Struct Multidisc Optim (2010) 41:219–241
DOI 10.1007/s00158-009-0420-2

RESEARCH PAPER

Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions

Songqing Shan · G. Gary Wang

Received: 7 November 2008 / Revised: 9 July 2009 / Accepted: 11 July 2009 / Published online: 7 August 2009

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 (i.e., black-box functions). The merger of these three challenges severely aggravates the difficulty and becomes a major hurdle for design optimization. This paper provides a survey on related modeling and optimization strategies that may help to solve *H*igh-dimensional, *E*xpensive (computationally), *B*lack-box (HEB) problems. The survey screens out 207 references including multiple historical reviews on relevant subjects from more than 1,000 papers in a variety of disciplines. This survey has been performed in three areas: strategies tackling high-dimensionality of problems, model approximation techniques, and 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

1 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

An earlier version of this work was published in *Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Sept. 10–12, 2008, Victoria, British Columbia, Canada.

S. Shan
Department of Mech. and Manuf. Eng., University of Manitoba, E2-327 EITC Building, 75A Chancellors Circle, Winnipeg, Manitoba, R3T 5V6, Canada

G. G. Wang (✉)
School of Engineering Science, Simon Fraser University, Surrey, British Columbia, V3T 0A3, Canada
e-mail: gary_wang@sfu.ca

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 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 min on an IBM RISC6000 7012 model 320 Planar workstation. If a two-factor 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 h for one evaluation (Gu 2001). Assuming ten variables with a two-factor full-factorial design, it needs 1,024 analyses and takes close to 12 years to complete.

The high dimensionality of input and output variables presents an exponential difficulty (i.e., 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 (e.g., 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, i.e., 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, 1997; Papalambros 1995; Papalambros and Michelena 1997, 2000). This kind of difficulty brought by the dimensionality of problems is known as the "curse-of-dimensionality." Mistree's research group referred to this difficulty as the "size of problem" 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 paper 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 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 Strategies tackling high-dimensionality

A spectrum of strategies 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 (e.g., parallel computing and increasing computer power), this section only reviews some of them that directly deal with high-dimensionality.

2.1 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 2009; Papalambros 1995; Papalambros and Michelena 1997, 2000). A technical map of decomposition methodology is provided in Fig. 1. The review is organized according to this map.

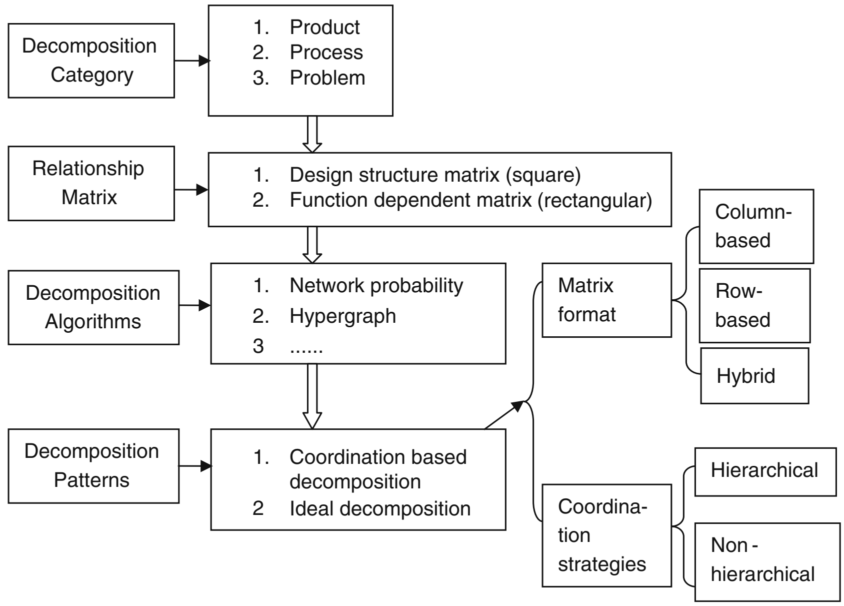

Fig. 1 An illustration of decomposition methodologies

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 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 the 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, 1997) advances the use of nonlinear optimization techniques in solving design problems. Such design optimization (e.g. 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 2009). 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, b), 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), 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 (i.e. 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. Decomposition often accompanies with parallel approaches to enhance the efficiency. Koch et al. (2000) proposed an approach to build partitioned, multilevel 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. (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, 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.

2.2 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. One category deals with a single response and the other deals with multiple responses.

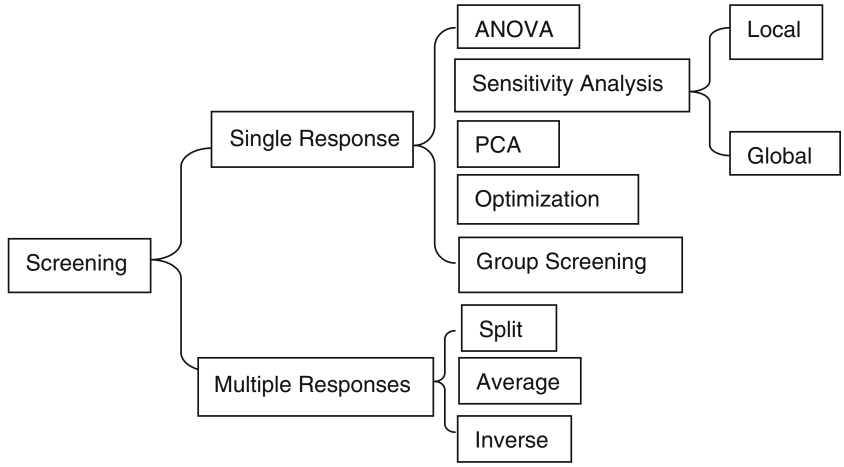

Fig. 2 Screening approaches

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 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 et al. (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 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 et al. (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 type of experiments. 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 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, b; 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) reviewed 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.

2.4 Space reduction

In modeling and optimizing a practical problem, ranges of design variables need to be determined. Combination of variable ranges defines the design space. In this paper, 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. (2001) developed a number of methods such as the adaptive response surface method (ARSM), 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, b) 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.

2.5 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, b) 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.

2.6 Summary remarks

Five main strategies tackling high-dimensionality are reviewed. Their pros and cons are summarized in Table 1. Among these methods, decomposition methodology is identified as the most promising tool for high dimensional problems, given its general 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.

3 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

Table 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; identification of 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

will then elucidate the relationship between modeling techniques and nature of underling functions to expose the oversight/flaws in current methods and indicate the direction for new model development.

3.1 Existing modeling techniques

Model approximation techniques involve two fields: computer design of 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 (i.e. experimental design); (2) to select a model to fit sampling points; (3) to choose a fitting method (e.g. 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 computer design of experiments (Sacks et al. 1989a, b; Steinberg and Hunter 1984) has been a few decades. The reviews on computer design of experiments can be found in the references (Chen et al. 2003, 2006; Crary 2002; John and Draper 1975; Steinberg and Hunter 1984). Multiple computer design of 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 et al. (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 (e.g. Jin et al. 2005; Morris and Mitchell 1995) combine the optimality approaches with the first category approaches (e.g. 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 merits. Those metrics of merits are summarized in Table 2.

Table 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 $N \bullet Var\left[\hat{f}(x)\right]/\sigma^2$ 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 3 Cost of some experimental designs

Experimental design	Condition (number of variables, $n = 30$)	Cost
Full factorial	Two level design	$2^{30} = 1.0737e9$
Fraction factorial	Half fraction	$\frac{1}{2} \times 2^{30} = 536,870,912$
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)

For computer design of 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 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.

In the modeling field, approximation models can be grouped into two categories: parametric models and non-parametric models as shown in Fig. 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.

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 (e.g. 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

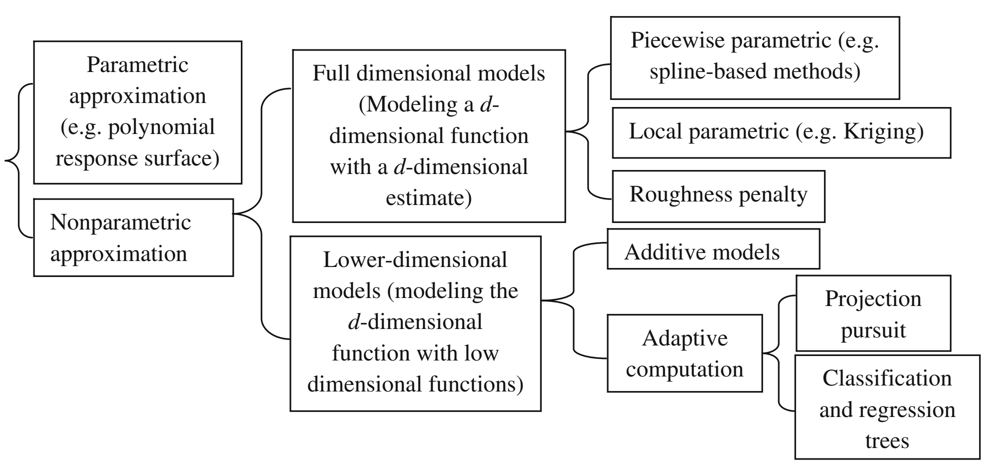

Fig. 3 Approximation models

Table 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 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

"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, 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, b), response surface models (Hill and Hunter 1966; Kaufman et al. 1996), support vector machine (Collobert and Bengio 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 (polynomial regression, multivariate adaptive regression splines, radial basis functions, and Kriging model), and Wang et al. (2006) compared meta-models (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 et al. (2006) described the pros and cons of some models. Meckesheimer 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 Tables 4 and 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.

3.2 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.

Table 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: $RAAE = \frac{\sum_{i=1}^{m}\lvert y_i - \hat{y}_i\rvert}{m \times STD}$	Usually correlated with MSE. A global error measurement
R Square: $R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}$	Usually correlated with MSE. A global error measurement
Predicted R-squares: $R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}$	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: $MAX = \max\lvert y_i - \hat{y}_i\rvert$, $i = 1, \cdots, m$	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; $\bar{y}$—the mean of the observed values; STD—standard deviation $STD = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2}{m}}$

A HDMR represents the mapping between the input variables $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ 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 follows:

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

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 always exact. The HDMR expands a d-dimensional function into summands 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 previous section. 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 highly 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:

$$f_0 = f(\mathbf{x}_0) \tag{2}$$

$$f_i(x_i) = f\left(x_i, \mathbf{x}_0^{\mathrm{i}}\right) - f_0 \tag{3}$$

$$f_{ij}(x_i, x_j) = f\left(x_i, x_j, \mathbf{x}_0^{\mathrm{ij}}\right) - f_i(x_i) - f_j(x_j) - f_0 \tag{4}$$

$$\begin{aligned} f_{ijk}(x_i, x_j, x_k) = f\left(x_i, x_j, x_k, \mathbf{x}_0^{\mathrm{ijk}}\right) - f_{ij}(x_i, x_j) \\ - f_{ik}(x_i, x_k) - f_{jk}(x_j, x_k) \\ - f_i(x_i) - f_j(x_j) - f_k(x_k) - f_0 \end{aligned} \tag{5}$$

...

$$\begin{aligned} f_{12\ldots n}(x_1, ..., x_n) = f(\mathbf{x}) - f_0 - \sum_i f_i(x_i) \\ - \sum_{ij} f_{ij}\left(x_i, x_j\right) - ... \end{aligned} \tag{6}$$

Where $\mathbf{x}_0^{\mathrm{i}}$, $\mathbf{x}_0^{\mathrm{ij}}$ and $\mathbf{x}_0^{\mathrm{ijk}}$ 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)

$$c = \sum_{i=0}^{l} \frac{n!}{(n-i)!i!} (s-1)^i \tag{7}$$

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 (1). If convergence of the Cut-HDMR expansion occurs at $\mathrm{L} \leq n$, 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 (7). This result is in strong contrast with the conventional view of exponential scaling of $\sim s^n$. It can be seen from (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

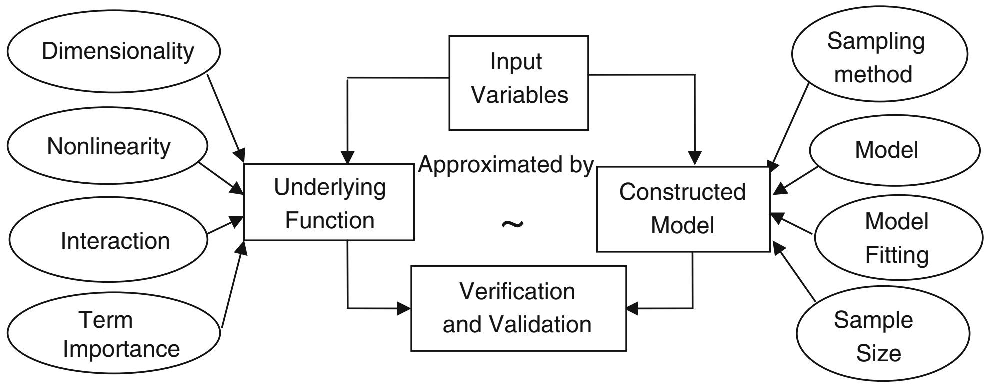

Fig. 4 Relationship among factors for approximation

component function as a sum of preconditioned low order Cut-HDMR component functions (Li et al. 2001b).

3.3 Relationship among factors for approximation

In the previous subsections, computer design of 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 (computer design of 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, i.e., dimensionality, nonlinearity, interactions among variables, and importance of terms, i.e., individual variables or a subset of interrelated variables. The relationship between features of the underlying problems and model approximation techniques is depicted in Fig. 4. In Fig. 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 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 (sampling method, model type, model fitting method and sample size), as well as the nature (the dimensionality, nonlinearity, interaction, and term importance) of the underlying functions. From this survey, it is observed that computer design of experiments and modeling techniques have been widely studied at the right side of Fig. 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. It is observed that there are few papers that studied the entire structure of the underlying function (the left side of Fig. 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. 4 (i.e. 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.

4 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.

4.1 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 (e.g. 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. 5: model approximation based techniques, and coarse-to-fine model based techniques.

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. 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.

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 coarse model or evaluated in important regions to improve the accuracy. There are several methods in this technique, as shown in Fig. 5, such as mapping, difference modeling, ratio modeling (Leary et al. 2003) and model fusion (Xiong et al. 2008). Mapping (Bakr et al. 1999a, b, 1998; Bandler et al. 1994, 1995a, b; 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. 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 ($r = \frac{f_f}{f_c}$ 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.

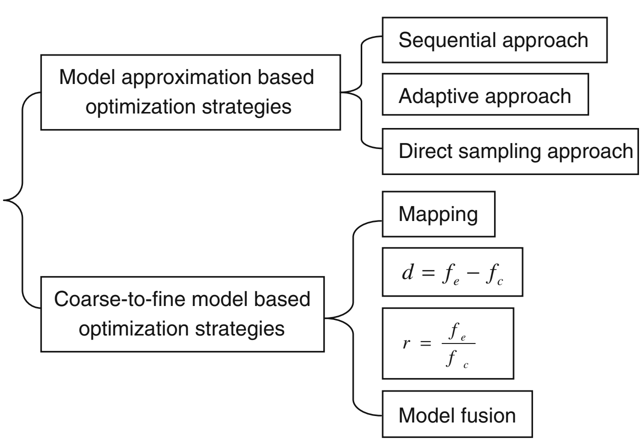

Fig. 5 Optimization strategies for computationally expensive problems

4.2 Non-gradient optimization algorithms

There are many well known optimization algorithms such as quasi-Newton methods (Arora et al. 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 (e.g., 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 paper 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 value. Björkman and Holmström (1999) implemented the DIRECT algorithm in Matlab™. 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 three or four variables in the electromagnetic field. Their approaches fall into the ones as shown in Fig. 6a, 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

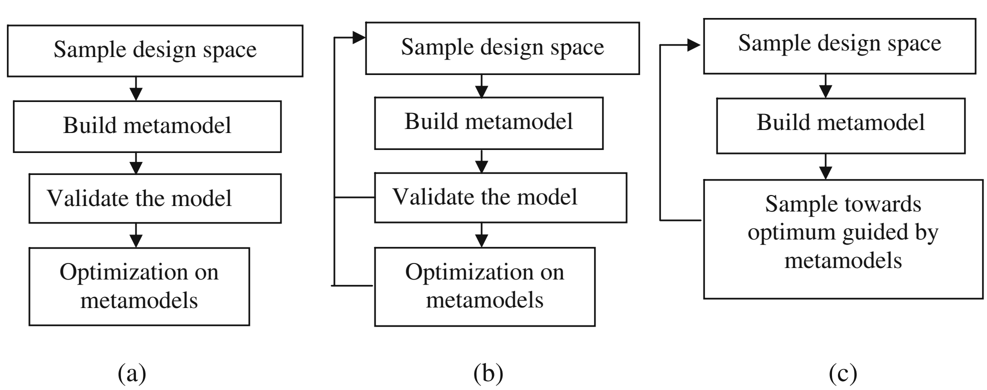

Fig. 6 MBDO strategies: **a** sequential approach, **b** adaptive MBDO, and **c** direct sampling approach (Wang and Shan 2007)

convergence. In Matlab$^{\text{TM}}$ 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 metaheuristics 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.

5 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 the above two challenges the high dimension 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 most promising 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, b). 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 (e.g. 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.

6 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.

Acknowledgments Funding supports from Canada Graduate Scholarships (CGS) and Natural Science and Engineering Research Council (NSERC) of Canada are gratefully acknowledged.

References

Ahn K-H, Chung WK (2002) Optimization with joint space reduction and extension induced by kinematic limits for redundant manipulators. In: Proceedings of the 2002 IEEE international conference on robotics & automation, Washington DC, 11–15 May

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 J Mech Des 118:486–489

An J, Owen A (2001) Quasi-regression. J Complex 17(4):588–607

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

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

Arora JS, Elwakeil OA, Chahande AI (1995) Global optimization methods for engineering applications: a review. Struct Optim 9:137–159

Audet C, Dennis JEJ (2004) A pattern search filter method for nonlinear programming without derivatives. SIAM J Optim 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 Trans Microwave Theor Tech 46(12):2412–2425

Bakr MH, Bandler JW, Georgieva N (1999a) An aggressive approach to parameter extraction. IEEE Trans Microwave Theor Tech 47(12):2428–2439

Bakr MH, Bandler JW, Georgieva N, Madsen K (1999b) A hybrid aggressive space-mapping algorithm for EM optimization. IEEE Trans Microwave Theor Tech 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 Trans Microwave Theor Tech 48(12):2297–2306

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

Bandler JW, Biernacki RM, Chen SH, Grobelny PA, Hemmers RH (1994) Space mapping technique for electromagnetic optimization. IEEE Trans Microwave Theor Tech 42(12):2536–2544

Bandler JW, Bienacki RM, Chen SH, Hemmers RH, Madsen K (1995a) Electromagnetic optimization exploiting aggressive space mapping. IEEE Trans Microwave Theor Tech 43(12):2874–2882

Bandler JW, Biernacki RM, Chen SH, Hemmers RH, Madsen K (1995b) Aggressive space mapping for electromagnetic design. In: IEEE MTT-S int. microwave symp. dig., Orlando, FL, 16–20 May

Bandler JW, Cheng QS, Dakroury SA, Mohamed AS, Bakr MH, Madsen K (2004) Space mapping: the state of the art. IEEE Trans Microwave Theor Tech 52(1):337–361

Banerjee I, Ierapetritou MG (2002) Design optimization under parameter uncertainty for general black-box models. Ind Eng Chem Res 41:6687–6697

Barry D (1986) Nonparametric Bayesian regression. Ann Stat 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. Springer, Berlin, pp 125–137

Bates RA, Buck RJ, Riccomagno E, Wynn HP (1996) Experimental design and observation for large systems. J R Stat Soc B 58(1): 77–94

Björkman M, Holmström K (1999) Global optimization using the DIRECT algorithm in Matlab. Adv Model Optim 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. Struct Optim 17(1):1–13

Bose RC, Bush KA (1952) Orthogonal arrays of strength two and three. Ann Math Stat 23(4):508–524

Box GEP (1957) Evolutionary operation: a method for increasing industrial productivity. Appl Stat 6:81–101

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

Browning TR (2001) Applying the design structure matrix to system decomposition and integration problems: a review and new directions. IEEE Trans Eng Manage 48(3):292–306

Byrd RH, Schnabel RB, Shults GA (1987) A trust region algorithm for nonlinearly constrained optimization. SIAM J Numer Anal 24(5):1152–1170

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

Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Stat Sci 10(3):273–304

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

Chan TF, Cong J, Kong T, Shinnerl JR (2000) Multilevel optimization for large-scale circuit placement. In: Proceedings of the 2000 IEEE/ACM international conference on computer-aided design, San Jose, California, 5–9 November

Chen Z (1991) Interaction spline models and their convergence rates. Ann Stat 19(4):1855–1868

Chen Z (1993) Fitting multivariate regression functions by interaction spline models. J R Stat Soc 55(2):473–491

Chen L, Li S (2005) Analysis of decomposability and complexity for design problems in the context of decomposition. ASME J Mech Des 127:545–557

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

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

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

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

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

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

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

Collobert R, Bengio S (2001) SVMTorch: support vector machines for large-scale regression problems. J Mach Learn Res 1:143–160

Crary SB (2002) Design of computer experiments for metamodel generation. Analog Integr Circuits Signal Process 32:7–16

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

Currin C, Mitchell T, Morris M, Ylvisaker D (1991) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. J Am Stat Assoc 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. In: 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 6 Nov 2008

Ding C, He X, Zha H, Simon HD (2002) Adaptive dimension reduction for clustering high dimensional data. In: The 2002 IEEE international conference on data mining (ICDM'02), Maebashi City, Japan, 9–12 December. IEEE, pp 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. Sage, London

Eldred MS, Hart WE, Schimel BD, Waanders BGVB (2000) Multilevel parallelism for optimization on MP computers: theory and experiment. In: Proceedings of the 8th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization, Long Beach, CA, September, AIAA-2000-4818

Eldred MS, Giunta AA, Waanders BGB (2004) Multilevel parallel optimization using massively parallel structural dynamics. Struct Multidisc Optim 27(1–2):97–109

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

Fadel GM, Riley MF, Barthelemy JM (1990) Two points exponential approximation method for structural optimization. Struct Multidisc Optim 2:117–124

Fang H, Horstemeyer MF (2006) Global response approximation with radial basis functions. J Eng Optim 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. Ann Stat 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. J Am Stat Assoc 76(372):817–823
Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Boston
Griensven AV et al (2006) A global sensitivity analysis tool for the parameters of multi-variable catchment models. J Hydrol 324: 10–23
Grignon P, Fadel GM (1994) Fuzzy move limit evaluation in structural optimization. In: The 5th AIAA/NASA/USAF/ISSMO fifth symposium on multidisciplinary analysis and optimization, Panama City, FL, 7–9 September, AIAA-94-4281
Gu L (2001) A comparison of polynomial based regression models in vehicle safety analysis. In: Proceedings of 2001 ASME design engineering technical conferences—design automation conference, Pittsburgh, PA, 9–12 September
Haftka RT (1991) Combining global and local approximations. AIAA J 29(9):1523–1525
Haftka RT, Scott EP, Cruz JR (1998) Optimization and experiments: a survey. Appl Mech Rev 51(7):435–448
Hamby DM (1994) A review of techniques for parameter sensitivity analysis of environmental models. Environ Monit Assess 32: 135–154
Harada T et al (2006) Screening parameters of pulmonary and cardiovascular integrated model with sensitivity analysis. In: Proceedings of the 28th IEEE EMBS annual international conference, New York City, USA, 30 Aug–3 Sept 2006
Hedayat AS, Sloane NJA, Stufken J (1999) Orthogonal arrays: theory and applications. Springer, New York
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
Hooker G (2004) Discovering additive structure in black box functions. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, Seattle, WA, USA, 22–25 August
Iman RL, Conover WJ (1980) Small sensitivity analysis techniques for computer models with an application to risk assessment. Commun. Stat, Theory and Methods A 9(17):1749–1842
Jiang T, Owen AB (2002) Quasi-regression for visualization and interpretation of black box functions. Stanford University, Stanford
Jiang T, Owen AB (2003) Quasi-regression with shrinkage. Math Comput Simul 62(3-6):231–241
Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodeling techniques under multiple modeling criteria. Struct Multidisc Optim 23(1):1–13
Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. In: The ASME 2002 design engineering technical conferences and computer and information in engineering conference, Montreal, Canada, 29 September–2 October
Jin R, Chen W, Sudjianto A (2004) Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design. In: SAE 2004 world congress, Detroit, MI, USA, 8–11 March, SAE 2004-01-0429
Jin R, Chen W, Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. J Stat Plan Inference 134(1):268–287
John RCS, Draper NR (1975) D-Optimality for regression designs: a review. Technometrics 17(1):15–23
Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. J Stat Plan Inference 26(2):131–148
Jones DR, Perttunen CD, Stuckman BE (1993) Lipschitzian optimization without the Lipschitz constant. J Optim Theory Appl 79(1):157–181
Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. J Glob Optim 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 8 August 2008
Kaski S (1998) Dimensionality reduction by random mapping: fast similarity computation for clustering. In: The neural networks proceedings, 1998. IEEE world congress on computational intelligence, Anchorage, AK, USA, 4–9 May
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. Comput Mech 18:112–126
Kaya H, Kaplan M, Saygin H (2004) A recursive algorithm for finding HDMR terms for sensitivity analysis. Comput Phys Commun 158:106–112
Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, Perth, WA, Australia, 27 Nov 1995, pp 1942–1948
Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc B 63(3):425–464
Kim HM, Michelena NF, Papalambros PY, Jiang T (2003) Target cascading in optimal system design. ASME J Mech Des 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. In: ASME advances in design automation
Koch PN, Simpson TW, Allen JK, Mistree F (1999) Statistical approximations for multidisciplinary design optimization: the problem of size. J Aircr 36(1):275–286
Koch PN, Mavris D, Mistree F (2000) Partitioned, multilevel response surfaces for modeling complex systems. AIAA J 38(5):875–881
Kodiyalam S, Sobieszczanski-Sobieski J (2000) Bilevel integrated system synthesis with response surfaces. AIAA J 38(8):1479–1485
Kokkolaras M, Mourelatos ZP, Papalambros PY (2006) Design optimization of hierarchically decomposed multilevel systems under uncertainty. ASME J Mech Des 128:503–508
Krishnamachari RS, Papalambros PY (1997a) Hierarchical decomposition synthesis in optimal systems design. ASME J Mech Des 119:448–457
Krishnamachari RS, Papalambros PY (1997b) Optimal hierarchical decomposition synthesis using integer programming. ASME J Mech Des 119:440–447
Kusiak A, Larson N (1995) Decomposition and representation methods in mechanical design. ASME J Mech Des 117(special 50th anniversary design issue):17–24
Kusiak A, Szczerbicki E (1992) A formal approach to specifications in conceptual design. ASME J Mech Des 114:659–666
Kusiak A, Wang J (1993) Decomposition of the design process. ASME J Mech Des 115:687–693
Lambert TJ III, Epelman MA, Smith RL (2005) A fictitious play approach to large-scale optimization. Oper Res 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. J Optim Eng 2:385–398
Leary SJ, Bhaskar A, Keane AJ (2003) A knowledge-based approach to response surface modeling in multifidelity optimization. J Glob Optim 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 Trans Compon Packag Technol 23(1):23–32

Li S (2009) Matrix-based decomposition algorithms for engineering application: survey and generic framework. Int J Prod Dev 9: 78–110
Li G, Rosenthal C, Rabitz H (2001a) High dimensional model representations. J Phys Chem 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. J Math Chem 30(1):1–30
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. J Phys Chem A 110:2474–2485
Lu SC-Y, Tcheng DK (1991) Building layered models to support engineering decision making: a machine learning approach. ASME J Mech Des 113:1–9
Marin FTS, Gonzalez AP (2003) Global optimization in path synthesis based on design space reduction. Mech Mach Theory 38: 579–594
Martin JD, Simpson TW (2005) Use of kriging models to approximate deterministic computer models. AIAA J 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 J 40(10):2053–2060
Michelena NF, Papalambros PY (1995a) A network reliability approach to optimal decomposition of design problems. ASME J Mech Des 117:433–440
Michelena NF, Papalambros PY (1995b) Optimal model-based decomposition of powertrain system design. ASME J Mech Des 117:499–505
Michelena NF, Papalambros PY (1997) A hypergraph framework for optimal model-based decomposition of design problems. Comput Optim Appl 8(2):173–196
Michelena N, Jiang T, Papalambros P (1995) Decomposition of simultaneous analysis and design models. In: Proceedings of the 1st world congress of structural and multidisciplinary optimization, pp 845–850
Michelena N, Papalambros P, Park HA, Kulkarni D (1999) Hierarchical overlapping coordination for large-scale optimization by decomposition. AIAA J 37(7):890–896
Mitchell TJ, Morris MD (1992) Bayesian design and analysis of computer experiments: two examples. Stat 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. J Stat Plan Inference 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. Wiley, Toronto
Nain PKS, Deb K (2002) A computationally effective multi-objective search and optimization technique using coarse-to-fine grain modeling (KanGal report no. 2002005). Indian Institute of Technology Kanpur, Kanpur
Oakley JE, O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc 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. Math Comput Simul 44:347–367
Owen AB (1992a) Orthogonal arrays for computer experiments, integration, and visualization. Stat Sinica 2:439–452
Owen AB (1992b) A central limit theorem for Latin hypercube sampling. J R Stat Soc 54(2):541–551
Owen AB (1998) Detecting near linearity in high dimensions. Stanford University, Stanford
Owen AB (2000) Assessing linearity in high dimensions. Ann Stat 28(1):1–19
Papalambros PY (1995) Optimal design of mechanical engineering systems. ASME J Mech Des 117(special 50th anniversary design issue):55–62
Papalambros PY, Michelena NF (1997) Model-based partitioning in optimal design of large engineering systems. In: Multidisciplinary design optimization: state-of-the-art. SIAM, pp 209–226
Papalambros PY, Michelena NF (2000) Trends and challenges in system design optimization. In: Proceedings of the international workshop on multidisciplinary design optimization, Pretoria, S. Africa, 7–10 August
Penha RML, Hines JW (2001) Using principal component analysis modeling to monitor temperature sensors in a nuclear research reactor. In: Proceedings of the maintenance and reliability conference (MARCON 2001), Knoxville, TN, 6–9 May
Pérez VM, Apker TB, Renaud JE (2002a) Parallel processing in sequential approximate optimization. In: The 43rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Denver, Colorado, 22–25 Apr, AIAA-2002-1589
Pérez VM, Renaud JE, Watson LT (2002b) Reduced sampling for construction of quadratic response surface approximations using adaptive experimental design. In: The 43rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Denver, Colorado, 22–25 Apr, AIAA-2002-1587
Queipo NV et al (2005) Surrogate-based analysis and optimization. Prog Aerospace Sci 41:1–18
Rabitz H, Alis ÖF (1999) General foundations of high-dimensional model representations. J Math Chem 25:197–233
Rabitz H, Alis ÖF, Shorter J, Shim K (1999) Efficient input–output model representations. Comput Phys Commun 117:11–20
Rao SS, Mulkay EL (2000) Engineering design optimization using interior-point algorithms. AIAA J 38(11):2127–2132
Rassokhin DN, Lobanov VS, Agratiotis DK (2000) Nonlinear mapping of massive data sets by fuzzy clustering and neural networks. J Comput Chem 22(4):373–386
Ratschek H, Rokne JG (1987) Efficiency of a global optimization algorithm. SIAM J Numer Anal 24(5):1191–1201
Regis RG, Shoemaker CA (2007a) Parallel radial basis function methods for the global optimization of expensive functions. Eur J Oper Res 182:514–535
Regis RG, Shoemaker CA (2007b) A stochastic radial basis function method for the global optimization of expensive functions. INFORMS J Comput 19(4):497–509
Renaud JE (1993) Second order based multidisciplinary design optimization algorithm development. Adv Des Autom 65-2:347–357
Renaud JE, Gabriele GA (1991) Sequential global approximation in non-hierarchic system decomposition and optimization. Adv Des Autom 32-1:191–200
Rodríguez JF, Renaud JE, Watson LT (1998) Trust region augmented Lagrangian methods for sequential response surface approximation and optimization. ASME J Mech Des 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. Stat Sci 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 Trans Comput 42(10):1222–1233

Sammon JW (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput 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 (eds) New development and applications in experimental design. Lecture notes-monograph series, vol 34. Institute of Mathematical Statistics, Hayward, pp 11–25

Shan S, Wang GG (2004) Space exploration and global optimization for computationally intensive design problems: a rough set based approach. Struct Multidisc Optim 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 J Mech Des 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 8 August 2008

Shin YS, Grandhi RV (2001) A global structural optimization technique using an interval method. Struct Multidisc Optim 22:351–363

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

Shorter JA, Ip PC, Rabitz HA (1999) An efficient chemical kinetics solver using high dimensional model representation. J Phys Chem 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 Trans Microwave Theor Tech 52(1):276–285

Simpson TW (2004) Evaluation of a graphical design interface for design space visualization. In: 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, Mauery TM, Korte JJ, Mistree F (1998) Comparison of response surface and kriging models for multidisciplinary design optimization. In: 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. Int J Reliab Appl 2(3):209–240

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

Simpson TW, Booker AJ, Ghosh D, Giunta AA, Koch PN, Yang RJ (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. Struct Multidisc Optim 27: 302–313

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

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

Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exper 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. J Biomed Inform 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. Struct Multidisc Optim 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. Ann Stat 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. In: The 9th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization, Atlanta, GA, 4–6 September, AIAA 2002-5622

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

Tang B (1993) Orthogonal array-based Latin hypercubes. J Am Stat Assoc 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. In: The international symposium on GIS, Istanbul-Turkey, 23–26 September

Tu J, Jones DR (2003) Variable screening in metamodel design by cross-validated moving least squares method. In: The 44th AIAA/ASME/ASCE/AHS structures, structural dynamics, and materials conference, Norfolk, Virginia, 7–10 April

Tunga MA, Demiralp M (2005) A factorized high dimensional model representation on the nodes of a finite hyperprismatic regular grid. Appl Math Comput 164:865–883

Tunga MA, Demiralp M (2006) Hybrid high dimensional model representation (HHDMR) on the partitioned data. J Comput Appl Math 185:107–132

Vanderplaats GN (1999) Structural design optimization status and direction. J Aircr 36(1):11–20

Wagner S (2007) Global sensitivity analysis of predictor models in software engineering. In: Proceedings of third international workshop on predictor models in software engineering (PROMISE'07), Washington, DC, USA. IEEE Computer Society

Wagner TC, Papalambros PY (1993) A general framework for decomposition analysis in optimal design. De-Vol. 65-2. Adv Des Autom 2:315–325

Wang H, Ersoy OK (2005) Parallel gray code optimization for high dimensional problems. In: Proceedings of the sixth international conference on computational intelligence and multimedia applications, Las Vegas, Nevada, 16–18 August

Wang GG, Shan S (2004) Design space reduction for multi-objective optimization and robust design optimization problems. SAE Trans 113:101–110

Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. ASME J Mech Des 129:370–389

Wang GG, Simpson TW (2004) Fuzzy clustering based hierarchical metamodeling for space reduction and design optimization. J Eng Optim 36(3):313–335

Wang GG, Dong Z, Aitchison P (2001) Adaptive response surface method—a global optimization scheme for computation-intensive design problems. J Eng Optim 33(6):707–734

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. J Phys Chem, A 107:4707–4716
Wang L, Shan S, Wang GG (2004) Mode-pursuing sampling method for global optimization on expensive black-box functions. J Eng Optim 36(4):419–438
Wang L, Beeson D et al (2006) A comparison of meta-modeling methods using practical industry requirements. In: The 47th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Newport, Rhode Island, USA, 1–4 May 2006
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 Trans Microwave Theor Tech 44(12):2495–2503
Weise T (2008) Global optimization algorithms theory and application. http://www.it-weise.de/projects/book.pdf. Accessed 7 Nov 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. Struct Multidisc Optim 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. Struct Multidisc Optim 23(6): 425–435
Wujek BA, Renaud JE (1998a) New adaptive move-limit management strategy for approximate optimization, Part 1. AIAA J 36(10):1911–1921
Wujek BA, Renaud JE (1998b) New adaptive move-limit management strategy for approximate optimization, Part 2. AIAA J 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 J Mech Des 130:111401. doi:10.1115/1.2976449
Ye KQ (1998) Orthogonal column Latin hypercubes and their application in computer experiments. J Am Stat Assoc 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 8 August 2008
Yoshimura M, Izui K (1998) Machine system design optimization strategies based on expansion and contraction of design spaces. In: 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 J Mech Des 126:217–224