Adaptive Response Surface Method – A Global Optimization

Scheme for Approximation-based Design Problems

G. Gary Wang*, Assistant Professor
Mechanical and Industrial Engineering
University of Manitoba
Winnipeg, MB, Canada, R3T 5V6
Tel: (204) 474-9463
Fax: (204) 275-7507
gary_wang@umanitoba.ca

Zuomin Dong, Professor
Mechanical Engineering
University of Victoria
Victoria, BC, Canada
Tel: (250) 721-8693
zdong@me.uvic.ca

Peter Aitchison, Professor
Applied Mathematics
University of Manitoba
Winnipeg, MB, Canada
Tel: (204) 474-9160
aitchis@cc.umanitoba.ca

Abstract

For design problems involving computation-intensive analysis or simulation processes, approximation models are usually introduced to reduce computation time. Most approximation-based optimization methods make step-by-step improvements to the approximation model by adjusting the limits of the design variables. In this work, a new approximation-based optimization method for computation-intensive design problems

* Corresponding Author
— the *adaptive response surface method* (ARSM), is presented. The ARSM creates quadratic approximation models for the computation-intensive design objective function in a gradually reduced design space. The ARSM was designed to avoid being trapped by local optimum and to identify the global design optimum with a modest number of objective function evaluations. Extensive tests on the ARSM as a global optimization scheme using benchmark problems, as well as an industrial design application of the method, are presented. Advantages and limitations of the approach are also discussed.

**Keywords**: Response Surface Method, Approximate Optimization, Global Optimization, Design Automation

**Introduction**

Today’s engineer design frequently involves complex and large systems. With the advances of Computer-Aided Design and Engineering (CAD/CAE) techniques, complex computer models and computation-intensive numerical analyses/simulations are often used to more accurately study the system behavior from many aspects and to guide design improvements. However, the high computational cost associated with these analyses and simulations prohibits them from being used as performance measurement tools in the optimization of a design. The design optimization process normally requires a large number of numerical iterations, each with one or more analysis calls, before the optimal solution is identified. The use of approximation models to replace the expensive computer analysis is a natural approach to avoid the computation barrier to the application of modern CAD/CAE tools in design optimization.
A number of approximation methods have been introduced in the past. Among them the *response surface method (RSM)* has attracted a growing interest in recent years. The RSM is one of the *design of experiments* (DOE) methods used to approximate an unknown function for which only a few values are computed [6, 29, 21, 28]. The RSM stems from science disciplines in which physical experiments are performed to study the unknown relation between a set of variables and the system output, or response, for which only a few experiment values are acquired. These relations are then modeled using a mathematical model, called *response surface*. Optimization based on the response surface is referred as experimental optimization. When experiments are inexpensive, experimental optimization can directly utilize methods developed for conventional analytical optimization. When experiments are expensive, the number of experiments required for the optimization must be minimized to reduce the total cost of the optimization. Haftka and his co-authors [17] pointed out that two factors distinguish experimental optimization from analytical optimization. The first is that experiments are cheaper to run in batches and the second is that it is difficult to obtain proper derivatives from experiments. More detailed comparison of these two types of optimization is in [17].

In engineering design, computation-intensive design analyses are treated as expensive computer “experiments” and thus experimental optimization can then be used for design optimization. Response surface models based on computer experiments are called *surrogates*. There are a large number of successful design examples using surrogates that
are archived in the literature [3, 42, 56, 38]. Applications of surrogates have also been widely studied in multidisciplinary design optimization. For instance, Renaud and Gabriele developed response surface approximations of multidisciplinary systems during concurrent subspace optimizations (CSSOs) [34, 35]; Chen et al. investigated the use of response surface approximations for robust concept exploration and concurrent design [8, 9]; and Korngold and Gabriele addressed discrete multidisciplinary problems using the RSM [24].

When employing surrogates in design optimization, the number of expensive computer experiments should be minimized to reduce computation cost, while ensuring improvements in the actual system through each step of the search. The surrogate related methods can be roughly categorized into two groups. The first group focuses on the better choice of an experiment planning scheme and a response surface model. The second group concentrates on the approximation model management in the optimization process to ensure a design optimum for the actual problem.

Experiment planning is an active research area. Standard DOE methods, such as factorial design and central composite are often used [8]. Other schemes include Taguchi’s method [52], D-optimal design [20] and Latin Hypercube designs [27]. Sacks et al. proposed a stochastic model to treat the deterministic computer response as a realization of a random function with respect to the actual system response [39, 40]. Neural networks have also been applied in generating the response surfaces for system approximation.
Representative researches include those reported in Swift and Batill [48, 49, 50, 51], Sellar, Batill and Renaud [44], Sellar and Batill [43], Sellar et al. [45], and Shi et al. [46].

The pioneer work on approximation model management in the optimization process was done by Box and Draper [5]. They developed a heuristic approach called evolutionary operation, which iteratively builds a response surface around the optimum from the previous iteration. Otto, Lanman and Patera introduced a qualitative surrogate validation in the experimental optimization [30]. Torczon and Trosset experimented with the use of merit functions to simultaneously improve both the quality of the approximation and the solution to the actual optimization problem [54]. A number of heuristic move-limit strategies have been developed for the approximate design optimization [15, 53, 74, 16, 33, 60, 61]. These methods vary the bounds of design variables in approximation iterations and differ from each other by different bound-adjustment strategies.

Related work in the mathematical community includes a proposed trust region method that guides the construction of the response surface using an accuracy index [10, 11, 12]. This idea was later incorporated into design applications by Dennis and Torczon [13], Lewis [25], Alexandrov et al. [1], and Rodríguez et al. [36]. The trust region method can work jointly with a response model and the global convergence can be achieved [10, 11]. The term, global convergence, means the convergence to a local minimum from any valid starting point. Convergence to the global minimum, which has the possible lowest/highest function value across the entire design space, is not guaranteed.
The new method presented in this paper falls into the second category of the surrogate related methods. The method was introduced to guide the approximation modeling in the optimization process to reach the global optimum with a limited number of computer experiments. The design problems that this method deals with are assumed having following characteristics:

- Evaluation of a particular design objective depends on the output of computer simulations involving large numbers of intermediate states or complicated computation processes;
- The evaluation of the design objective is computationally intensive and expensive.
- Evaluation results from computer analysis or simulation are deterministic, i.e., rerunning the simulation with the same inputs gives identical output. Thus, any difference between a model prediction and the result evaluated from the computer experiment at the same design point is because of the approximation inaccuracy of the response model.

**Problem Description**

The standard form of a non-linear optimization problem is

\[
\begin{align*}
\text{Minimize} & \quad f(\bar{x}) \\
\overline{x} & = [x_1, \cdots, x_n] \\
\text{subject to} & \quad h_j(\bar{x}) = 0, \quad (j = 1, \cdots, J)
\end{align*}
\]
\[ g_k(\bar{x}) \leq 0, \quad (k = 1, \cdots, K) \quad (3) \]
\[ x_{L,i} \leq x_i \leq x_{U,i} \quad (i = 1, \cdots, n) \quad (4) \]

The lower and upper bounds \( x_{L,i} \) and \( x_{U,i} \) in Equation 4 are initial variable bounds imposed by the designer.

In applications with the surrogate, approximations of both the objective and the constraints are often needed because of the associated computation intensity in evaluating these functions. In this paper for the ease of illustration, it is assumed that all constraint functions be in simple forms and only the objective function requires approximation due to its computation intensive evaluation process. For constraints requiring approximation, surrogates of the constraints can be constructed and incorporated similarly in the design optimization. The surrogate problem is formulated as

\[
\text{Minimize} \quad \tilde{f}(\bar{x})
\]

\[ \bar{x} = [x_1, \cdots, x_n] \]

\text{subject to} \quad h_j(\bar{x}) = 0, \quad (j = 1, \cdots, J) \quad (6)

\[ g_k(\bar{x}) \leq 0, \quad (k = 1, \cdots, K) \quad (7) \]
\[ x_{L,i} \leq x_i \leq x_{U,i} \quad (i = 1, \cdots, n) \quad (8) \]
where, \( \tilde{f}(\vec{x}) \) is the surrogate to the actual computation-intensive objective function. The proposed Adaptive Response Surface Method (ARSM) solves the original design problem (defined by Equations 1 to 4) through the construction and solution of a series of surrogate problems (specified by Equations 5 to 8).

**Adaptive Response Surface Method**

Response Surface Method

The standard RSM first employs an experimental strategy to generate design points in the design space, then applies either the first-order model or the second-order model to approximate the unknown system. The first and second order models have forms as given in Equations 9 and 10.

\[
y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{9}
\]

\[
y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i<j}^{n} \beta_{ij} x_i x_j \tag{10}
\]

where \( \beta_i \), \( \beta_{ii} \), and \( \beta_{ij} \) represent regression coefficients; \( x_i, (i = 1 \cdots n) \) are design variables and \( y \) is the response. The ARSM employs the second-order polynomial function (Equation 10) as the response surface model. It is the simplest nonlinear model in contrast to Sacks’ stochastic model and neural networks [39, 40, 44].

Many experimental strategies, including the full and fractional factorial designs, central composite designs (CCD) [28], Taguchi [52], and D-optimality design [20], can be used for
planning experimental designs. The CCD is probably by far the most widely used experimentation strategy for the second-order model fitting. Although CCD is not the most efficient planning strategy, it is adopted in the ARSM due to its successful use in other RSM applications. Even though the ARSM emphasizes the approximation model management during the optimization rather than the experimental design, a better planning method may overcome limitations of the ARSM imposed by the CCD when solving large-scale design problems. This will be further discussed in the later sections.

Conceptual Illustration of the ARSM

The concept of the ARSM can be illustrated through a single variable cubic function, \( f(x) = 2x^3 - 32x + 1 \), which is used to represent a computation-intensive design function. In the designated design space \([-3, 5]\), two local minima of the function exist, namely, (-3, 43) and (2.31, -48.3). Based upon CCD, five experimental points are obtained to approximate the “unknown” curve, as illustrated in Fig. 1. The experiment points are (-3, 43), (-1, 31), (1, -29), (3, -41), and (5, 91). If one applies a second-order model to carry out the approximation, the first fitted function using the least square method will be \( f(x) = -24.2 - 10.8x + 6x^2 \). The minimum of this fitted model is at \( x = 0.9 \) while the actual global optimum of the design function is at \( x = 2.31 \), as shown in Fig. 1. To improve the approximation accuracy, the ARSM introduces a threshold, i.e., a cutting plane, to the objective function. In this illustration, 20 is used as the threshold and the part of the design space corresponding to the function value above 20 is discarded. The new design space is now between the points (-1.96, 20) and (3.76, 20). The ARSM then applies
the CCD again over the reduced design space \([-1.96, 3.76]\), and produces a second fitted model \( f(x) = -10.06 - 22.95x + 5.40x^2 \), also shown in Fig. 1. This second fitted model yields a much better approximated optimum at \( x = 2.13 \), close to the true design optimum at \( x = 2.31 \). By continuing this process, the approach can quickly locate the global optimum of the design at the point \((2.31, -48.3)\).

Algorithm of the ARSM

As shown in Fig. 2, given design variables, objectives, constraints, and the initial design space, experimental designs (points in the design space) are generated according to a formal planning strategy, e.g. CCD. Values of the objective function are evaluated through the computation intensive analysis and simulation processes. The quadratic response surface model, or surrogate, is fitted to the data using the usual least square method. Since the constraints in Equations 6 and 7 may take any nonlinear forms, standard optimization methods, such as Lagrange Multipliers or line search, may fail. In the ARSM, a global optimization algorithm, simulated annealing, is used to find the global optimum. Following this step, the value of the actual objective function at the optimum of the surrogate is calculated through an evaluation of the computation-intensive objective function. If the value of the actual objective function at the surrogate optimum is better than the values at all other experimental designs, the point is added to the set of experimental designs for the following iteration because the point represents a promising search direction. All experimental designs and the accepted model optimum are recorded in a design library. After each design iteration, a threshold, or cutting plane, is used to reduce the design space, as described in the following sub-section. In this algorithm, the
second highest value of the objective function in the set of experimental designs is chosen as the threshold. If this second highest value cannot help to reduce the design space, the next highest value of the design function will be used, and so on. The termination criteria of the process are described in the sub-section: Termination Criteria.

Identification of the Reduced Design Space

For a single variable function, calculation of the reduced design space can be easily carried out by intersecting “the cutting plane” to the fitted function. Identification of the reduced design space for a multiple-variable design task is more challenging. Given the threshold of the objective function, the bounds of all design variables are identified through the construction of subsidiary optimization problems.

Assume the design problem has a total of \( n \) design variables, \( x_1, x_2, \cdots, x_n \), the ARSM model is as in Equation 10 and the threshold for the objective function is \( y_0 \), the range of design variable, \( x_k \) (\( k = 1, 2, \cdots, n \)), can be obtained by optimizing \( x_k \) with respect to all other design variables. Rearranging the response surface model given in Equation 10 gives a quadratic function in \( x_k \), with respect to other design variables in the coefficients as in Equation 11,

\[
\beta_{kk} x_k^2 + (\beta_k + \sum_{i=1,\neq k}^{n} \beta_{ik} x_i + \sum_{i>k}^{n} \beta_{ii} x_i) x_k + [\sum_{i=1,\neq k}^{n} (\beta_{ii} x_i + \beta_{ij} x_j) + \sum_{i<j,\neq k}^{n} \sum_{j=1,\neq k}^{n} \beta_{ij} x_j + \beta_0 - y_0] = 0
\]

(11)

which can be written as

\[
a x_k^2 + b x_k + c = 0,
\]
where,

\[ a = \beta_k \]

\[ b = \beta_k + \sum_{i=1,k}^{n} \beta_{ik}x_i + \sum_{i>k}^{n} \beta_{ii}x_i \]

\[ c = \sum_{i=1,k}^{n} (\beta_{ii}x_i^2 + \beta_i x_i) + \sum_{i<j,k}^{n} \sum_{j=1,k}^{n} \beta_{ij}x_ix_j + \beta_0 - y_0 \]  

(12)

The two solutions of \( x_k, x_{k,1} \) and \( x_{k,2} \), are given by

\[
\begin{align*}
    x_{k,1} & = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \\
    x_{k,2} & = \frac{-b \mp \sqrt{b^2 - 4ac}}{2a}
\end{align*}
\]

(a \neq 0)  

(13)

The reduced range of \( x_k \) is identified by finding the minimum and maximum of \( x_k \) with respect to all other design variables within their bounds, \( x_i \in [x_{l,i}, x_{u,i}] \). Two subsidiary optimization problems are formulated as:

For the lower limit of \( x_k \),

\[ \text{Minimize} \quad \min \{ x_{k,1}, x_{k,2} \} \]  

(14)

subject to

\[ x_{l,i} \leq x_i \leq x_{u,i} \quad (i = 1, \cdots, k-1, k+1, \cdots, n) \]  

(15)

where \( x_{l,i} \) and \( x_{u,i} \) are, respectively, the lower and upper limits of \( x_i \) from the previous model.

For the upper limit of \( x_k \),

\[ \text{Minimize} \quad -\max \{ x_{k,1}, x_{k,2} \} \]  

(16)
subject to

\[ x_{l,i} \leq x_i \leq x_{u,i} \quad (i = 1, \ldots, k - 1, k + 1, \ldots, n) \]  

These optimization problems involve non-linear objective functions that have no guaranteed unimodality. The simulated annealing global optimization method is thus applied for their solutions [22].

It is to be noted that the above procedure applies well to functions whose overall shape is convex, even though the function itself is non-convex, as shown in Fig. 3. An overall convex function \( f(x) \) is one that there is a convex function \( g(x) \) and a constant \( d \) such that \( |f(x) - g(x)| \leq d \) for most \( x \) in the domain and where \( d \) is substantially less than the total variation in the function \( f(x) \). For this case, the approximation function will generally be convex. The obtained design space, defined by the new bounds of each design variable, will thus be smaller than the previous design space. The reduced design space is then used to plan the experimental designs for the next iteration and a new surrogate will be constructed.

If the overall shape of the objective function is concave, the approximation function will be concave too and the minimum of the design problem will likely be at or close to the boundary of the design space. In this case, the above-described space identification procedure will not enable the ARSM to converge to a global optimum. To detect this situation, one needs to check at the first design iteration whether the obtained surrogate
optimum is in the new design space. If it is not, that implies that the overall function shape is concave.

Termination Criteria

In the ARSM, the search process terminates when the difference between the upper bound and lower bound for all variables becomes negligible. The termination criterion can be written as

$$|x_{i,d} - x_{u,d}| \leq \varepsilon, \ i = 1, \cdots, n$$

(18)

where, $\varepsilon$ is a small number specified by the designer. Also, the designer can terminate the process whenever a satisfactory design is found.

Tests on the ARSM

Even though there is not a standard set of test problems for global optimization algorithms, a number of well-known problems are widely used [14, 18, 31, 41, 55]. The proposed ARSM method was tested using these benchmark optimization problems. Among those, the Goldstein and Price function is a high-order polynomial function with an unusual shape, and is frequently used to test new optimization algorithms for their convergence characteristics [14]. This function rises up to higher than $10^9$ and has four local minima in a small region and has the following form:

$$f(x_1, x_2) = (1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2)) [30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1 x_2 + 27x_2^2)]$$

(19)
where, \( x_1, x_2 \in [-2, 2] \). We assume that the Goldstein and Price function acts as a complex function that requires intensive computation for each function evaluation.

The tests on the ARSM cover two aspects:

- How close does it come to the global optimum of the test function?
- How many function evaluations are needed for the algorithm to converge?

The comparison was made to the response surface method that only one surrogate was built, referred to here as "one-time RSM", in order to demonstrate the advantage of the ARSM in which surrogates are gradually improved. For the Goldstein and Price function, the one-time RSM results in a function value of 425.66, which is far from the design optimum of 3.0 (Table 1). On the other hand, the ARSM continues to improve the fitting accuracy and systematically approaches the global optimum. Calculation results after each design iteration are listed in Table 1. At the seventh iteration, the ARSM reached a solution very close to the analytic global optimum. The search could be terminated as a satisfactory result was obtained. Calculation results from additional iterations are listed as well to illustrate the potential improvement on the accuracy of the design optimum.

The search efficiency (in terms of required number of function evaluations) and the accuracy of the optimization are compared with another global optimization scheme, *simulated annealing* (SA), in Table 2.
Table 1. Optimization of the GoldStein and Price function using the ARSM

<table>
<thead>
<tr>
<th>Design Iteration</th>
<th>Design Space</th>
<th>Cutting Value</th>
<th>Optimum Design @</th>
<th>Design Func. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>X1: [-2, 2], X2: [-2, 2]</td>
<td>316600</td>
<td>(2.00, 0.53)</td>
<td>425.66</td>
</tr>
<tr>
<td>#2</td>
<td>X1: [-1.24, 2], X2: [-1.50, 2]</td>
<td>211478</td>
<td>(2.00, 0.53)</td>
<td>425.66</td>
</tr>
<tr>
<td>#3</td>
<td>X1: [-1.24, 2], X2: [-1.50, 1.76]</td>
<td>211478</td>
<td>(2.00, 0.53)</td>
<td>425.66</td>
</tr>
<tr>
<td>#4</td>
<td>X1: [-1.24, 2], X2: [-1.50, 1.39]</td>
<td>-4000.00</td>
<td>(2.00, 0.53)</td>
<td>425.66</td>
</tr>
<tr>
<td>#5</td>
<td>X1: [-1.24, 2], X2: [-1.50, -0.72]</td>
<td>41444.9</td>
<td>(0.012, 0.72)</td>
<td>63.200</td>
</tr>
<tr>
<td>#6</td>
<td>X1: [-1.24, 1.66], X2: [-1.50, -0.49]</td>
<td>8855.98</td>
<td>(0.21, -0.99)</td>
<td>26.880</td>
</tr>
<tr>
<td>#7</td>
<td>X1: [-1.24, 1.18], X2: [-1.50, -0.49]</td>
<td>5177.48</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#8</td>
<td>X1: [-1.24, 0.88], X2: [-1.50, -0.34]</td>
<td>2224.69</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#9</td>
<td>X1: [-1.24, 0.74], X2: [-1.21, -0.34]</td>
<td>287.420</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#10</td>
<td>X1: [-0.68, 0.51], X2: [-1.21, -0.34]</td>
<td>200.00</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#11</td>
<td>X1: [-0.64, 0.29], X2: [-1.02, -0.34]</td>
<td>80.36</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#12</td>
<td>X1: [-0.33, 0.29], X2: [-1.02, -0.53]</td>
<td>34.38</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#13</td>
<td>X1: [-0.28, 0.29], X2: [-1.02, -0.77]</td>
<td>28.09</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#14</td>
<td>X1: [-0.23, 0.20], X2: [-1.02, -0.77]</td>
<td>10.00</td>
<td>(-0.03, -0.99)</td>
<td>3.21</td>
</tr>
<tr>
<td>#15</td>
<td>X1: [-0.10, 0.14], X2: [-1.02, -0.87]</td>
<td>N/A</td>
<td>(-0.00, -0.99)</td>
<td>3.11</td>
</tr>
</tbody>
</table>

Table 2. A performance comparison between the SA and the ARSM

(on Goldstein and Price function)

<table>
<thead>
<tr>
<th>Global Optimization Method</th>
<th>Obtained Design Optimum</th>
<th>Optimal Design Function Value</th>
<th>Minimum Obtained at Iteration #</th>
<th>Number of Function Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARSM</td>
<td>(-0.01, -0.99)</td>
<td>3.11</td>
<td>15</td>
<td>150</td>
</tr>
<tr>
<td>SA</td>
<td>(0.00, -1.00)</td>
<td>3.00</td>
<td>11,276</td>
<td>11,276</td>
</tr>
</tbody>
</table>

If five minutes of computation time were needed to carry out one computation-intensive function evaluation, then the total computation time required by the SA would be forty days, in comparison to twelve hours required by the ARSM. The computation time associated with the design space reduction at each design iteration in the ARSM is in 1~2
seconds on a Sparc Ultra SUN workstation and thus is negligible. The ARSM optimization could be stopped after the seventh or eighth iteration with an obtained design minimum at 3.21, very close to the actual minimum and with far fewer computations. Thus the time needed to acquire the design optimum using the proposed ARSM is more suitable for a real design optimization task than the SA method. If distributed computation is used to perform experimental designs, the time can be further reduced.

Following the same procedure of minimizing the Goldstein and Price function, the ARSM was also tested with a number of other minimization problems listed below, where \( n \) is the total number of variables [14, 18, 31, 41, 55].

- **Six-hump camel-back function (SC),** \( n = 2 \).
  
  \[
  f_{sc}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4, \ x_{1,2} \in [-5, 5]
  \]  
  (20)

- **Branin function (BR),** \( n = 2 \).
  
  \[
  f_{b}(x) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos x_1 + 10, \ x_1 \in [-5, 10], \ x_2 \in [0, 15]
  \]  
  (21)

- **Generalized polynomial function (GF),** \( n = 2 \).
  
  \[
  f_{sb}(x) = u_1^2 + u_2^2 + u_3^2
  \]  
  (22)

  \[
  u_i = c_i - x_1(1 - x_2^i), \ i = 1, 2, 3
  \]  
  (23)

  \[
  c_1 = 1.5, \ c_2 = 2.25, \ c_3 = 2.625
  \]  
  (24)
• Rastrigin function (RS), \( n = 2 \).

\[
f_{RS}(x) = x_1^2 + x_2^2 - \cos 18x_1 - \cos 18x_2, \quad x_{1,2} \in [-1, 1]
\]  

(25)

• Griewank function (GN), \( n = 2 \).

\[
f_{GN}(x) = \sum_{i=1}^{n} x_i^2 / 200 - \prod_{i=1}^{n} \cos(x_i / \sqrt{i}) + 1, \quad x_{1,2} \in [-100, 100]
\]

(26)

• Hesse function (HE), \( n = 6 \)

\[-f_{HE}(x) = 25(x_1 - 2)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 + (x_4 - 4)^2 + (x_5 - 1)^2 + (x_6 - 4)^2\]

(27)

\[x_{1,2} \geq 0, \quad 1 \leq x_3 \leq 5, 0 \leq x_4 \leq 6, 1 \leq x_5 \leq 5, 0 \leq x_6 \leq 10, 2 \leq x_1 + x_2 \leq 6, x_1 - 3x_2 \leq 2, \]
\[4 \leq (x_3 - 3)^2 + x_4, \quad 4 \leq (x_5 - 3)^2 + x_6\]

(28)

• Hartman function (HN), \( n = 6 \).

\[
f_{HN}(x) = -\sum_{i=1}^{4} c_i \exp[-\sum_{j=1}^{n} \alpha_{ij} (x_j - p_{ij})^2], \quad x_i \in [0, 1], \quad i = 1, \cdots, n
\]

(29)

where

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \alpha_{ij}, \quad j = 1, \cdots, 6 )</th>
<th>( c_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 3 17 3.5 1.7 8</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>.05 10 17 0.1 8 14</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>3 3.5 1.7 10 17 8</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>17 8 .05 10 0.1 14</td>
<td>3.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( i )</th>
<th>( p_{ij}, \quad j = 1, \cdots, 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.1312 .1696 .5569 .0124 .8283 .5886</td>
</tr>
<tr>
<td>2</td>
<td>.2329 .4135 .8307 .3736 .1004 .9991</td>
</tr>
<tr>
<td>3</td>
<td>.2348 .1451 .3522 .2883 .3047 .6650</td>
</tr>
<tr>
<td>4</td>
<td>.4047 .8828 .8732 .5743 .1091 .0381</td>
</tr>
</tbody>
</table>
The test results were compared to those obtained through the *one-time RSM* and simulated annealing method, as listed in Table 3, where GP stands for the Goldstein and Price function. The third column of Table 3 lists the number of local minima of each test function within its given design space.

As shown in Table 3, the ARSM can converge fairly close to the global optimum (verified by the analytical solution) with a modest number of function evaluations. For functions RS, GN and HE, as the optimum is obtained at the first fitting, the advantage of the ARSM over the *one-time RSM* is not revealed. In other cases, the *one-time RSM* yields poor solutions for functions GP, SC, BR, GF and HN, as shown in Table 3. The ARSM did not find the global optimum for the function BR because the global optimum was cut-off during design space reduction. This problem is further discussed later in the Discussion section of this paper. For function HN, the ARSM identified its global optimum with a relatively large number of function evaluations. At present the total number of design function evaluations of the ARSM is determined by the CCD for experimental design planning. A drawback of the CCD is that the number of experimental designs (function evaluations) increases exponentially with the number of design variables. The ARSM can reasonably be used at this time for design optimization problems with less than ten design variables. Nevertheless the total number of design function evaluations required by the ARSM is significantly smaller than that of the *simulated annealing* method.
### Table 3. Summary of tests on the ARSM

<table>
<thead>
<tr>
<th>Test Function</th>
<th>Number of Variables</th>
<th>Number of Local Minima</th>
<th>Global Optimum Obtained</th>
<th>Number of Function Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Analytic Solution</td>
<td>ARSM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>One-time RSM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Simulated Annealing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARSM</td>
</tr>
<tr>
<td>GP</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3.21</td>
</tr>
<tr>
<td>SC</td>
<td>2</td>
<td>6</td>
<td>-1.032</td>
<td>-0.866</td>
</tr>
<tr>
<td>BR</td>
<td>2</td>
<td>3</td>
<td>0.398</td>
<td>2.099</td>
</tr>
<tr>
<td>GF</td>
<td>2</td>
<td>(\geq 5)</td>
<td>0</td>
<td>0.609</td>
</tr>
<tr>
<td>RS</td>
<td>2</td>
<td>50</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>GN</td>
<td>2</td>
<td>500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HE</td>
<td>6</td>
<td>18</td>
<td>-310</td>
<td>-304.5</td>
</tr>
<tr>
<td>HN</td>
<td>6</td>
<td>(\geq 3)</td>
<td>-3.32</td>
<td>-3.32</td>
</tr>
</tbody>
</table>

**Application Example: Optimal Design of a Fuel Cell Stack and System**

In all, three industrial design examples were studied. Initial testing of the ARSM was with the well-known two-bar structure design problem [8, 58]. Then the second application of the ARSM was to design a mechanical component of multiple functions [9]. For the purpose of brevity, these two design examples have been omitted from this article. The third application of the ARSM is the design optimization for a fuel cell stack and system with seven design variables. A fuel cell is an electric power-generating element based on the controlled reaction of fuel, typically hydrogen, and an oxidant, typically oxygen. The fuel cell system has significant potential to be used as a “clean” alternative to automotive engines.
Design Objectives and Constraints

The operation of a fuel cell system involves very complex mass and heat transfer processes. Design optimization has to be carried out at a stack and system level to optimize the performance and minimize the costs of fuel cells as an “engine”. The performance measures include the United States Advanced Battery Consortium Dynamic Stress Test (USABC DST) efficiency \([37, 62]\), net power output, volumetric power density, and gravitational power density. The system cost measures include the operational cost, material cost, manufacturing cost and assembly cost. The fuel cell, as an engine for vehicles, has to fit within the given space of a passenger car. According to \([47]\), the space is limited to \(0.5 \, m \times 1 \, m^2\). In addition, the structural integrity of a fuel cell stack is to be ensured. This work is to optimize the system design based on a patented fuel cell stack structure, the Tri-Stream External Manifolding and Radiator Stack (TERS). The system performance modeling involves a large number of non-linear equations and iterative simulation processes. The cost model is strongly interlaced with the performance model. Detailed fuel cell system performance and cost modeling is in \([32, 57]\). The performance measures and the system cost are evaluated through calling a system operation simulation process.

Design Variables

Fig. 4 illustrates a group of chosen design variables. These variables include:

1. **Operational variable**: air stoichiometry, \(\text{airSt}\);
2. **Configuration variable**: number of fuel cells per stack, \(nCell\);
3. **Geometric dimensions**: 
• Stack width (assumed square cross-section), $stackW$,
• Supporting column width (assumed square cross-section), $colW$,
• Height of the panel (fin), $finH$,
• Thickness of the stack end plate, $tEnd$, and
• Thickness of manifold covers, $tMani$.

Based upon a sensitivity study, the thickness of the endplate and the thickness of the manifolding behave independently of other design variables, and so are not included in the optimization.

Optimization Process

The optimization was carried out using five variables: $airSt$, $stackW$, $colW$, $nCell$, and $finH$, denoted as $x_1$, $x_2$, $x_3$, $x_4$, and $x_5$, respectively. A heuristics-based design given by Pastula [32] is used as the base case design for comparison. This base case is represented by a point in the design space: $\bar{x} = [2, 177.8, 19.05, 24, 9]$. The range of the five variables was set by the designer of the TERS fuel cell system as shown in Table 4.

<table>
<thead>
<tr>
<th>Design Variable</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Stoichiometry, $airSt$</td>
<td>1.3 ~ 2.5</td>
</tr>
<tr>
<td>Stack Width, $stackW$</td>
<td>100 ~ 240 mm</td>
</tr>
<tr>
<td>Column Width, $colW$</td>
<td>10 ~ 30 mm</td>
</tr>
<tr>
<td>Number of Cells per Stack, $nCell$</td>
<td>10 ~ 130 mm (Layout (b)) or 10 ~ 60 mm (Layout (a))</td>
</tr>
<tr>
<td>Fin Height, $finH$</td>
<td>4 ~ 15 mm</td>
</tr>
</tbody>
</table>

The number of cells per stack is a function of the stack layout. Two possible stack layouts are given in Fig. 5. Intuitively, layout (b) can reduce the total number of stacks and present a better solution. However, both layouts are included in the design optimization.
The system performances and cost of the fuel cell system were normalized and summed up into a single design objective function using the weighted-sum method with equal weights. The space constraint was incorporated in the optimization process while the structure integrity was verified at the obtained optimum after the optimization. The optimal designs were obtained for both layouts after two design iterations of each case using the ARSM. Detailed process and the experimental design data can be found in [57]. The results of optimization are given in Table 5.

Table 5. A comparison of the two system layouts

<table>
<thead>
<tr>
<th>Layout</th>
<th>Efficiency (%)</th>
<th>Net Power (kW)</th>
<th>Power Density (w/l)</th>
<th>Power Density (w/kg)</th>
<th>System Cost Reduction (%)</th>
<th>Length (m)</th>
<th>Space (m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>53.345</td>
<td>57.249</td>
<td>171.427</td>
<td>137.323</td>
<td>7.172</td>
<td>0.499</td>
<td>0.27</td>
</tr>
<tr>
<td>(b)</td>
<td>54.163</td>
<td>57.453</td>
<td>173.328</td>
<td>143.051</td>
<td>11.527</td>
<td>0.998</td>
<td>0.14</td>
</tr>
<tr>
<td>Base</td>
<td>46.3</td>
<td>51.0</td>
<td>121.2</td>
<td>100.1</td>
<td>0</td>
<td>0.33</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Significant performance improvement and cost reduction were achieved by both layouts over the base case design while satisfying the space constraint. Layout (b) presents a better design solution and the result confirms our design intuition.

For Layout (b), another type of optimization problem was formulated. This formulation was to use the cost as the objective function, subjected to the space constraint and the performance constraint that all the performance measures were required to be better than a base case design performances. This formulation represents the industrial need because the cost reduction is of major concern today while the fuel cell performances are satisfactory. The optimization results are given in Table 6. For the first formulation, the ARSM led to a balanced performance and cost design solution at \( x_{\text{min}} = [2 \ 100 \ 10 \ 114 \ 4] \) after two design iterations. This result is a little different from that in Table 5 because after the optimization for layout comparison, the air stoichiometry was intentionally fixed at a conservative number [57]. For the second formulation,
a minimum cost design at $x_{\text{min}}=[2 135 10 114 4]$ after nine iterations was reached by using the ARSM. Both designs are superior to the base case design. A considerable 43 percent increase in system power density or 16 percent reduction on system cost could be achieved. The entire design optimization was completed in hours.

Table 6. Design improvements over the base (reference) design

<table>
<thead>
<tr>
<th>Designs</th>
<th>Efficiency Increase (%)</th>
<th>Net Power Increase (%)</th>
<th>Volumetric Power Density Increase (%)</th>
<th>Gravimetric Power Density Increase (%)</th>
<th>System Cost Reduction (%)</th>
<th>Number of Design Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario I</td>
<td>12.02</td>
<td>6.20</td>
<td>43.49</td>
<td>43.49</td>
<td>12.07</td>
<td>2</td>
</tr>
<tr>
<td>Scenario II</td>
<td>6.09</td>
<td>3.83</td>
<td>40.59</td>
<td>40.75</td>
<td>16.12</td>
<td>9</td>
</tr>
</tbody>
</table>

Discussions

The proposed *adaptive response surface method* was introduced to carry out computation-intensive design optimizations. The ARSM applies an iterative process to progressively improve the quadratic approximation around the minimum of a complex or unknown design function. The space reduction ensures the convergence of the optimization process for overall convex shape functions as the objective function value is forced to reduce at each iteration.

The ARSM inherits the advantages from the traditional RSM, including the facilitation of distributed computation and providing insights into complex design problems. The RSM also enables design optimization to be independent of any CAD/CAE systems. Various modeling, analysis and simulation processes can be utilized to perform experimental designs and be integrated through the RSM with minimum coding effort. This feature allows broader applications and imposes fewer constraints on industrial users.
The ARSM also has its own distinct merits. It is an efficient global search scheme for approximation-based design optimization. It requires significantly fewer function evaluations for the convergence of design optimization, compared to the probabilistic global optimization algorithm, simulated annealing. In this sense, the ARSM increases the range of problem for which global optimization of computation-intensive design problems is feasible and practical. In addition, the ARSM facilitates the automation of the design process with proper interface between various function evaluation modules and the ARSM unit.

As any other optimization algorithms, the ARSM also has its limitations. In rare cases, the search algorithm may miss the true global optimum, as illustrated in Fig. 6. During the reduction of the design space, the true global optimum at point 2 will be ruled out and point 1 will be identified as the global optimum. Unless an experiment had been performed within the neighborhood of point 2, the real global optimum would be missed in the search. This is the case that the ARSM failed to find the global optimum of the Branin function in the testing. To avoid this case, a reasonably small initial design space should be chosen, if this is possible, and an appropriate experimental planning method with experimental points close enough to avoid missing the minimum. Secondly, as described in the section of Identification of the Reduced Design Space, the ARSM fails to converge to the global optimum when the overall shape of the objective function is concave. Finally, a better method of selecting threshold values needs to be developed to replace the current ad hoc method.

The ARSM focuses on the approximation model management during the optimization process. To solve practical industrial design problems, it can be combined with the sequential screening method [5, 2] to eliminate unimportant design variables and with a better experimental design planning technique to tackle large-scale design problems [23].
Conclusions

The Adaptive Response Surface Method (ARSM) introduced in this work focuses on the approximation model management for computation-intensive design optimization problems. The method is introduced for the optimization of an industrial design problem that has complex objective and constraint functions, and requires a computation intensive analysis/simulation process to evaluate these functions. The proposed ARSM algorithm maps the objective function value to a gradually reduced design space for functions of overall convex shape. As illustrated in benchmark tests and the industrial design application, the ARSM can converge to a global design optimum with a modest number of function evaluations. This new method makes the global optimization of approximation-based design problems feasible. Further work may improve its efficiency and ability to handle large-scale engineering design problems.

Acknowledgement

Financial support from the Natural Science and Engineering Research Council of Canada (NSERC), British Gas plc., and Ballard Power Systems, Inc., as well as technical assistance from Ballard and NGFT Research Laboratory staffs are gratefully acknowledged.

References


List of Figures

Fig. 1 A conceptual illustration of the ARSM
Fig. 2 The flowchart of the ARSM
Fig. 3 A non-convex function with an overall convex shape
Fig. 4 Chosen design variables
Fig. 5 Two possible spatial layouts of stacks
Fig. 6 The case in which ARSM may miss the global optimum
Fig. 1 A conceptual illustration of the ARSM
Fig. 2 The flowchart of the ARSM
Fig. 3 A non-convex function with an overall convex shape
Fig. 4 Chosen design variables
Fig. 5 Two possible spatial layouts of stacks
Fig. 6 The case in which ARSM may miss the global optimum