# Bayesian Meta-Modeling of Engineering Design Simulations: A Sequential Approach with Adaptation to Irregularities in the Response Behavior

*A. Farhang-Mehr*
Postdoctoral Research Associate

*S. Azarm*[1]
Professor

Department of Mechanical Engineering
University of Maryland
College Park, Maryland 20742 U.S.A.

## Abstract

Among the current meta-modeling approaches, Bayesian-based interpolation models have received significant attention in the literature. A Bayesian model is valid for an entire range of design space. Also, a Bayesian model has the ability to adapt to the behavior of a response function and thus obtain a more accurate meta-model with a fewer number of experiments. However, the current adaptive methods in the literature are mainly based on the assumption that some variables are more important (or sensitive) than others and accordingly less sensitive variables can be weighted less or ignored. This dramatically limits the scope and applicability of these models since in many practical cases none of the variables can be ignored or weighted less than others for the entire range of design space. A more pragmatic model is one that identifies regions of the design space where more experiments are needed.

In this paper, a new Bayesian meta-modeling approach is developed that designs and performs sets of experiments in a sequentially adaptive manner. In order to achieve the best possible meta-model, the approach adaptively utilizes the information obtained from previous experiments, builds interim meta-models, and identifies "irregular" regions of the design space in which more experiments are needed. The behavior of the interim meta-model is then quantified as a spatial function and incorporated into the next stage of the design to sequentially improve the accuracy of the obtained meta-model. The performance of the meta-modeling approach is demonstrated using numerical and engineering examples.

[1] Corresponding author- phone: (301)405-5250, fax: (301)314-9477, e-mail: azarm@umd.edu

## 1. Introduction

While the use of high-fidelity simulation models such as finite element or computational fluid dynamics models for optimization of engineering design problems has become very common, the computational expense of conducting numerous runs of such models remains for the most part an unresolved issue. One approach to reduce the computational expense is to approximate the simulation model by a less expensive meta-model (see, e.g., Roux et al., 1998, for a review of meta-modeling techniques in engineering). Among meta-modeling techniques, interpolative approaches such as Bayesian meta-modeling or kriging have been gaining significant attention (Sacks et al. 1989; Currin et al. 1991; Koehler and Owen 1996; and Martin et al. 2003).

A typical approximation technique consists of two phases: (1) *Design Of Experiments* (DOE), in which a sample of experiments in the design space is selected, and (2) *meta-modeling*, in which the response values from the DOE are evaluated and used to build a reasonably accurate approximation for a response function. The existing DOE techniques can be classified into two major groups: The first class of techniques, often referred to as *classical* DOE, account for inherent randomness in the behavior of the model and are mostly appropriate for physical experimentation with inherent measurement errors (e.g., Myers and Montgomery, 1995). The second class of techniques, often referred to as *space filling*, are especially appropriate for deterministic computer simulations. Examples of this second class of techniques include Random Latin Hypercube, Orthogonal Array, Integrated Mean-Squared Error, MaxiMin, MiniMax and Maximum Entropy (Booker et al. 1999). For a review of experimental design and analysis of deterministic computer experiments refer to Sacks et al. 1989, and Kohler and Owen 1996.

The meta-modeling approach in this paper is of space filling and global[2] type that makes it particularly suitable for global optimization of deterministic computer simulation models. It is also sequentially adaptive in order to achieve the best possible accuracy with fewer experiments, as discussed in the following section.

### 1.1 Meta-Modeling with Sequential Adaptation: Motivation and State of the Art

Sacks et al. (1989) classified sequential meta-modeling techniques into two major categories: sequential with and without adaptation. A sequentially adaptive approach designs a new sample of experiments based on all previous responses and the knowledge obtained from the resulting interim meta-model (e.g., Sacks et al. 1989; Sasena et al. 2002; Martin and Simpson 2002; and Turner et al. 2003). Several recent studies reveal that sequential adaptive techniques are in general much more promising than non-adaptive approaches (Jin et al. 2002; Pacheco et al. 2003; Turner et al. 2003). Bayesian approaches such as kriging (see Section 1.3) are particularly appropriate for adaptive meta-modeling because they can be easily updated with new information. Osio and Amon (1996) developed perhaps one of the first adaptive kriging approaches. Osio and Amon's approach as well as other similar approaches (e.g., Jin et al. 2002; Turner et al. 2003) are based on the assumption that some variables are more important than others, i.e., the response is more sensitive with respect to some variables. These variables are then weighted accordingly, or the less sensitive variables are suppressed. It can be argued (e.g., Jin et al. 2002; Kleijnen and Beer 2003; and Kleijnen et al. 2003) that such approaches cannot be effective unless there are ignorable variables in the problem. This dramatically limits the scope and applicability of these approaches since in most practical cases none of the variables can be ignored or weighted less than others.

[2] A global meta-model is valid in the entire range of design space (See Bartelemy and Haftka 1993, for classification and review of approximation techniques based on their ranges of validity). Examples include kriging and radial basis functions.

The adaptive approach of this paper, in contrast, is applicable in general[3]. In this approach, the behavior of the response function is assessed using a spatial function (i.e., function of the position vector, **x**, the input design space). This spatial function can then be used in subsequent stages of meta-modeling to automatically conduct more experiments in *irregular* regions of the design space where more information about the behavior of the response is needed (Farhang-Mehr and Azarm 2002).

The issue of irregularity in the response function behavior is described next.

### 1.2 Adaptation to Irregularities in the Response Function

The proposed approach of this paper adapts to the irregularities in the response function behavior (as defined quantitatively in Section 2.3) and sequentially designs more experiments accordingly. To further demonstrate this point, consider the example of Figure 1. The daggers on the x-axis represent a maximum-entropy single stage design of 15 experiments. (Maximum entropy design is fully described in Section 1.3. See also Koehler and Own, 1996, for a review of maximum-entropy design technique.) After evaluating the response values -- shown as bullets on the response curve -- the interpolating surrogate model is constructed (solid line).

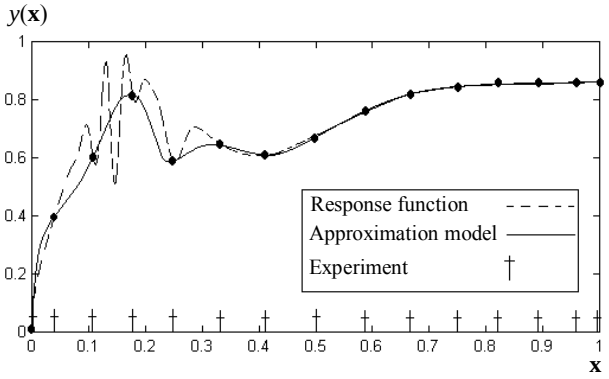



Figure 1: A design of 15 experiments for a deterministic response function and the resulting approximation model

[3] As discussed later in this section, this approach is even applicable to problems with only one design variable, as well as to problems with multiple variables none of which is insensitive (or ignorable). Note that neither of these classes of problems can be efficiently approximated using variable-weighting approaches.

As shown in Figure 1, the experiments in this single stage design are distributed symmetrically along the design interval with a slight emphasis towards the corners. A single-stage (non-adaptive) maximum entropy design is not problem specific and does not depend on the response function's behavior. Therefore, there are exactly 7 experiments in the interval [0,0.5) as well as the interval (0.5,1]. However, by looking at the actual response function, one notices that $y(x)$ is not behaving the same everywhere in the design space. The left half interval [0,0.5) consists of many local optima -- located tightly together (i.e., forms an irregular region) -- while a large portion of the domain in the right half of the interval, i.e., (0.5,1], is monotonic. Thus, to approximate the behavior of the response function in the irregular region of the domain more experiments must be conducted in the left half of the interval. In contrast, note that fewer experiments are needed in the less irregular region (i.e., the right half) of the domain to obtain a desired accuracy. A single-stage maximization of the entropy completely ignores this fact. Moreover, none of the previous adaptive meta-modeling techniques that are based on weighting or suppressing variables (e.g., Osio and Amon 1996; or Jin et al. 2002) are applicable to this problem because the variable x is sensitive in only a portion of its range, and behaves regularly in the rest of its domain[4]. The proposed adaptive meta-modeling approach of this paper addresses this issue.

Finally, note that the proposed approach of this paper aims at finding surrogate models that are globally valid throughout the design space. In other words, no particular region of the design space is assumed to be more attractive than others because of the value of the response. (For example, although a lower cost is usually very desirable, the proposed approach does not place more experiments in areas of the design space where cost is lower.) Therefore, the only

[4] Also, since there is only one variable in this example, suppressing or weighting approaches are not applicable.

criterion that is considered for adaptation is irregularity in the response function behavior and not the value of the response function itself. In some situations, however, additional factors other than irregularity should be considered. In minimization of a function, as an example, regions with lower response values are of much more interest. Several methods have been developed that employ a combination of approximation and optimization techniques to converge to the optimum of a function using meta-models (e.g., Jones et al. 1998; Sasena et al. 2000; and Wang et al. 2001). The main difference between such techniques and the approach of this paper is that we are interested in global meta-models that predict the response function everywhere in the design space and not just around areas of interest (e.g., close to an optimum). In multi-objective optimization, for instance, we are interested in meta-modeling of the entire range of each objective function and therefore, a global technique such as the proposed method of this paper can be of much interest[5].

The organization of the rest of this paper is as follows: In the next section, Section 1.3, a brief review of kriging and maximum entropy design terminology is given which will be used later in Section 2 to describe our proposed approach. Section 1.4 is a discussion of the computational complexities involved with kriging and maximum entropy designs. The proposed approach is presented in Section 2. Sections 3.1 and 3.2 present two demonstration examples: 1) an analytical function with interesting properties as shown in Figure 1; and 2) a crashworthiness model of front-end of a pickup truck. The concluding remarks are given in Section 4.

[5] Note that in multi-objective optimization, one cannot focus only on the optima of each individual objective function. In fact, points that are Pareto-optimal in a multi-objective sense often correspond to a compromise between objectives and not to optima of individual objective functions. In such cases, an approximation technique that focuses primarily on the optima of individual functions may miss many of Pareto-optimal solutions that do not correspond to the optimal of individual functions. This issue will be revisited in the second examples of this paper (Section 3.2.2).

### 1.3 Terminology of Bayesian Meta-Modeling and Maximum Entropy Design

The main idea behind the Bayesian approach is simple: the prior and posterior knowledge (i.e., before and after conducting experiments) about the response function can be modeled as random processes (e.g., Currin et al. 1991; Chaloner and Verdinelli 1995; Koehler and Own 1996; and Pacheco et al. 2001). Consider a deterministic response function, $y(\mathbf{x})$, with a vector of $p$ input variables, denoted as $\mathbf{x}$. Assume that each element of this vector is bounded between 0 and 1. So, $\mathbf{x} \in [0,1]^p$. A grid is constructed in this space. This essentially reduces the design space into a finite set of $p$-tuples in $[0,1]^p$, denoted by $U$. The unknown (or actual) response function, $y$, is fully explored if and only if $y(\mathbf{x})$ is evaluated for all $\mathbf{x} \in U$. Note that an exhaustive exploration of the design space is impossible for most of the real-world (computation intensive) simulation codes.

Design of Experiments: A design of $n$ experiments, denoted by $D_n$, is defined as a subset of $U$ at each point of which an experiment is run and the corresponding response function $y$ is evaluated (a total of $n$ experiments). The complement of $D$ in $U$ is also denoted as $\overline{D}$. The goal of the approximation is to predict the responses at $\overline{D}$ by conducting experiments in $D$.

Prior Process: Suppose that the prior knowledge for a deterministic function $y_i$ at a point $\mathbf{x}_i$ is represented by a normal process[6], $Y_i$. This prior distribution has the expected value and variance of $E(Y_i)=\mu_i$; and $Var(Y_i)=\sigma_{ii}$. However, $Y_i$ and $Y_j$ (i.e., prior distributions at $\mathbf{x}_i$ and $\mathbf{x}_j$) are not statistically independent and their covariance is: $Cov(Y_i, Y_j)=\sigma_{ij}$. Therefore, the prior knowledge about a given design, $D_n$, is represented by a multivariate normal process, $Y_D$. The observed $y_D$ is a realization of $Y_D$ with the expected value of $E[Y_D] = \boldsymbol{\mu}_D = [\mu_i]$.

[6] The normal assumption is made to simplify the computation of Bayesian posteriors. The same approach can be applied to non-normal priors (See Koehler and Own, 1996, for a general formulation).

Covariance Matrix: The elements of $Y_D$ are statistically dependent and the $n$x$n$ covariance matrix is denoted by $\mathbf{V}_{DD}$=Cov[$Y_D$, $Y_D$] = $[\sigma_{ij}]_{nxn}$ ; $\mathbf{x}_i, \mathbf{x}_j \in D_n$.

Posterior Process: After performing $n$ experiments in $D_n$ and observing $y_D$, the posterior distribution of $\overline{D}$ is represented by $Y_{\overline{D}|D}$, which is also multivariate normal. The posterior covariance matrix is denoted by: $\mathbf{V}_{\overline{D}\overline{D}|D} = \mathrm{Cov}[Y_{\overline{D}}, Y_{\overline{D}} \mid y_D] = [\sigma_{ij|D}]; \quad \mathbf{x}_i, \mathbf{x}_j \in \overline{D}_n$, where (Koehler and Owen 1996):

$$\mathbf{V}_{\overline{D}\overline{D}|D} = \mathbf{V}_{\overline{D}\overline{D}} - \mathbf{V}_{\overline{D}D}\mathbf{V}_{DD}^{-1}\mathbf{V}_{\overline{D}D}^{T} \tag{1}$$

Bayesian Meta-Modeling (Kriging): A Bayesian estimate for $y$ in $\overline{D}$ after observing $y_D$ is the mean of the posterior distribution (which minimizes a quadratic loss function, Casella and Berger 1990):

$$\hat{\mathbf{y}}_{\overline{D}} = \boldsymbol{\mu}_{\overline{D}|D} = \boldsymbol{\mu}_{\overline{D}} + \mathbf{V}_{\overline{D}D}\mathbf{V}_{DD}^{-1}(\mathbf{y}_D - \boldsymbol{\mu}_D) \tag{2}$$

where $\mathbf{y}_D$ is the vector of responses, i.e., $\mathbf{y}_D$=$[y_1, \ldots y_n]^T$. $\boldsymbol{\mu}_{\overline{D}|D}$ is a vector whose elements are the mean of the posterior processes of $\overline{D}$, and $\hat{\mathbf{y}}_{\overline{D}}$ estimates the response values for points in $\overline{D}$.

Stationary Assumption: In the absence of prior knowledge, it is assumed that the prior means and variances are the same everywhere in the design space (i.e., a "non-informative" assumption). Moreover, the covariance of two points is only a function of their distance and does not depend on their position in the design space. This is, in some sense, a non-informational and non-discriminatory assumption.

Correlation Function: The covariance matrix is stationary provided that $\sigma_{ij} = \sigma^2 R(\|\mathbf{x}_i - \mathbf{x}_j\|)$, where $R(.)$ is the correlation function which is monotonically decreasing with the distance between two points. Moreover, $R(.)$ satisfies $R(0)=1$, thus $\sigma_{ii}=\sigma^2$. The mean of the prior

processes are also assumed to be identical: $\mu_i=\mu;\ \forall \mathbf{x}_i \in U$. For the rest of this paper, we assume a Gaussian correlation function:

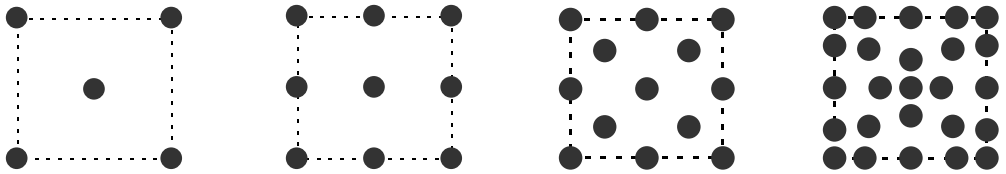
$$R(d) = e^{-\theta d^2} \tag{3}$$

where $d$ is the Euclidian distance between two points, and $\theta$ is a problem dependent constant which is not usually known a priori. Appropriate values for $\theta$, $\mu$ and $\sigma$ are set according to the designer's experience or knowledge of the response function smoothness, or by maximum likelihood estimators (see also, Mardia and Marshall 1984; Koehler and Own 1996).

Maximum Entropy Design: Shannon (1948) introduced the abstract notion of information entropy that has since found many applications in different fields. Lindley (1956) interpreted Shannon's entropy concept as the amount of information obtained by a Bayesian observation of a dependent parameter: an experiment that yields a maximum reduction in the expected entropy is the most informative experiment. Assuming normal priors, this criterion (also known as D-optimality criterion, see Roux et al. 1998) is equivalent to maximization of the determinant of the prior covariance matrix, i.e.,

$$\underset{D_n}{\text{Maximize}}\ det\ (\mathbf{V}_{DD}) \tag{4}$$

After estimating $\mu$ and $\sigma$, the maximum entropy design can be determined as follows: a subset of $U$ of size $n$ is selected such that the determinant of $\mathbf{R}_{DD}$ is maximized. According to Equation 4, this subset is the maximum entropy $D_n$ and has the highest expected value of information retrieval from the unknown response function. All experiments in $D_n$ are run, $y_D$ is observed, and an approximation model is built accordingly. Figure 2 depicts maximum entropy designs in a two-dimensional design space, with different number of experiments. Note that maximum entropy design slightly emphasizes the boundaries.



(a) 5 experiments (d) 9 experiments (c) 13 experiments (b) 25 experiments

Figure 2: Single-stage maximum entropy design of experiments (2-dimensional)

### 1.4 Computational Complexity of Non-Adaptive Entropy Optimal Design

In order to find the entropy optimal design (recall Equation 4), one can exhaustively calculate the determinant of all possible subsets of size $n$ in $U$ and compare these values -- an optimal $D_n$ is a subset that maximizes determinant of the covariance matrix. However, this brute force implementation may become computationally infeasible in the sense that the number of possible subsets of $U$ grows rapidly with the cardinality of $U$. Assume $U$ is a grid of $N$ nodes in the design space and a design of size $n$ is desired. The number of possible subsets of size $n$ in $U$ is therefore $\binom{N}{n}$. Thus, unless the grid is coarse or the dimension of the input vector is low, an exhaustive search becomes very time consuming. Note that there is no need for calculating the actual (perhaps computationally expensive) response function values during this process, nevertheless, $\binom{N}{n}$ grows so rapidly that renders computing even the simplest surrogate models computationally infeasible. As such, Currin et al. (1991) suggest an algorithm that successively augments new experiments to the existing design. In that algorithm, they took advantage of the Shewry and Wynn's result (1987) for one-point augmentation to an existing $n$-design, as discussed next.

**Shewry and Wynn's (1987) Augmentation:** If one desires to augment one more experiment to an existing set of experiments, the new experiment must be conducted at a point, namely $\mathbf{x}_i \in \overline{D}_n$,

with the largest variance of the posterior distribution. In other words, the best $\mathbf{x}_i$ to conduct a new experiment is the one at which $\sigma_{ii|D_n}$ is maximum.

In the algorithm suggested by Currin et al. (1991), experiments are augmented one-by-one to the current set according to the Shewry and Wynn's augmentation result. That is, a multiple-hiker search is conducted over $U$ to identify $\mathbf{x}_i \in \overline{D}_n$ with maximum $\sigma_{ii|D_n}$, as follows: Consider $n$ ascending 'hikers' that start from the current design, $D_n$, where the posterior variances are zero and move in the grid, one step at a time, in order to maximize the posterior variance. (This is very similar to a hill-climbing algorithm with multiple starting points.) Each of these hikers is allowed to move in $2p$ directions in the grid (2 directions along each edge of the grid). In each step, we evaluate the increase or decrease in the function for all possible $2p$ directions (at most $2pn$ evaluations). Each hiker then moves to the next node in the grid that has the highest variance. If two hikers meet at a node, they merge and continue as a single hiker. The algorithm continues until all hikers are at local maxima. Among them, the one with the maximum posterior variance is augmented to the current set of experiments. Although this algorithm does not guarantee obtaining the global maximum, nevertheless, it dramatically reduces the computational burden of finding the most informative (or close to the most informative) experiment.

In the following section, the proposed adaptive meta-modeling approach of this paper is presented.

## 2. Sequential Meta-Modeling with Adaptation to Irregularities

The new approach is based on the observation that while the actual response function is not known a priori, as more experiments are designed and performed, more information is gained

about the behavior of the response function. This information is then used to identify 'irregular' regions of the design space and to design the next set of experiments accordingly. Note that a good approximation model cannot be achieved without iteratively using this information towards the design of next set of experiments. Obviously, the stationary assumption is no longer applicable since the correlation of two points is not only a function of the relative distance but also dependent upon the characteristics of the regions in which the experiments are conducted and how well the function behaves in those regions. This information should be updated sequentially based on the data gathered from previous experiments, as explained in the following subsections.

**2.1 Assumptions**

It is assumed that the stationary assumption holds at the beginning of the process. A block of size $m$, denoted as $B_m$, is defined as a set of $m$ experiments whose elements (i.e., the individual experiments) are designed and performed in a single stage. In other words, all of the $m$ experiments in an $m$-block are designed by maximization of entropy, but once the block is designed, all of its experiments are conducted one after another -- without using the response value of one experiment to update the prior information of the next experiments. Accordingly, a single-stage maximum entropy design is basically a single block of experiments. As mentioned before, a single block of experiments may not be the best design since it does not take advantage of the information obtained during the meta-modeling process to place the next experiments in the regions where they are more informative. Basically, the geometry of a single block of experiments does not depend on the response function that is being investigated. In the proposed algorithm of this section, however, after each block is designed, the corresponding experiments

are run, the values of the response function are assessed and used to determine those regions of the design space in which new experiments are expected to be more informative.

As mentioned in Section 1.3, the covariance of prior distributions at two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is a decreasing function of their distance, that is

$$\mathrm{Cov}(Y_i, Y_j) = \sigma_{ij} = \sigma^2 R(\|\mathbf{x}_i - \mathbf{x}_j\|) \quad \text{and} \quad \mathrm{Var}(Y_i) = \sigma_{ii} = \sigma^2 \qquad ; \mathbf{x}_i, \mathbf{x}_j \in D \tag{5}$$

where $R(.)$ can be determined from Equation 3. This can be interpreted as follows: conducting an experiment at point $\mathbf{x}_i$ transmits some information about the response at point $\mathbf{x}_j$. For example, if the expected value of the prior distribution at $\mathbf{x}_i$ turns out to be an underestimation after observing $y_i$, i.e., $y_i > \mathrm{E}(Y_i)$, because of the positive correlation it is more likely that $\mathrm{E}(Y_j)$ is also an underestimation for the response function at an adjacent point $\mathbf{x}_j$ (i.e., based only on the information provided by the experiment conducted at $\mathbf{x}_i$). This can be thought of as the influence of conducting an experiment at $\mathbf{x}_i$ on its neighboring points. However, this influence decreases with the distance from $\mathbf{x}_i$ due to the decreasing correlation function. Points that are located in the vicinity of $\mathbf{x}_i$ are highly correlated with $Y_i$, and therefore observing $y_i$ has a significant impact on their posterior distributions. On the other hand, points that are located far from $\mathbf{x}_i$ are uncorrelated or loosely correlated and thus not influenced by $\mathbf{x}_i$. This is because $R(d)$ approaches zero as $d$ approaches infinity. Now consider two experiments $A$ and $B$ in Figure 3.

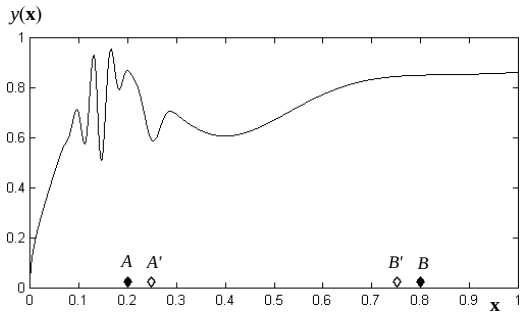



Figure 3: The uncertainty grows faster with distance in irregular regions

Both of these experiments influence their neighborhoods. If the response function is known a priori, one can say that experiment *A* is located in a neighborhood where *y*(x) is multi-modal (or irregular) with many close local optima. The influence of the experiment *A* on point *A′* is therefore relatively weak. In other words, conducting an experiment at *A* and observing $y_A$ does not say much about $y_{A'}$. On the other hand, experiment *B* conveys much more information about *B′*, since *B* is located in a less irregular region of the response function. So, the correlation decays slower with distance (or equivalently uncertainty grows slower) in the less irregular region. And, more experiments must be conducted in the multi-modal region to enable a more accurate modeling of the response function. In contrast, not that many experiments are needed in the less irregular region.

Clearly, irregularity should be quantified and formally incorporated into the design criterion, as discussed later in our approach. However, we first present a theorem that is used later as a basis for the proposed approach.

**2.2 Regional Adaptivity Theorem**

This theorem is proved as a way for incorporating spatial behavior functions (e.g., irregularity) into the maximum entropy design criterion. Suppose a set of *n* experiments has already been designed according to the entropy criterion. The theorem asserts that addition of a new experiment that has low correlation with previous experiments is in fact very informative (i.e., yields greater reduction in posterior entropy).

**Regional Adaptivity Theorem:** Suppose $D_n$ has already been designed and evaluated. Assuming Gaussian priors, conducting a new experiment at $\mathbf{x}_i \in \overline{D}_n$ is more informative than conducting another one at $\mathbf{x}_j \in \overline{D}_n$, if $\sigma_{ik} < \sigma_{jk}$; $\forall \mathbf{x}_k \in D_n$ and $\sigma_{ii}=\sigma_{jj}=\sigma^2$.

**Proof:** For Gaussian priors:

$$\sigma_{ii|D} = \sigma_{ii} - \mathbf{V}_{iD}^T \mathbf{V}_{DD}^{-1} \mathbf{V}_{iD} \quad \text{and} \quad \sigma_{jj|D} = \sigma_{jj} - \mathbf{V}_{jD}^T \mathbf{V}_{DD}^{-1} \mathbf{V}_{jD} \tag{6}$$

where $\mathbf{V}_{iD} = [\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{in}]^T$ is the covariance vector of $\mathbf{x}_i$ and $D_n$. Moreover, $\sigma_{ik} < \sigma_{jk}$; $\forall \mathbf{x}_k \in D_n$, and $\sigma_{ii} = \sigma_{jj} = \sigma^2$. Since $\sigma_{ij}$'s are all positive numbers, we have: $\sigma_{ii|D_n} > \sigma_{jj|D_n}$. Therefore, using Shewry and Wynn's augmentation result (Section 1.4), we conclude that conducting an experiment at $\mathbf{x}_i$ yields a greater reduction in posterior entropy than $\mathbf{x}_j$ and the theorem follows. □

An immediate and intuitive result from the above theorem is that addition of a new experiment that is located far from already existing experiments is very informative (i.e., due to small covariance between the new experiment and already existing ones). In fact, the maximum entropy criterion tends to maximize the distances among experiments and place new experiments in remote regions where the correlation with the existing experiments is minimum (and in some sense, the uncertainty is maximum). The most important aspect of this theorem, however, is that it provides a way to account for irregularities in the response function. According to this theorem, assigning a weaker correlation in the irregular regions of the design space increases the informational worth of an experiment conducted in those regions. A maximum entropy design is basically a design that maximizes the information content of a set of experiments and thus places more experiments in the regions where the correlation decays faster. This follows our previous observation that there must be more experiments conducted in the irregular portion of the domain in Figure 1, while not that many are needed in the less irregular part. In reality, however, we do not know the behavior of the actual response function upfront. Indeed, there is no information about the response function at the beginning of the process. This is the main motivation for introducing the new adaptive meta-modeling technique, as described next.

## 2.3 Approach

Suppose there exist enough time and computational resources to conduct a total of $n$ experiments. These experiments are designed and performed in several blocks, $k$ blocks of size $m$. The overall $n$-design then will be:

$$D_n = B_m^1 \cup ... \cup B_m^i \cup ... \cup B_m^k ; \qquad n = km \tag{7}$$

where $B_m^i$ is the $i$-th $m$-block, designed and performed after all previous ($i$-1) blocks are designed and performed.

Suppose that prior to the first $m$-block, i.e., $B_m^1$, no information is available about the response function and as reasoned before the stationary assumption is applicable for the prior distribution. After performing all experiments in $B_m^1$, an interim surrogate model is built based on these responses, $y_m^1$. The interim surrogate model provides a preliminary insight into the behavior of the response function.

Now that an initial understanding of the behavior of the unknown response function is obtained, one would like to take advantage of this knowledge in the design of the next set of experiments. One way to incorporate this knowledge is to use Regional Adaptivity Theorem (recall Section 2.2) and penalize the correlations in those regions of the design space where the function is irregular. In other words, our goal is to identify those **x**'s in $U$ for which the correlation decays faster with distance because of the presence of irregularity in the response function. The definition of function irregularity and how to update the covariance matrix accordingly is subjective in nature. Note that the correlation function itself is subjectively defined and the best choice is generally never known before solving the problem. In the following, a strategy is presented that is based on the multi-modality of the initial (or interim)

surrogate model. Note that, as demonstrated in the example of Figure 1, multi-modality in certain regions of the domain is a very important property that prompts for more experiments in those regions. As will become clear later in this section, if the function contains no local optima in $U$ (excluding the boundaries of the domain), the proposed sequential approach results in the same design as non-adaptive maximum entropy augmentation of experiments (where each block of experiments is designed at once, without revising the Bayesian priors in between blocks.)

The first step before designing a new set of experiments is to obtain all (or as many as possible) local optima of the interim surrogate model (which is constructed based on the previous blocks of experiments).[7] We refer to a node that is not located on the boundary of the grid and its approximated response function value is strictly higher or lower than all immediate neighboring nodes as an 'interior local optimum'. We also define a 'flat optimum set', denoted by $S \subset U$, as a connected set of two or more nodes with the same approximated response function value, strictly higher or lower than all other immediate neighbors that are not included in $S$. (The term 'connected set' refers to a set, each of its elements is an immediate neighbor of at least one other element in the set.) Clearly, the nodes of a flat optimum set are weakly optimum, because there exists at least one immediate neighbor in the flat optimum set with the same approximated response function value. In general, one node is selected randomly from a flat optimum set as a representative of that set. Finally, set $P$ is defined as the union of all interior local optima, including the representatives from all flat optimum sets.

In the following, we use $P$ to quantify the irregularity in the neighborhood of a point, namely $\mathbf{x}_i \in U$.

[7] Note that the approximation model is always considerably less expensive to compute compared to the original response function, and thus the optimization techniques that require many function calls are computationally feasible to apply. The implementation and computational complexity issues are discussed in Section 2.5. That section also includes an algorithm to find as many local optima as possible without considerable computational effort.

**Definition 1:** The Characteristic Certainty Width (CCW) at a point $\mathbf{x}_i \in U$, denoted by $L(\mathbf{x}_i)$ or $L_i$, is defined as the length of the diagonal of the smallest hyper-rectangle in the design space that encloses $\mathbf{x}_i$ and whose two opposite vertices are any two local optima in $P$. Moreover, we define $L_0$ as the length of the diagonal of the design space, which is a hyper-rectangle itself. It is then assumed that: $L(\mathbf{x}_i) = L_0$, if no such rectangle can be found that encloses $\mathbf{x}_i$.

The intuitive interpretation of this definition is shown in Figure 4. The set of optima, $P$, is marked by solid bullets in a 2-dimensional design space. Points A to E represent 5 candidate points in $U$ for conducting the next experiment. From Definition 1, we have $L(\mathbf{A})=L(\mathbf{E})=L_0$; $L(\mathbf{B})=L_1$; $L(\mathbf{C})=L_2$; $L(\mathbf{D})=L_3$. For a one-dimensional design space, CCW of a point is the distance of the two optima that bracket that point.

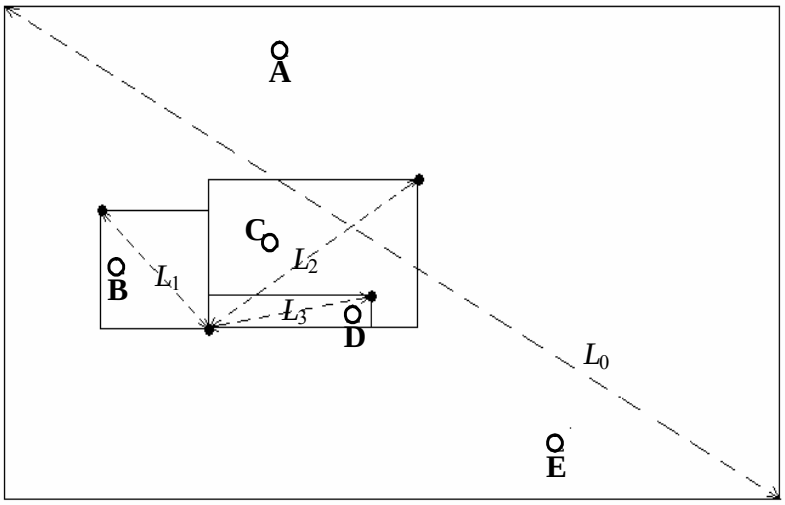



Figure 4: Characteristic certainty width

We use $L$ as a measure of regularity in the behavior of the function. A large $L$ implies that there is a wide unimodal region. As mentioned earlier, fewer experiments are needed for such a region in the design space, as opposed to an irregular region with many tightly located local optima. In other words, the correlation tapers off slower with distance in a region with a larger $L$. This corresponds to a larger covariance. Therefore, $(L_i/L_0)$ is used as a correcting factor that

incorporates irregularities of the design space and updates the covariance of two points accordingly, as follows: (compare to Equation 5)

$$\sigma'_{ij} = \sigma_{ij}(\frac{L_i}{L_0})(\frac{L_j}{L_0}) = \left[\sigma(\frac{L_i}{L_0})\right]\left[\sigma(\frac{L_j}{L_0})\right] R\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|\right); \quad i \neq j \tag{8}$$

where $R(.)$ is obtained from Equation 3. Intuitively, the correlation of two points is corrected by a factor that is proportional to irregularity at those two points: $[L(\mathbf{x}_i)/L_0][L(\mathbf{x}_j)/L_0]$. If this factor is large, uncertainty grows faster with distance (i.e., covariance of two points tapers off faster). Hence, according to the Regional Adaptivity Theorem, updating the covariance matrix using the above equations and maximization of entropy automatically places more experiments in the irregular regions as the next block of experiments, because of their higher informational worth (greater entropy reduction). In contrast, in less irregular regions where the local optima are located far from each other, the covariance decays slowly with distance (or uncertainty grows slowly). The quantity $L/L_0$ is larger in those regions and thus covariances are not considerably reduced, if reduced at all (recall $L_i=L_0$ if $\mathbf{x}_i$ is not bracketed between two optima in $P$). Note that if we find none or only one local optimum inside the design space we have $L=L_0$, and therefore this strategy basically results in the same design as obtained from non-adaptive maximum entropy approach (where each block of experiments is designed at once, without revising the Bayesian priors in between blocks.) Indeed, a uni-modal response function with one or no interior local optimum is considered regular everywhere in the design space and therefore the covariances are not reduced at all; i.e., $(L/L_0)=1$.

### 2.4 Step-by-Step Approach

A step-by-step description of the proposed approach is given below.

***Step 1*** – Since there is no initial information about the response function, the first block of experiments, $B_m^1$, is designed with the stationary assumption and according to entropy criterion of Equation 4. The initial *m*x*m* covariance matrix is constructed using Equation 5. The standard deviation, $\sigma$, is estimated via the maximum likelihood estimators.

***Step 2*** – In the *k*-th iteration we have: $D_{k\times m} = B_m^1 \cup ... \cup B_m^k$. All new experiments are observed, i.e., obtain $y_{k\times m}$.

***Step 3*** – An interim approximation model is constructed based on the responses $y_{k\times m}$ from all previous experiments (using kriging, recall Equation 2).

***Step 4*** – The interim approximation model is optimized and all (or as many as possible) local optima are obtained to form set *P*.

***Step 5*** – The characteristic certainty width, $L(\mathbf{x}_i)$, is calculated for all $\mathbf{x}_i \in \overline{D}_{k\times m}$ (i.e., complement of $D_{k\times m}$.)

***Step 6*** – A *q*x*q* covariance matrix is constructed where $q=m(k+1)$. The first *mk*x*mk* rows and columns of the matrix correspond to the experiments that were already designed and performed in the previous iterations, i.e., $D_{k\times m}$. These entries are updated according to Equation 8 using the new CCW's. There are *m* rows and *m* columns remaining in this matrix that have to be determined, as discussed next.

***Step 7*** – The remaining *m*x*m* rows and columns of the matrix correspond to the next *m*-block, $B_m^{k+1}$. The elements of this block are selected from $\overline{D}_{k\times m}$ such that the determinant of the *q*x*q* matrix is maximized. Again, these new entries follow Equation 8. The result of this maximization is a set of *m* new points that marks *m* new experiments, i.e., $B_m^{k+1}$.

***Step 8*** – Go to step 2 and continue until a total of $n = k \times m$ experiments are designed and performed. The final approximation model is constructed based on these *n* experiments. Should one decides to conduct more experiments (i.e., by assigning more computational resources), this algorithm can be run again to design new experiments, starting from the existing approximation model.

### 2.5 Computational Complexity; and Revised Multiple Hiker Algorithm

There are basically two main sources of computational burden in the proposed approach (i.e., other than the evaluations of the potentially expensive response function):

(i) Step 1 and Step 7, in which a block of experiments are chosen such that it maximizes the determinant of the covariance matrix. In the test example of the next section, this is done exhaustively by searching the set of all subsets of size *m* in *U*. However, exhaustive search is not possible for large *U*'s. Therefore, the multiple-hiker algorithm of Section 1.4 can be used with some modifications to determine the optimum block of experiments in Steps 1 and 7.

(ii) Step 4, in which the local optima of the approximation model is obtained. The straightforward approach is to exhaustively search the grid of *U* for those nodes that are local minima (and/or maxima). Suppose *U* is *p*-dimensional and contains *N* points, thus each point in *U* has at most 2*p* neighbors (points on the boundary have fewer neighbors). To find all local optima, one should examine all points in *U* and compare the values of the approximation model. Although an approximation model is not costly to compute, for high-dimensional design spaces with fine grids, an exhaustive search may become very time-consuming or even computationally prohibitive. Therefore, to reduce the computational effort, we revise the multiple-hiker algorithm to include *n* descending-hikers (i.e., a total of 2*n* hikers). These descending hikers will locate local minima (local maxima are located using ascending hikers.) Note that although this

algorithm significantly reduces the computational cost of the process, it may not necessarily obtain all optima. As mentioned before, a full set of optima is not guaranteed in general without an exhaustive search of *U*. However, obtaining a reasonable number of local optima (as computational resources permit) should provide a good understanding of irregularities in the response[8]. As more and more local optima are detected, the irregular regions are better identified, which in turn increases the accuracy of the approximation. Once all optima are obtained, CCW can be easily calculated at each point, say **x**, in the design space by finding a pair of optima in *P* that encloses **x** as closely as possible (i.e. finding the rectangle with the smallest diameter that contains **x**).

In the next section, the performance of our approach in section 2.4 is demonstrated using two test examples.

## 3. Examples

Two test examples are presented in this section. The first test example is a mathematical function (demonstrated graphically in Figure 1). The second test example involves a computationally-expensive crashworthiness simulation model.

### 3.1 Numerical Example

The formulation of the response function of Figure 1 is given below:

$$\begin{aligned} y(\mathrm{x}) &= (1-e^{-2\sqrt{\mathrm{x}}}) + 6\mathrm{x}e^{-7\mathrm{x}}\sin(10\mathrm{x}) - 0.2\mathrm{e}^{-2000(\mathrm{x}-0.25)^2} \\ &+ 60\min(0,|\mathrm{x}-0.14|-0.08)^2[\ln(\mathrm{x}+0.2)+1.5\sin^2(85\mathrm{x})] \end{aligned} \tag{9}$$

Here, we assume that there is just sufficient computational power (or time) to perform 15 response function evaluations. A single-stage maximum entropy design and kriging approach for this problem was discussed in Section 1.2 and demonstrated in Figure 1. In the following

[8] If fewer than two local optima are located, the application of the new approach results in the same design as that of a single-stage maximum entropy approach.

subsection, to demonstrate the application of the proposed approach, we sequentially design a total of 15 experiments as 3 blocks of size 5 experiments and compare the results with the single-stage maximum entropy approach in Section 3.1.2.

### 3.1.1 Application of the Proposed Approach

A step-by-step application of the proposed sequential approach to the mathematical function of Equation 9 is described next.

***Step 1 (1st iteration)*** – Similar to the single-stage maximum entropy design, the first block of 5 experiments is chosen exhaustively from all possible subsets of size 5 in $U$ such that the determinant of the covariance matrix is maximum. The small arrows on the x-axis in Figure 5 illustrate $B_5^1$ for this response function. Note that one could alternatively use the multiple-hiker algorithm for one-point augmentation of experiments, and obtain an optimum 5-design. This latter approach results in a slightly different, yet good enough design while it significantly reduces the computational burden.

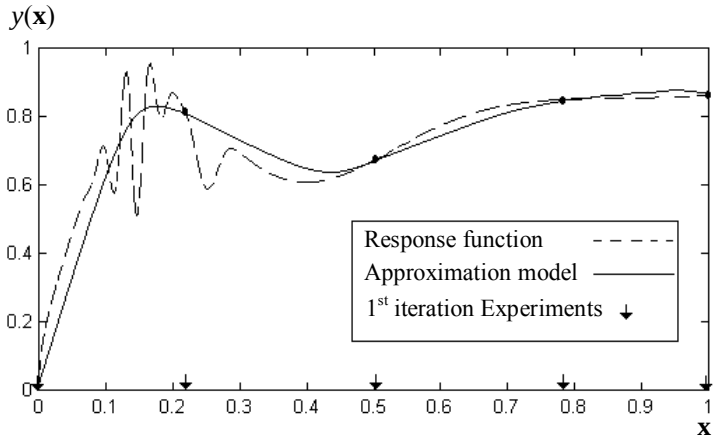



Figure 5: The initial block of 5 experiments and the resulting intermediate approximation model

***Steps 2, 3, and 4 (1st iteration)*** – The surrogate model of the current design is shown in Figure 5. The intermediate surrogate model has three local optima inside the design space: $P_1$={0.17; 0.44; 0.96}.

***Step 5 (1st iteration)*** – $L(\mathrm{x})$ is computed to be:

$$L(\mathrm{x}) = \begin{cases} 0.44 - 0.17 = 0.27 & 0.17 < \mathrm{x} < 0.44 \\ 0.96 - 0.44 = 0.52 & 0.44 < \mathrm{x} < 0.96 \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

***Step 6 (1st iteration)*** – A 10x10 covariance matrix is constructed. The first 5 rows and columns correspond to the covariance of 5 experiments designed in the first block (small arrows in Figure 5).

***Step 7 (1st iteration)*** – The second 5-block of experiments is designed such that the determinant of the 10x10 matrix is maximized. Note that the entries of this matrix are obtained from Equation 8, and using the CCW of Equation 10. The optimum second block is found by searching the set of subsets of size 5 in $\overline{B}_5^1$ exhaustively, and comparing their determinants. (As in Step 1, one could alternatively use a one-point augmentation approach to reduce the computational burden.) This yields $B_5^2$ which is a 5-block that results in the highest determinant of the 10x10 covariance matrix. (The first 5 rows and columns are already known, i.e., $B_5^1$.) Figure 6 shows the first and second 5-block of experiments (a total of 10 experiments). This figure clearly shows that more experiments are designed in the irregular region (between the first two optima in $P_1$ where $L$ is small).

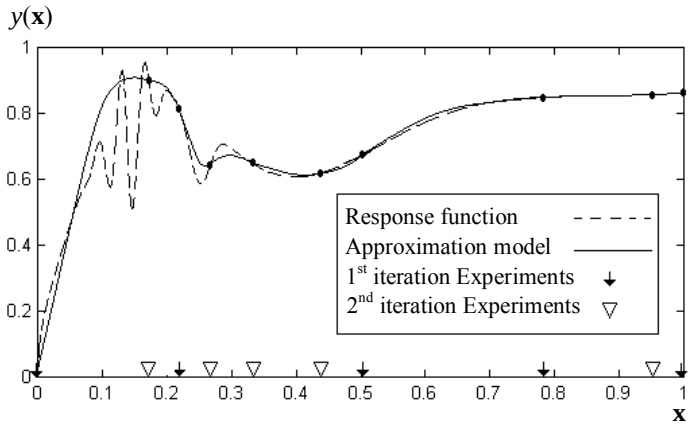



Figure 6: The first iteration and the corresponding approximation model

***Step 8 (1st iteration)*** – There are a total of 10 experiments. So we repeat Steps 2 through 7 to design 5 more experiments, as in the following.

***Steps 2 and 3 (2nd iteration)*** – As before, the new experiments, i.e., $B_5^2$, are evaluated and an interpolating surrogate model is constructed (Figure 6). The new surrogate model is more accurate in the irregular region.

***Steps 4 and 5 (2nd iteration)*** – Optimization of this surrogate model yields: $P_2$={0.15; 0.26; 0.30; 0.41}. Hence, CCW(x) is:

$$L(\mathrm{x}) = \begin{cases} 0.11 & 0.15 < \mathrm{x} < 0.26 \\ 0.04 & 0.26 < \mathrm{x} < 0.30 \\ 0.11 & 0.30 < \mathrm{x} < 0.41 \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

***Steps 6 and 7 (2nd iteration)*** – As before, the new 5-block is determined such that it maximizes the determinant of a 15x15 covariance matrix whose first 10 rows and columns correspond to the experiments designed in the first iteration. The entries of the matrix are evaluated from Equation 8 using the CCW of Equation 11. Figure 7 demonstrates the new 5-block, i.e., $B_5^3$, along with the previous blocks, $B_5^1$ and $B_5^2$, and the new approximation model.

***Step 8 (2nd iteration)*** – we have obtained a total of 15 experiments. The algorithm stops.

Note that both designs in Figures 1 and 7 contain 15 experiments. However, the approximation model from the proposed approach (Figure 7) shows a dramatic improvement as compared to that of a single-stage maximum entropy design (Figure 1). This is because of the tendency of the proposed approach to distribute experiments in the design space with an emphasis on the irregular regions of the domain. The accuracy of the approximation can be improved even further should one decides to continue. As more experiments are conducted, the approximation model approaches the actual response function, revealing the irregular regions in which because of the faster growth of uncertainty more experiments are needed to provide a more accurate approximation model.

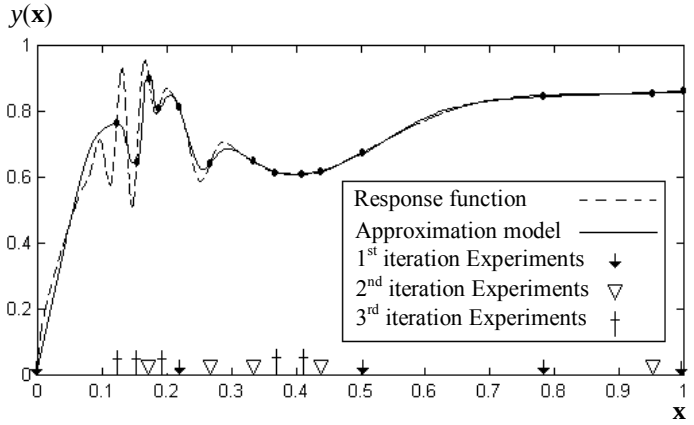
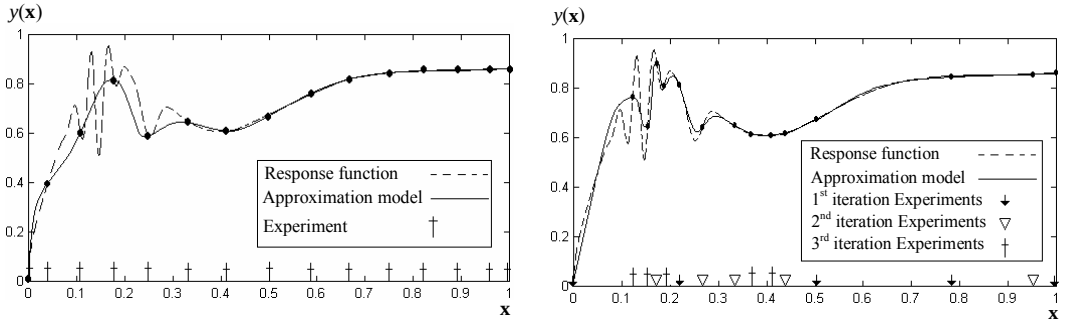



Figure 7: The second iteration and the corresponding approximation model (15 experiments total)

### 3.1.2 Comparison of Results and Discussion

Figure 8 shows the final surrogate models obtained from the proposed approach as well as a typical maximum entropy approach (Section 1.3). To numerically verify the accuracy of the above approximation models, one can select a random sample of points in the design space and compute the deviation between the actual response function and the approximation model. The

Root-Mean-Square-Error (RMSE) of the meta-model can then be estimated: $RMSE = \left\{\sum[\hat{y}(\mathbf{z}_i) - y(\mathbf{z}_i)]^2 / m\right\}^{1/2}$, where $m$ is the size of the random sample. We selected a random sample of 40 points in $[0,1]^2$, and estimated the normalized RMSE (RMSE divided by sample mean) for both approaches.




(a) RMSE = 0.0436 (b) RMSE = 0.0229

Figure 8: Side-by-side comparison of: (a) single-stage maximum entropy/kriging; and (b) the proposed sequential approach

It can be observed that:

- The proposed approach automatically places more experiments in the irregular regions (where more experiments are needed), and therefore creates a meta-model that is much more accurate on average (lower RMSE).
- For functions with small local irregularities (such as the above test example), the proposed approach may not be able to detect the irregular regions by conducting the initial experiments. For example, in Figure 5, the initial sample of experiments almost missed the irregular region in the left-half of the variable range. In such cases, the proposed approach will behave similar to single-stage maximum entropy design (because the irregularities are not detected, and therefore, the covariance matrix is not revised). However, as more and more experiments are designed in subsequent stages, the probability of detecting irregular regions increases. As irregular regions are detected, the

covariance matrix is revised adaptively and the proposed approach starts to deviate from entropy approach and obtain more accurate results. For example, in Figure 6, the second sample of experiments ($2^{nd}$ stage) detected the irregular region to some extent, and designed the $3^{rd}$ stage of experiments accordingly. The final meta-model in Figure 8(b), therefore, is clearly more accurate than the single stage approach in the irregular region.

### 3.2 Crashworthiness Design of Bumper-Rail Assembly of a Pickup Truck

By most measures, the simulation of a crash event involving a typical vehicle is a computationally intensive task. A complete detailed computer model of a passenger vehicle typically involves $10^5 - 10^6$ degrees of freedom and one performance evaluation may require many hours or days of computer time. The complexity of the problem makes design optimization of even a small component of a vehicle a challenging undertaking. Approximation techniques, therefore, are commonly used to handle such problems (e.g., Craig et al. 2003). A detailed multi-purpose finite element model of a 1994 Chevrolet C-2500 pick-up truck was developed at the National Crash Analysis Center at George Washington University (Bedewi et al. 1996). This model is the first of its kind developed specifically to address vehicle safety issues, including front and side performance. As shown in Figure 9, the bumper-rail assembly from this model is used in this analysis[9].

[9] The model is developed by Professor A. Diaz and Graduate Student A. Ravisekar at Michigan State University (Farhang-Mehr et al. 2003).

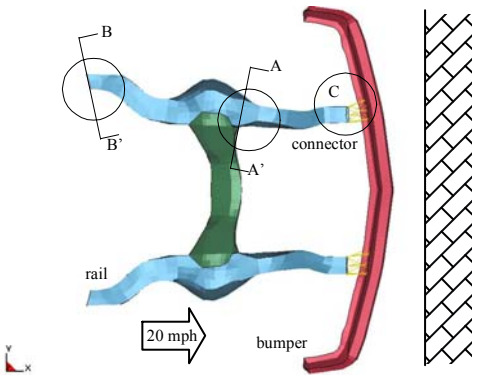



Figure 9: The bumper-rail assembly

The assembly consists of the bumper, the left and right rails, and the cross rail connector. The rail mountings, which connect the bumper and the rails, were replaced with connectors modeled by beam elements whose purpose is simply to engage/disengage the rail and the bumper. Lumped masses are attached at the rear end of the rails, at section B-B'. The assembly is moving forward at a 20 mph when it hits a rigid wall. The analysis was performed at Michigan State University using a standard finite element package for large deformation and impact analysis (Farhang-Mehr et al. 2003). Each analysis run cost about 15 minutes on a Sun Ultra 80 workstation. In design for enhanced crashworthiness, the objective is to improve the protection of the passenger, e.g., by controlling the accelerations experienced by the passenger and the deformation of the structure in the immediate vicinity of the passenger. In our case there are two response functions to be approximated: (1) the maximum force, $F$, transmitted through the rail (measured at section A-A'); and (2) the maximum (X-) displacement $D$ of the section at B-B'. The simulation has two input design variables:

$x_1$ = collapse strength of connector material ($MPa$); and $x_2$ = sheet metal thickness of rail forward of section A-A' ($mm$).

The first variable acts as a "switch" that controls the timing of the failure of the mountings that connect the bumper and the rail. The second variable has a strong effect on both

the amount of deformation and the accelerations transmitted to the rear of the vehicle. The upper and lower bounds for the two variables of the problem, $x_1$ (collapse strength) and $x_2$ (sheet metal thickness), are used for normalization: $z_1 = (x_1–1)/69$; and $z_2 = (x_2–2)/3$. In the next section, the proposed meta-modeling technique of this paper is used to construct surrogate models for this problem. The results are compared with typical maximum entropy approach in Section 3.2.2, followed by discussions.

### 3.2.1 Application of the Proposed Approach

A 50x50 grid is constructed in the design space (a total of 2500 nodes). We also limit our experimentation for each objective to 39 experiments, as 3 blocks of 13 experiments each. At the beginning of the approximation process the non-informative (stationary assumption) holds. Figures 10 (a-c) shows the obtained meta-model from the first, second, and third stages of the design respectively. It is clear from these figures that as the meta-modeling progresses and the meta-model becomes more accurate, more experiments are conducted in the irregular regions (regional adaptivity).

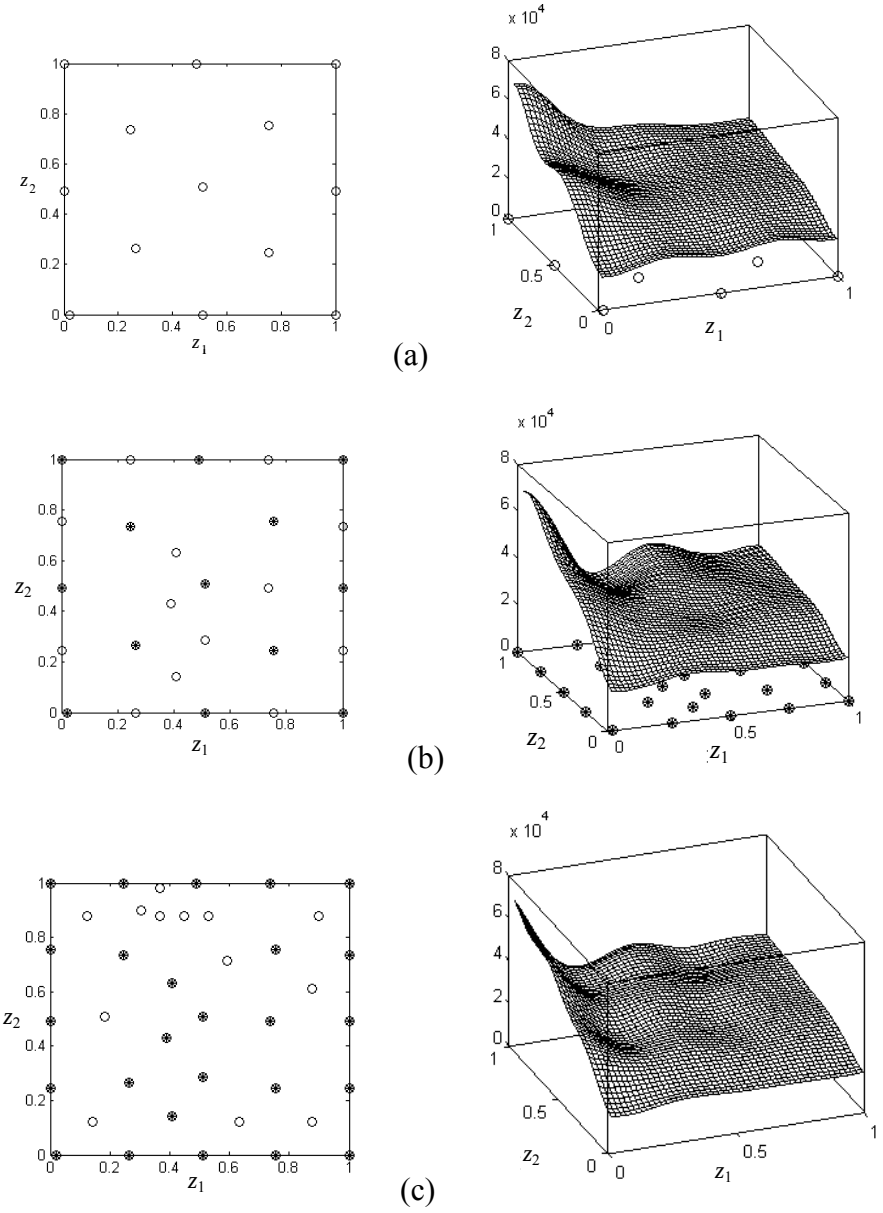



Figure 10 (a), (b), and (c): Meta-models for the maximum force response function ($f_1$) based on 13, 26, and 39 experiments, respectively. Solid circles mark the first block of 13 experiments (already performed) and hollow circles are new experiments in next stage. For the final meta-model, it is estimate that: $RMSE(f_1)/Mean(f_1) \sim 5.1\%$ (using a random sample of 40 experiments)

In a similar fashion for the maximum displacement response function ($f_2$), a total of 39 experiments is designed as 3 blocks of 13 experiments each. Figure 11 shows the final design and the resulting approximation model.

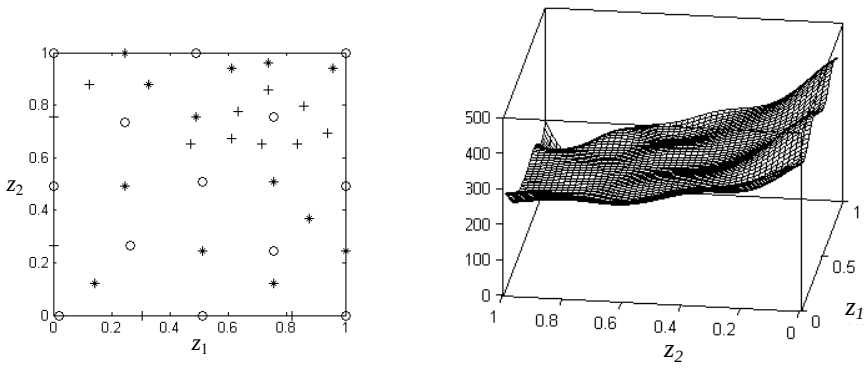
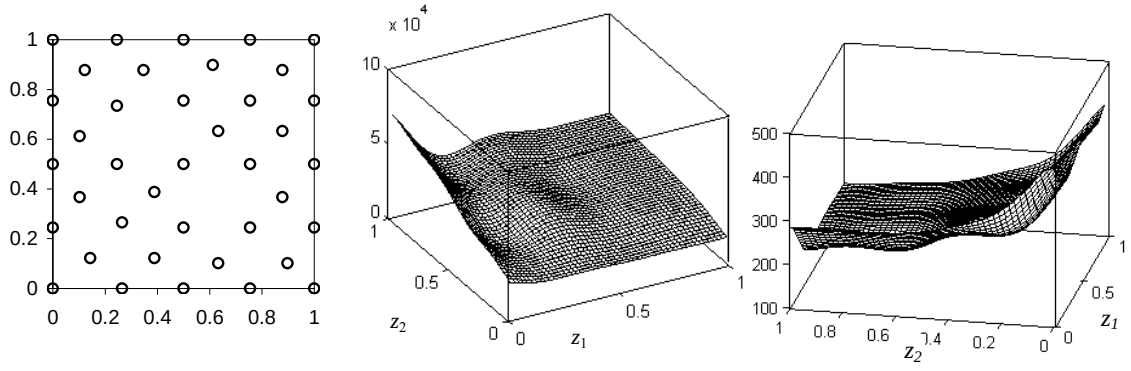



Figure 11: Meta-model for the maximum displacement response function ($f_2$) based on 39 experiments. Hollow circle, asterix, and plus sign, mark the 1st , 2nd, and 3rd blocks of experiments respectively. For the final meta-model, it is estimate that: $RMSE(f_2)/Mean(f_2) \sim 6.3\%$ (using a random sample of 40 experiments)

To measure the accuracy of the obtained meta-models, normalized RMSE's (i.e. RMSE divided by the mean value) are estimated using a random sample of 40 points in $[0,1]^2$: $RMSE(f_1)/Mean(f_1) \sim 5.1\%$; and $RMSE(f_2)/Mean(f_2) \sim 6.3\%$.

### 3.2.2 Comparison of Results and Discussions

Figures 12 (a) and (b) show the design of experiments and the obtained meta-models for maximum force and maximum displacement responses using a typical (single-stage) maximum entropy design approach.

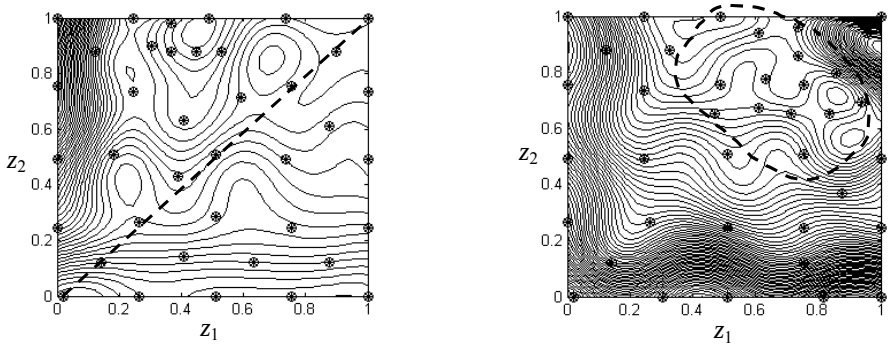



(a) (b) (c)

Figure 12: (a) Maximum entropy design (single-stage); (b) kriging model for maximum force response function ($RMSE(f_1)/Mean(f_1) \sim 7.3\%$); (c) kriging model for maximum displacement response function ($RMSE(f_2)/Mean(f_2) \sim 8.9\%$)

Comparing these meta-models with those of Figures 10(c) and 11 (with the same number of experiments), one may observe the following:

- In this example, the proposed approach produced meta-models that are much more accurate on average (lower RMSE's).
- Figures 13 (a) and (b) show the contour plots of the final meta-models for $f_1$ and $f_2$, respectively. The adaptive nature of this approach can be easily observed in these figures: The concentration of experiments is higher in the irregular regions of the design space where uncertainty is higher (larger variance) due to the fluctuations in the response function. This is a key feature of the proposed approach that helps obtain a more accurate approximation model with fewer experiments.




(a) Maximum Force ($f_1$) (b) Maximum Displacement ($f_2$)

Figure 13: There is a higher concentration of experiments in the irregular regions of the design space where uncertainty is higher. For example in (a), compare the number of experiments in the upper-left and lower-right triangles

Finally, there is a clear tradeoff between $f_1$ and $f_2$ (Compare Figures 10(c) and 11). As mentioned in Section 1.2, in multi-objective optimization problems with conflicting objectives, the set of Pareto-optimal solutions do not necessarily correspond to the optima of the individual objectives. For such problems, global meta-models that are valid throughout the range of the design space (such as the proposed approach of this paper) are of particular interest.

## 4. Concluding Remarks

In this paper, a new sequential maximum-entropy design approach was introduced. This approach takes advantage of the information obtained during previous experiments, builds interim meta-models, and evaluates the irregular behavior of the response function at each point in the design space. The new approach then automatically adapts to the response function behavior by emphasizing the irregular regions of the design space and designing the next set of experiments accordingly. Adaptation to response function behavior is perhaps the most important advantage of the new approach that helps obtain a better representation of the actual response function with fewer experiments. For the two test cases of this paper, the new approach performed significantly better than other approaches, identifying the irregular regions of the

design space and spreading new experiments accordingly. Note that although this approach can help reduce the number of response function evaluations, it may utilize CPU time for internal computations within the algorithm (i.e., optimization of the interim approximation model in Step 4, as well as maximization of entropy in Steps 1 and 7.) Overall, if the response function is very time-consuming to compute, this extra effort to optimally place expensive experiments in the design space should provide a dramatic improvement in the accuracy of the approximation model.

## 5. Acknowledgements

This research is supported in part by the National Science Foundation Grant DMI-0112767. This support is gratefully acknowledged. We would also like to acknowledge Professor A. Diaz and his graduate student A. Ravisear from Michigan State University for providing us with the crashworthiness model of Section 3.2.

## 6. References

Barthelemy, J.-F. M., Haftka R. T., 1993, "Approximation Concepts for Optimum Structural Design - a Review," Structural Optimization, Vol. 5, pp. 129-144.

Bedewi, Z. A., Kan, A., Marzougui., D., 1996, "Validation of a Non-linear Finite Element Vehicle Model Using Multiple Impact Data," ASME Winter Annual Congress and Exposition, Atlanta, GA.

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

Casella, G., Berger, R. L., 1990, Statistical Inference, Duxbury Press, Belmont, California.

Chaloner, K., Verdinelli, I., 1995, "Bayesian Experimental Design: A Review," Statistical Science, Vol. 10, pp. 237-304.

Craig, K. J., Stander, N., Balasubramanyam, S., 2003, "Worst-Case Design in Head Impact Crashworthiness Optimization", International Journal for Numerical Methods in Engineering, Vol. 57, pp. 795-817.

Currin, C., Mitchell, M., Morris, M., Ylvisaker, D., 1991, "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments," Journal of the American Statistical Association, Vol. 86, pp. 953-963.

Farhang-Mehr, A., Azarm, S., 2002, "A Sequential Information-Theoretic Approach to Design of Computer Experiments", CD-ROM Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, GA.

Farhang-Mehr, A., Azarm, S., Diaz, A., Ravisekar, A., 2003, "Bayesian Approximation-Assisted Optimization Applied to Crashworthiness Design of a Pickup Truck," CD-ROM Proceedings of the ASME International Design Engineering Technical Conferences, Chicago, IL.

Jin, R., Chen, W., Sudjianto, A., 2002, "On Sequential Sampling for Global Metamodeling in Engineering Design", CD-ROM Proceedings of IDETC ASME Design Automation Conference, Montreal, Canada.

Jones, D. R., Schonlau, M., Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black Box Functions", Journal of Global Optimization, Vol. 13, pp. 455-492.

Kleijnen, J. P. C., Sanchez, S. M. , Lucas, T., Cioppa, T. M., 2003, "A User's Guide to the Brave New World of Designing Simulation Experiments," Working Paper, Department of Information Management/Center for Economic Research (CentER), Tilburg University, Tilburg, The Netherlands.

Kleijnen, J. P. C., Beers, W. C. M. van, 2003, "Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodeling," Discussion Paper 33, Tilburg University, Tilburg, The Netherlands.

Koehler, J. R., Owen, A. B., 1996, Computer Experiments, Volume 13 of Handbook of Statistics, Elsevier Science, New York, NY.

Lindley, D. V., 1956, "On a Measure of the Information Provided by an Experiment," The Annals of Mathematical Statistics, Vol. 27, pp. 986-1005.

Mardia, K. V., Marshall, R. J., 1984, "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," Biometrika, Vol. 71, pp. 135-146.

Martin, J. D., Simpson, T. W., 2002, "Use of Adaptive Metamodeling for Design Optimization", 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, GA.

Martin, J. D., Simpson, T. W., 2003, "A Study on the Use of Kriging Models to Approximate Deterministic Computer Models," CD-ROM Proceedings of the ASME International Design Engineering Technical Conferences - Design Automation Conference, Chicago, IL.

Myers, R. H., Montgomery, D. C., 1995, "Response Surface Methodology: Process and Product Optimization Using Designed Experiments", John Wiley & Sons, New York.

Osio, I. G., Amon, C.H., 1996, "An Engineering Design Methodology With Multistaqge Bayesian Surrogates and Optimal Sampling," Research in Engineering Design, Vol. 8, pp. 189-206.

Pacheco, J. E., Amon, C. H., Finger, S., 2001, "Developing Bayesian Surrogates for Use in Preliminary Design", CD-ROM Proceedings of ASME Design Engineering Technical Conference, Design Theory and Methodology, Pittsburgh, PA.

Roux, W. J., Stander, N., Haftka, R. T., 1998, "Response Surface Approximations for structural optimization", International Journal for Numerical Methods in Engineering, Vol. 42, pp. 517-534.

Sacks, J., Welch, W. J., Mitchell, T. J., Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Statistical Science, Vol. 4, pp. 409-435.

Sasena, M., Parkinson, P., Goovaerts, P., Papalambros, P., Reed, M., 2002, "Adaptive Experimental Design Applied to an Ergonomics Testing Procedure," CD-ROM Proceedings of ASME, IDETC Conference, Montreal, Canada.

Sasena, M. J., Papalambros, P.Y., Goovaerts, P., 2000, "Metamodeling Sampling Criteria in a Global Optimization Framework," 8th AIAA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA.

Shannon, C. E., 1948, "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, pp. 379-423 and 623-656 (July and October).

Shewry, M. C., Wynn, H. P., 1987, "Maximum Entropy Sampling," Journal of Applied Statistics, Vol. 14, pp. 165-170.

Turner, J. C., Campbell, M.I., Crawford, R. H., 2003, "Generic Sequential Sampling for Meta-Model Approximations", CD-ROM Proceedings of the ASME International Design Engineering Technical Conferences - Design Automation Conference, Chicago, IL.

Wang, G. G., Dong, Z., Aitchison, P., 2001, "Adaptive Response Surface Method - A Global Optimization Scheme for Computation-intensive Design Problems", Journal of Engineering Optimization, Vol. 33(6), pp. 707-734.