8.5
	68022	4	24	7.47	13.78		68061	8	50	4.48	8.5
	68023	5	25	7.47	16.54	InID: input variable order for modeling					
	68024	6	26	7.47	16.84						

5.10. References

[1]   Myers, R. H. and Montgomery, D., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* John Wiley and Sons, Inc., Toronto.

[2]   Box, G. E. P., Hunter, J. S. and Hunter, W. G., 2005, *Statistics for Experimenters: Design, Innovation, and Discovery* John Wiley & Sons, Inc., Hoboken, New Jersey.

[3]   Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Statistical Science, 4(4), pp. 409-435.

[4]   Simpson, T. W., Mauery, T. M., Korte, J. J. and Mistree, F., 2001, "Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," AIAA Journal, 39(12), pp. 2233-2241.

[5]   Jin, R., Chen, W. and Simpson, T. W., 2001, "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria," Structural and Multidisciplinary Optimization, 23(1), pp. 1-13.

[6]   Chen, V., C. P., Tsui, K.-L., Barton, R. R. and Meckesheimer, M., 2006, "A Review on Design, Modeling and Applications of Computer Experiments," IIE Transactions, 38, pp. 273-291.

[7]   Clarke, S. M., Griebsch, J. H. and Simpson, T. W., 2005, "Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses," Transactions of ASME, Journal of Mechanical Design, 127(6), pp. 1077-1087.

[8] Giunta, A. A. and Watson, L. T., 1998, "A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models," *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, American Institute of Aeronautics and Astronautics, Inc., St. Louis, MO, Vol. 1, pp. 392-401.

[9] Viana, F. A. C., Haftka, R. T. and Steffen, V. J., 2009, "Multiple Surrogates: How Cross-Validation Errors Can Help Us to Obtain the Best Predictor," Structural and Multidisciplinary Optimization, 39(4), pp. 439-457.

[10] Acar, E. and Rais-Rohani, M., 2009, "Ensemble of Metamodels with Optimized Weight Factors," Structural and Multidisciplinary Optimization, 37, pp. 279-294.

[11] Goel, T., Haftka, R. T., Shyy, W. and Queipo, N. V., 2007, "Ensemble of Surrogates," Structural and Multidisciplinary Optimization, 33, pp. 199-216.

[12] Kuczera, R. C. and Mourelatos, Z. P., 2009, "On Estimating the Reliability of Multiple Failure Region Problems Using Approximate Metamodels," ASME Journal of Mechanical Design, 131(12).

[13] Yang, R. J., Wang, N., Tho, C. H. and Bobineau, J. P., 2005, "Metamodeling Development for Vehicle Frontal Impact Simulation," Transactions of the ASME, Journal of Mechanical Design, 127, pp. 1014-1020.

[14] Apley, D. W., Liu, J. and Chen, W., 2006, "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," Transactions of the ASME, Journal of Mechanical Design, 128, pp. 945-958.

[15] Georgopoulou, C. A. and Giannakoglou, K. C., 2009, "Metamodel-Assisted Evolutionary Algorithms for the Unit Commitment Problem with Probabilistic Outages," Applied Energy, 87(2010), pp. 1782-1792.

[16] Shan, S., Zhang, W., Myrna Cavers and Wang, G. G., 2010, "Application of Metamodeling Techniques to Power Transfer Capability Analysis of Manitoba-Ontario Electrical Interconnections," *Proceedings of The Canadian Society for Mechanical Engineering Forum 2010*, Victoria, British Columbia, Canada, June 7-9, 2010.

[17] Shan, S. and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-dimensional Design Problems with Computationally-Expensive Black-box Functions," Structural and Multidisciplinary Optimization, 41(2(2010)), pp. 219-241.

[18] Shan, S. and Wang, G. G., 2010, "Metamodeling for High-dimensional Simulation-Based Design Problems," Journal of Mechanical Design, 132(5), pp. 051009-1-051009-11.

[19] Shan, S. and Wang, G. G., 2010, "Turning Black-box into White Functions," Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Montreal, Quebec, Canada pp. August 15-18, 2010, Paper No. DETC201028958.

[20] Sobol', I. M., 1993, "Sensitivity Estimates for Nonlinear Mathematical Models," Mathematical Modeling & Computational Experiment, 1(4), pp. 407-414.

[21] Rabitz, H. and Al1s, Ö. F., 1999, "General Foundations of High-Dimensional Model Representations," Journal of Mathematical Chemistry, (25), pp. 197-233.

[22] Power technologies, I., 2000, "Iplan Program Manual."

[23] Lin, Y., Allen, J. K. and Mistree, F., 2002, "Metamodel Validation with Deterministic Computer Experiments," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 4-6 September 2002, Atlanta, Georgia, AIAA 2002-5425.

[24] Meckesheimer, M. and Booker, A. J., 2002, "Computationally Inexpensive Metamodel Assessment Strategies," AIAA Journal, 40(10), pp. 2053-2060.

[25] Kleijnen, J. P. C. and Sargent, R. G., 2000, "A Methodology for Fitting and Validating Metamodels in Simulation," European Journal of Operational Research, 120(2000), pp. 14-29.

[26] Gorissen, D., Couckuyt, I., Laermans, E. and Dhaene, T., 2010, "Multiobjective Global Surrogate Modeling, Dealing with the 5-Percent Problem," Engineering with Computers, 26, pp. 81-98.

[27] Kleijnen, J. P. C. and Deflandre, D., 2006, "Validation of Regression Metamodels in Simulation: Bootstrap Approach," European Journal of Operational Research, 170(2006), pp. 120-131.

[28] Santos, M. I. R. d. and Nova, A. c. M. O. P., 2006, "Statistical Fitting and Validation of Non-Linear Simulation Metamodels: A Case Study," European Journal of Operational Research, 171 (2006), pp. 53-63.

Chapter 6

Concluding Chapter

In this chapter, prominent points presented in the previous chapters are recapitulated with a discussion for future work.

6.1. Summary

Metamodeling techniques provide an effective approach to enhance the analysis of simulation results and expedite simulation-based optimization. This dissertation has surveyed the-state-of-the-art of metamodeling techniques and optimization methodologies for problems of high-dimensionality. The survey exposes that the existing metamodeling techniques are not only limited to low dimensionality and are also developed with little regard to the nature of the underlying systems. It is pointed out that the metamodeling techniques should emphasize the exploration of characteristics of the underlying systems and exploit these characteristics for metamodeling. We believe that sampling and metamodels should help to reveal the underlying problems rather than being "blind." The survey identifies the high-dimensional computationally-expensive black-box function (HEB) problems as a great hurdle for metamodeling and simulation-based optimizations. The "curse of dimensionality" is a bottleneck in the development, application, and integration of metamodeling techniques and optimization methodologies. Two promising approaches to solve HEB problems, mapping and decomposition, have been identified.

Aiming at solving HEB problems, this dissertation integrates Radial Basis Function (RBF) with High-Dimensional Model Representation (HDMR) into a new model, RBF-HDMR. The RBF-HDMR, compared with other metamodels, has a well-grounded theory. It is the first metamodel of its kind to tackle the HEB problem by taking advantage of the characteristics of underlying systems. The proposed RBF-HDMR and accompanying sampling adaptively explores and models an unknown underlying system and uses characteristics of an unknown underlying system such as linearity/nonlinearity and correlation relationships in variables. The RBR-HDMR has attractive features. The RBF-HDMR successfully circumvents the "curse of dimensionality". Multiple mathematical examples are used to test RBF-HDMR. Comparison with other sampling methods and metamodels demonstrates the effectiveness and efficiency of RBF-HDMR.

The dissertation extends from the initial development of RBF-HDMR to high-order component term, and also develops methods to "uncover" the black-box function. A structure matrix is proposed to support the explicit functional form of black-box function. A correlation matrix is proposed to explain the correlation relationship in the black-functions. The extended RBF-HDMR first "uncover" black-box functions. Two theorems are derived to support the metamodel constructing cost reduction if high-order correlation of variables exists and plays an important role in the black-box functions. Multiple examples show the turning-black-to-white capacity of RBF-HDMR.

The proposed RBF-HDMR has been applied to power transfer capability analysis of Manitoba-Ontario electrical interconnections. Three cases with 50 variables are modeled. The results further display the capability of constructing large-scale engineering problems. In addition, we found the deficiency of some performance metrics and propose

to assess a metamodel by the use of sets of performance metrics. The sets of performance metrics reflect the metamodel more thoroughly and avoid the impact brought by the deficiency of individual performance metrics.

This dissertation identifies high-dimensional computationally-expensive black-box function (HEB) problems as a practical-but-not-solved problem and "curse of dimensionality" as a bottleneck. A novel metamodel RBF-HDMR has been proposed to overcome this bottleneck. The proposed RBF-HDMR is successfully applied to engineering application cases. The dissertation represents the first work in tackling HEB problems in engineering design and opens doors to many possible research paths and applications.

6.2. Future Research

The RBF-HDMR is not only a high-dimensional metamodel, but also represents a general modeling methodology for HEB problems. Therefore, combining other types of metamodels (for example, support vector regression and Kriging) with the HDMR concepts are worthy to be explored.

The current RBF-HDMR does not address noises in simulation models. The noises exist in expensive analysis and simulation. Modeling noisy HEB problems have a practical significance in industry. RBF-HDMR needs to be further developed to tackle noisy problems. In addition, there are scenarios with existing simulation data. The function of RBF-HDMR using existing data is worthy to develop.

This research concentrates on metamodeling construction and sampling. Optimization for HEB problems is a natural extension of our current work. A decomposition-coordination optimization based on the proposed RBF-HDMR is to be developed.

Validation often receives little attention in the metamodeling community. It is an important part of metamodeling. Existing performance metrics have their own pros and cons. We have proposed two sets of performance metrics to assess metamodels. The efficiency of validation approaches is expected to be enhanced by statistical approaches. How to use performance metrics and statistics to effectively and efficiently validate metamodels, especially for HEB problems, will be a challenging and ongoing research topic.

Other Publications during Ph. D. Period

[1] **Shan, S**., Zhang, W., Myrna Cavers and Wang, G. G., 2010, "Application of Metamodeling Techniques to Power Transfer Capability Analysis of Manitoba-Ontario Electrical Interconnections," *Proceedings of The Canadian Society for Mechanical Engineering Forum 2010*,   Victoria, British Columbia, Canada, June 7-9, 2010.

[2] **Shan, S**. and Wang, G. G., 2009, "Reliable Space Pursuing for Reliability-Based Design Optimization with Black-box Performance Functions," Chinese Journal of Mechanical Engineering (English), **22**(1), pp. 27-35.

[3] Wang, H., **Shan, S**., Wang, G. G. and Li, G., 2009, "Mps-Based Ls-Svr Metamodeling Technique for Crashworthiness Optimization," Transactions of the ASME, Journal of Mechanical Design, **in revision**.

[4] Wang, G. G. and **Shan, S.**, 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," Transactions of the ASME, Journal of Mechanical Design, **129**, pp. 370-389.