Michael J. Sasena
Emmeskay, Inc., Plymouth, MI 48170
e-mail: msasena@umich.edu

Matthew Parkinson
University of Michigan,
Department of Biomedical Engineering,
Ann Arbor, MI 48109-2125
e-mail: mparkins@umich.edu

Matthew P. Reed
University of Michigan Transportation Research Institute, Associate Research Scientist,
Ann Arbor, MI 48109-2150
e-mail: mreed@umich.edu

Panos Y. Papalambros
University of Michigan,
Department of Mechanical Engineering,
Ann Arbor, MI 48109-2125
e-mail: pyp@umich.edu

Pierre Goovaerts
Biomedware, Inc. and PGeostat, LLC,
Ann Arbor, MI 48103
e-mail: goovaerts@biomedware.com

Improving an Ergonomics Testing Procedure via Approximation-based Adaptive Experimental Design

Adaptive design refers to experimental design where the next sample point is determined by information from previous experiments. This article presents a constrained optimization algorithm known as superEGO (a variant of the EGO algorithm of Schonlau, Welch, and Jones) that can create adaptive designs using kriging approximations. Our primary goal is to illustrate that superEGO is well-suited to generating adaptive designs which have many advantages over competing methods. The approach is demonstrated on a novel human-reach experiment where the selection of sampling points adapts to the individual test subject. Results indicate that superEGO is effective at satisfying the experimental objectives. [DOI: 10.1115/1.1906247]

Introduction

An experimental run, be it a physical experiment or done on a computer, can be viewed as a black-box function. Input values are sent to the black-box, and the associated outputs are returned. Design of experiments (DOE) studies how to select the set of input values at which to evaluate the black-box in order to explore its behavior efficiently. There are many ways to quantify how well a particular design achieves that goal. Traditional designs such as factorial and composite designs, Latin hypercubes and orthogonal arrays attempt to spread points around the design space according to different measures of spatial distribution. The reader is referred to Montgomery [1] for more background on classical design of experiments.

In some applications, a particularly useful approach is to determine the set of input values sequentially, taking into account information from the previous experiments. These so-called *adaptive designs* can select one point at a time or be performed in stages, where several new inputs are chosen at each stage.

Adaptive designs are complicated to generate because they must solve an optimization problem to determine the next set of input values. The topic is currently an active area of research for the experimental design community [2–6] and specialized algorithms have been developed [7]. However, because adaptive design involves solving an optimization problem, general nonlinear programming techniques can be used to generate the designs. This has led to interest from a wider optimization community as well [8–15].

Traditionally, optimization algorithms have been used to find the set of inputs that result in the best objective and constraint values. In the case of adaptive design, however, the goal is not a single solution point. Rather, it is the search history itself that is of interest. The goal is to find a *set* of points that collectively give useful information. While nearly any optimization algorithm can solve the former problem, not many are well suited to the latter. Increasingly popular approximation-based methods can perform well at both.

The framework used here to create adaptive designs is a constrained nonlinear optimization algorithm known as superEGO [13,16,17], a variation of the EGO algorithm of Schonlau, Welch, and Jones [18]. It uses approximations of the data to search for the next iterate. In the case of EGO and superEGO, the approximation consists of a kriging model [19]. The search behavior is controlled by the infill sampling criterion (ISC) which quantifies the benefit of sampling a specific location. It must be numerically maximized to determine the next iterate. In our experience, Jones' DIRECT algorithm [20,21] worked well at solving this auxiliary global optimization problem. The kriging models are iteratively updated with the new samples until the termination criterion is reached. The flowchart of the superEGO algorithm is shown in Fig. 1.

In much of the research on adaptive sampling, measures such as the integrated mean-squared error or maximum mean-squared error are used as the criterion for selecting the next sample point [22,23]. The study presented in this work takes advantage of the flexibility of superEGO by defining sampling criteria specifically tailored to the goals of the application. In addition, superEGO is able to incorporate constraint information to prevent injury to the test subject.

The remainder of the paper is organized as follows. We begin by briefly describing kriging and its application in this work. We then describe the application study, a test procedure for an ergonomics experiment. The methods used to address the problems are explained, and the results are shown. We conclude by highlighting the contributions in this work and possible avenues for future research.

Contributed by the Design for Manufacturing Committee for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received August 12, 2003; revised December 12, 2004. Associate Editor: D. Kazner.

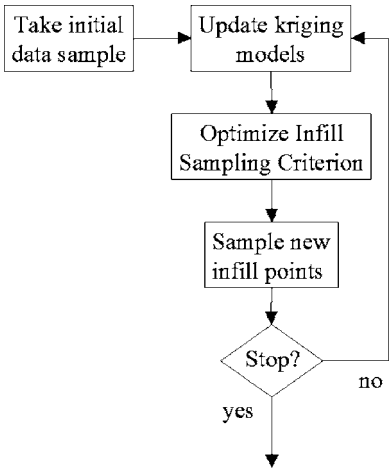

Fig. 1 Flowchart of the superEGO algorithm

An Overview of Kriging

We begin this section with an introduction to the approximation method used. Much of the explanation follows Schonlau's derivation [18]. We will also describe a method for fitting models that are noninterpolating.

Kriging Formulas. There has been much interest lately in a form of curve fitting and prediction known as kriging. Originally developed for use in mining and soil sciences in the 1950's, it grew into the field of study called geostatistics. In the late 1980's, statisticians introduced Design and Analysis of Computer Experiments (DACE) by adapting the kriging techniques to deterministic, simulation-based data [23]. This work, like most kriging applications outside of civil and environmental engineering, uses the DACE approach to kriging.

A kriging metamodel, which takes the form $Y(\mathbf{x})=f(\mathbf{x})+Z(\mathbf{x})$, is comprised of two parts: a polynomial $f(\mathbf{x})$ and a functional departure, $Z(\mathbf{x})$, from that polynomial. We will assume a constant term polynomial, that is $f(\mathbf{x})=\beta$. This simplification is often used and provides sufficiently accurate results because of the flexibility of $Z(\mathbf{x})$ [22]. More precisely, $Z(\mathbf{x})$ represents uncertainty in the mean of $Y(\mathbf{x})$ with $E(Z(\mathbf{x}))=0$ and

$$\mathrm{Cov}(Z(\mathbf{w}),Z(\mathbf{x}))=\sigma^2 R(w,x), \tag{1}$$

where σ^2 is a scale factor (the process variance) that can be tuned to the data, and $R(\mathbf{w},\mathbf{x})$ is known as the spatial correlation function (SCF).

The choice of the SCF determines how the metamodel fits the data. There are many choices of $R(\mathbf{w},\mathbf{x})$ to quantify how quickly and how smoothly the function moves from point $\mathbf{x}$ to point $\mathbf{w}$. One of the most common SCF's (shown here for one-dimensional points w and x) used in DACE is

$$R(w,x)=e^{-\theta|w-x|^p}, \tag{2}$$

where $\theta>0$ and $0<p\leqslant 2$. As with all choices of the SCF, the function tends to zero as $|w-x|$ increases.

To extend the SCF to several dimensions, common practice is to multiply the correlation functions. The so-called *product correlation rule* applied to Eq. (2) results in

$$R(\mathbf{w},\mathbf{x})=\prod_{k=1}^{d} e^{-\theta^k|w^k-x^k|^{p^k}}, \tag{3}$$

where the superscript $k=1$ to d refers to the dimension.

For n sample points, let $\mathbf{R}$ be the $n\times n$ correlation matrix with element i,j defined by $R(\mathbf{x}_i,\mathbf{x}_j)$ as in Eq. (3) and let

$$\mathbf{r}_\mathbf{x}=[R(\mathbf{x}_1,\mathbf{x}),R(\mathbf{x}_2,\mathbf{x}),\ldots,R(\mathbf{x}_n,\mathbf{x})]^t \tag{4}$$

be the $n\times 1$ vector of correlations between the point at which to predict, $\mathbf{x}$, and all the previously sampled data points $\mathbf{x}_i$. We can now define the kriging prediction equation as

$$\hat{y}(\mathbf{x})=\hat{\beta}+\mathbf{r}_\mathbf{x}^t\mathbf{R}^{-1}(\mathbf{y}-\mathbf{1}\hat{\beta}), \tag{5}$$

where $\mathbf{y}$ is the vector of sampled output data, $\mathbf{1}$ is a vector of ones, and

$$\hat{\beta}=(\mathbf{1}^t\mathbf{R}^{-1}\mathbf{1})^{-1}\mathbf{1}^t\mathbf{R}^{-1}\mathbf{y} \tag{6}$$

is the generalized least squares estimator of β.

A very useful feature of kriging is the availability of the *kriging variance* or mean squared error (MSE), an estimate of the uncertainty in the prediction of $\hat{y}(\mathbf{x})$.

$$\mathrm{MSE}[\hat{y}(\mathbf{x})]\equiv\hat{s}^2(\mathbf{x})=\hat{\sigma}^2(1-\mathbf{r}_\mathbf{x}^t\mathbf{R}^{-1}\mathbf{r}_\mathbf{x}), \tag{7}$$

where $\hat{\sigma}^2$ is the generalized least squares estimator of the process variance of Eq. (1) computed as

$$\hat{\sigma}^2=\frac{1}{n}(\mathbf{y}-\mathbf{1}\hat{\beta})^tR^{-1}(\mathbf{y}-\mathbf{1}\hat{\beta}). \tag{8}$$

The most difficult aspect of kriging is fitting the model parameters (θ and p) that describe the covariance function of Eq. (3). In the DACE approach, models are fit with maximum likelihood estimation (MLE) by numerically minimizing the equation

$$\mathrm{MLE}=\frac{1}{n}(n\ln\hat{\sigma}^2+\ln(\det\mathbf{R})), \tag{9}$$

which is a function of only the model parameters.

Smoothed Kriging Models. The approach to kriging described above results in an interpolating surface. That is, at the sampled data locations, the predictor goes exactly through the data, and the kriging variance returns to zero. While this is appropriate for many engineering applications, there are cases where a noninterpolating version of the kriging model would be very beneficial-for example, when the data are highly noisy. To allow kriging to smooth the data, an additional parameter is introduced into the SCF of Eq. (3) to account for the measurement error, $E(\mathbf{x})$. The kriging model now takes the form $Y(\mathbf{x})=f(\mathbf{x})+Z(\mathbf{x})+E(\mathbf{x})$, and the SCF appears as

$$R(\mathbf{w},\mathbf{x})=(1-nugget)\prod_{k=1}^{d} e^{-\theta^k|w^k-x^k|^{p^k}}, \tag{10}$$

with $0\leqslant nugget<1$. The name refers to the nugget effect model used in geostatistics to describe the discontinuity of the covariance function as $|w-x|$ approaches zero. The scaling factor (1-nugget) is the same as the ratio $\sigma_Z^2/(\sigma_Z^2+\sigma_E^2)$ defined in the DACE literature [22,23]. In either case, the additional parameter ranges between zero and one and must be fit via MLE along with θ and p. We prefer the notation of Eq. (10) simply because it is easier to view the value of *nugget* which is usually quite close to zero (e.g., 1e-5) rather than $\sigma_Z^2/(\sigma_Z^2+\sigma_E^2)$ which is usually quite close to one (e.g., 0.99999). Typically, a single scaling factor is applied to the covariance model. However, one can fit a separate *nugget* parameter for each dimension as

$$R(\mathbf{w},\mathbf{x})=\prod_{k=1}^{d}(1-nugget^k)e^{-\theta^k|w^k-x^k|^{p^k}}, \tag{11}$$

so that more smoothing is applied in some directions than others.

A *nugget* value of zero means that the predictor goes exactly through the data. Increasing the *nugget* value results in smoothed models because collocated points no longer have perfect correlation. A *nugget* value of one implies no correlation in the data, and the kriging model degenerates to the global trend $f(\mathbf{x})$. It is im-

portant to note that the scaling factor (1-*nugget*) is *not* applied to the diagonal of the **R** matrix, but it *is* applied to the entire $\mathbf{r_x}$ vector even when a prediction point coincides with a sample point [19]. This is done because a point is perfectly correlated with itself (diagonal of **R** matrix consists of ones), but would not be perfectly correlated with a second sample point at the same location ($\mathbf{r_x}$ vector) in the presence of measurement error.

The *nugget* parameter can also be used to improve the stability of the kriging model. When superEGO is used for design optimization, points begin to cluster around the optimum, making the **R** matrix nearly singular and difficult to invert. Our experience has been that very small values for *nugget* tend to improve the conditioning without overly smoothing the data. It can be fit via MLE along with the other model parameters, but we have found that fixing the value at $nugget = 10^{-12}$ works well across many different examples.

A simple one-dimensional example is shown to more easily visualize the smoothing effect. Consider

$$y = (x - 5)^2 + \epsilon, \tag{12}$$

where $0 \leqslant x \leqslant 1$ and ϵ is a uniformly distributed random number (i.e., a white noise term) between $\pm\frac{1}{2}$. Nineteen sample points are taken: 11 equally spaced points and 8 additional points near the minimum. Using MLE, kriging models are fit to $[\theta, p]$ for the interpolating and to $[\theta, p, nugget]$ for the smoothing covariance models of Eqs. (3) and (10), respectively. Figure 2 shows the predicted value of the interpolating (solid line) and smoothed (dashed line) kriging models. In the area where the noise is prominent, the interpolating model oscillates in order to pass through each of the data samples, whereas the other model is much smoother in a least squares sense.

Figure 3 shows the kriging variance. For both the interpolating and smoothed models, the variance is lower in the middle because more data are available in that region. A consequence of the smoothing effect is that the variance model no longer returns to zero at the sample points. This is because the smoothing effectively increases the uncertainty in the model everywhere. In this example, there is little difference between the variance of the smoothed model at the sample points and the variance in locations far from the data. Because of this feature, one must be cautious when using the kriging variance as a measure of the uncertainty in smoothed kriging models. It is recommended that the interpolating covariance model of Eq. (3) rather than Eq. (10) be used in conjunction with Eq. (7) for a more useful estimate of the model uncertainty.

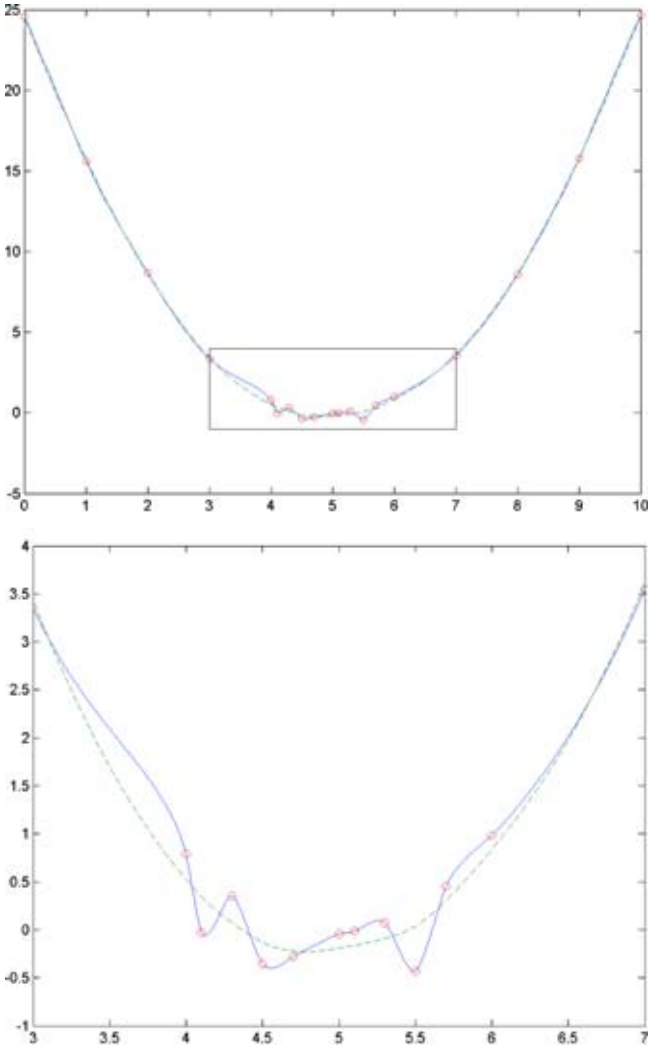

Fig. 2 Kriging predictions for interpolating (solid) and smoothed (dashed) models (top). A close up of the boxed region is shown at bottom. Data points are shown as circles.

Case Study

In this section the tools described above are used to create adaptive designs for an ergonomics experiment which explores the reach range of a seated test subject. A subjective, and therefore noisy, measure of the difficulty of reaching a specific target is obtained. SuperEGO creates a model of the difficulty and adaptively determines where to gather additional data from the subject.

Problem Description. Figure 4 shows the setup for the experiment. A subject is seated in a mock-up of an industrial or workplace environment. A computer controls a reach target to raise and lower it, to move it nearer to or farther from the subject and to rotate the chair. Once the target is positioned, the subject is asked to reach to the target and depress the button for 2 s. The subject then rates the difficulty of the reach on a scale of 1–10, with 10 being the most difficult. Any target location that is unreachable is rated an 11. Maintenance of balance, the amount of exertion or degree of awkward posture required, and distance to the target all contribute to the difficulty of the reach. The experiment has two main purposes: to form a predictive model of a subject's reach difficulty at any location, and to estimate the maximum reach range in any direction.

The original testing procedure collected data in six chair rotation positions, each of which was sampled along five discrete rays. The length of the rays were calculated before the experiment by estimating the maximum reach length of the subject. This is a quantity commonly measured and indexed for ergonomics designers. Targets were then placed at 5%, 25%, 50%, 70%, 85%, 95%, 100%, and 105% of the estimated maximum reach along five discrete rays. The chair was simultaneously rotated to one of six discrete planes. A 40 point sample from the original ray-based approach is shown in Fig. 5 for the 90 deg (lateral) plane. The plot is shown from a perspective behind a subject extending their right arm in a lateral direction. The [0,0] reference is taken as the h-point, an ergonomic reference point at the mid-hip level along the body centerline. The box in the lower left is the area around the test subject, and the arc shows a radial distance of 1600 mm from the h-point. Two procedures were in place: a short test and a long test. The short test sampled 100–150 reaches and required about 1.5 h. The long test sampled 220–270 reaches and required about 3 h.

The experimental requirements clearly indicate adaptive design as an attractive approach, but can one justify the added complication? It would be much simpler to generate a generic space-filling design via classical DOE methods (e.g., orthogonal arrays, factorial designs, etc.). Because people vary in stature, one would have to scale the DOE for each subject. Doing so is not necessarily

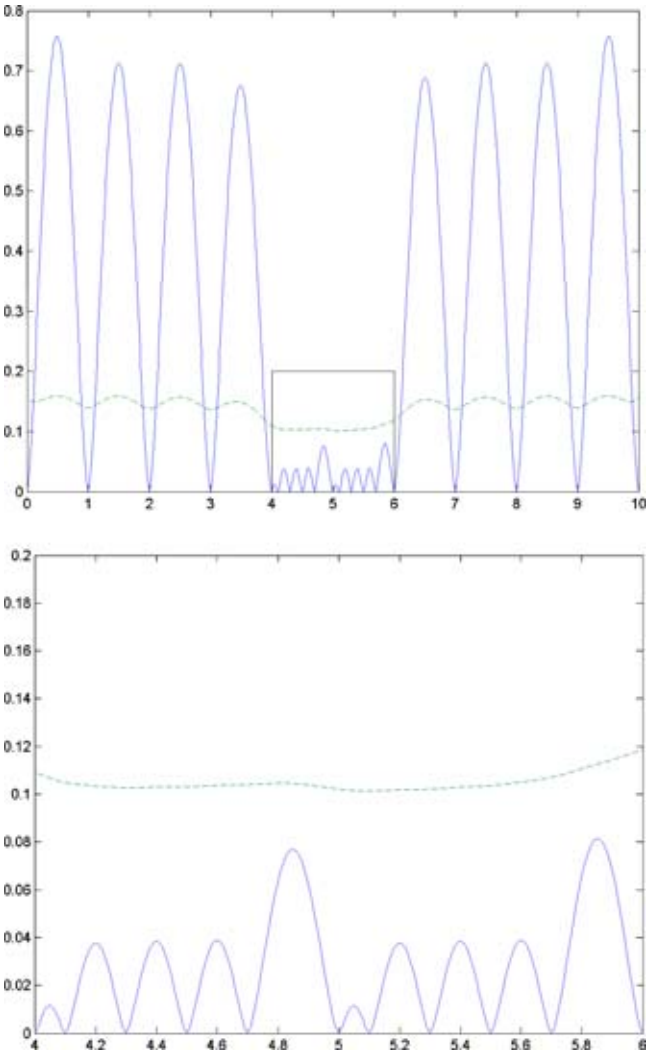

Fig. 3 Kriging variance for interpolating (solid) and smoothed (dashed) models (top). A close up of the boxed region is shown at bottom.

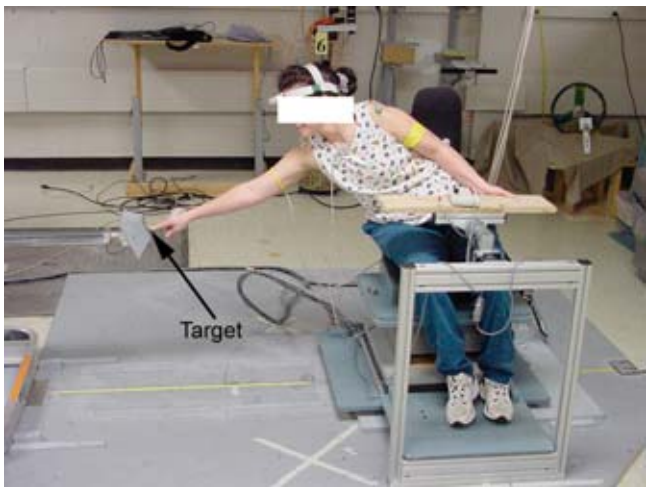

Fig. 4 Example of a test subject reaching for the target. The chair can be rotated and the target can be moved up and down, left and right.

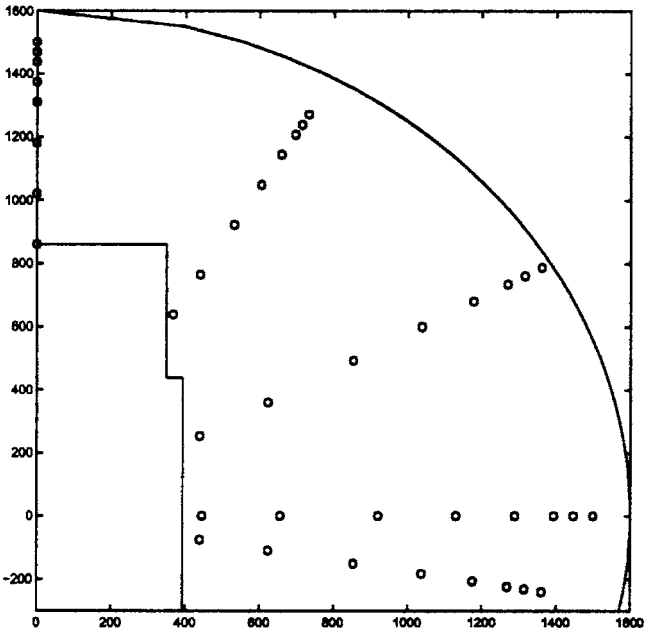

Fig. 5 Example of the static, ray-based experimental design, shown for the lateral plane

reliable due to the number of physical characteristics that determine reach difficulty ratings. Individuals also differ in their physical limitations and flexibility, therefore two subjects of identical size may rate identical target locations differently. In addition, there is an interest in subject-specific information (e.g., a subject's maximum reach range). Thus, it is not practical to create a scalable, static design that would fulfill all the needs of this experiment. For these reasons, adaptive sampling was selected as an appropriate tool for the test procedure.

The researchers wished to reduce the time required for subject testing without losing accuracy in the models obtained from the data. Reducing the time meant a reduction in the number of sample points taken for each subject. The maintenance of accuracy meant a sample procedure able to produce a space-filling design by adapting to the individual test subject. The researchers were interested in more than simply spreading the data points around the design space, however. They wished to obtain a data set that had a good distribution of points in the low, middle, and extreme reach difficulty ranges, as defined by Table 1. Zone 1 corresponds to easy reaches, which tend to be in the middle of the design space. Zone 2 is for medium reach ranges, which occur typically in the area where the subjects begin to lean out of the chair to reach the target. Zone 3 is for points just within and just outside the maximum reach range of the subject.

In summary, the experimental goals were as follows:

(1) Gather data that adapts to the test subject;
(2) Gather data at least as good as ray-based approach:

   (a) maximize spacing between samples to provide a good global model;
   (b) distribute points evenly in the 3 reach zones;
   (c) sample points around the maximum reach range;

Table 1 Ranges chosen for 3D testing

Zone	Range of reach difficulty
1	0–6.5
2	6.5–9.5
3	9.5–11

(3) Use constraint information to avoid hitting the subject or wasting time;
(4) Reduce total experiment time to 1 h (100 reaches).

To fulfill these objectives, a series of experiments were conducted using adaptive sampling to create more efficient experimental designs. Next, we will describe how superEGO was able to achieve these goals.

SuperEGO Implementation. The goal of the adaptive experiment was to gather relatively few data (about 100) for a globally accurate model while sampling many points near the maximum reach range. For the model to be globally accurate, the experimental design must be space-filling, regardless of the subject's stature or flexibility.

Various sampling criteria were implemented in a pilot study [24]. After some experimentation, a new sampling routine was devised that leveraged the flexible nature of superEGO:

$$\mathrm{ISC} = \begin{cases} \hat{s}^2(\mathbf{x}), & \text{if } (y_{\min} + \mathrm{tol}) \leqslant \hat{y}(\mathbf{x}) \leqslant (y_{\max} - \mathrm{tol}) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $\hat{y}(\mathbf{x})$ is the predicted reach difficulty, and $y_{\min}$ and $y_{\max}$ define the range of difficulty values to search over. For example, setting $y_{\min}$ and $y_{\max}$ to 0 and 6.5, respectively, would cause superEGO to search for the location of maximum model uncertainty within zone 1. At each iteration, the data was sorted into the three bins according to Table 1, and $y_{\min}$ and $y_{\max}$ were set to the boundaries of the zone with the fewest data points. SuperEGO then searched for the location of maximum uncertainty for the range of reach difficulty that had been least sampled thus far.

In practice, the location of maximum variance often occurred along the zone boundaries. For example, suppose an iterate is attempting to produce a sample point in zone 2. The location of maximum variance could have a predicted reach difficulty value of 9.5, the upper limit of that zone. Due to the modeling inaccuracy, the target might actually be out of the subject's reach, resulting in another data point in zone 3 instead of the desired zone 2. This effect typically manifested itself as a lower than expected percentage of data points within zone 2 and a higher than expected percentage of data in zone 3. To combat this effect, a tolerance was imposed to shrink the zone boundaries away from each other [tol in Eq. (13)]. During testing, tol was set to 0.5 for the first 20 iterations, then zero afterwards, once the kriging models were predicting more accurately. By including the tolerance factor, the new sample point was located in the intended zone more frequently.

Two constraints were also incorporated into the solution of the ISC problem at each iteration: the encroachment and the predicted out-of-reach constraints. The first kept the target from getting too close to the subject and was represented by the rectangular region in the lower left corner of Fig. 5. The second prevented placing the target in locations that were clearly out of reach of the subject, thus wasting time. It modeled the out-of-reach zone by a kriging model of the reach difficulty and therefore adapted to the subject. Locations where the reach difficulty was predicted to be above 10 were considered infeasible.

During the experiments, a *smoothed* version of the kriging model was used because the data were extremely noisy. This yielded a more well-behaved model of the reach difficulty. However, an *interpolating* model was used for the kriging variance to quantify the uncertainty in the reach difficulty model. Because the sampling criteria used the kriging variance to create a space-filling design, it was critical that the variance tend toward zero at the sampled locations and increase monotonically as the distance from neighboring samples increased. Thus, the interpolating version (i.e., nugget=0) was selected to model the uncertainty in the predictions. We have taken advantage of the flexibility of superEGO to use smoothed kriging models in the constraints while using the interpolating model to calculate variance estimates.

Generally speaking, the MLE fitting process of the kriging models is not very robust [13]. Small changes in the data set can unpredictably lead to poor fits. For the case study, a kriging model of the subjects' perceived reach difficulty was successfully fit using the initial data set, then the model parameters $[\theta, p, \mathrm{nugget}]$ were held fixed for the rest of the experiment. Doing so reduced the waiting time between reaches and yielded more reliable reach difficulty models. Fixing the covariance model parameters does not imply that the resulting kriging model is fixed, but only determines how the model interpolates the data. It still changed shape as data were added.

In summary, the following auxiliary optimization problem was solved for each iteration:

$$\max_{\mathbf{x}} \mathrm{ISC}(\mathbf{x}) \tag{14}$$

$$\text{subject to: } g_1: \mathbf{x} \notin \text{Encroachment zone}$$

$$g_2: \hat{y}(\mathbf{x}) \leqslant 10 \tag{15}$$

Predictions for $\hat{y}(\mathbf{x})$ used Eq. (10) for the covariance model while predictions for $\hat{s}^2(\mathbf{x})$ used Eq. (3). The solution to Eq. (14) yielded the location to place the target next. The subject reached for the target and rated the difficulty. The $\mathbf{R}$ matrix of the kriging model was then updated to incorporate the new data (without refitting the model parameters), and the process was repeated until the specified number of reaches were performed.

SuperEGO requires an initial data set to begin. Frequently, a random Latin hypercube is used. For this study, however, we manually created a static set of 24 reaches spread out over a range of locations. The reaches were the same normalized position for each subject, but were scaled according to their measured maximum overhead reach. As mentioned above, scaling a static DOE *a priori* would not satisfy the goals of the experiment, but it did work well enough to initialize the kriging model.

The initial samples were selected to be somewhat space filling while including a number of points around the perimeter of the feasible space. It is a well known feature of kriging that the variance tends to be large around the outer edge of the design space, which is less densely sampled. Populating points in this area at the onset aims to reduce the number of far away samples chosen after initialization.

Once the initial model was fit, adaptive sampling was performed until approximately 100 reaches had been sampled. The chair in which they were seated was allowed to rotate to any arbitrary angle and the target could be positioned anywhere, resulting in a fully three dimensional data collection protocol. By comparison, the original procedure only sampled in six discrete planes (i.e., chair rotations), each of which contained five discrete rays. The resulting data could then be used to create an approximate model for predicting the reach difficulty of the subject. Figure 6 shows an example of the smoothed kriging model made from one subject's reach data.

Results

A group of 49 test subjects performed the adaptive experiment as described above. The data was compared to data obtained earlier from 38 test subjects who had performed the original static experiment, i.e., the ray-based approach. The Problem Description section states that the second objective of the experiment was to obtain a data set "at least as good as the ray-based approach." Here we compare the two data sets using three metrics: separation distance (goal 2a), zone balance (goal 2b), and distance to the maximum reach boundary (goal 2c).

The first comparison measured how far the sample points were spread out around the design space. Large separation distances are preferred because interpolating models built on space-filling designs tend to predict better. For a given subject, the norm distance

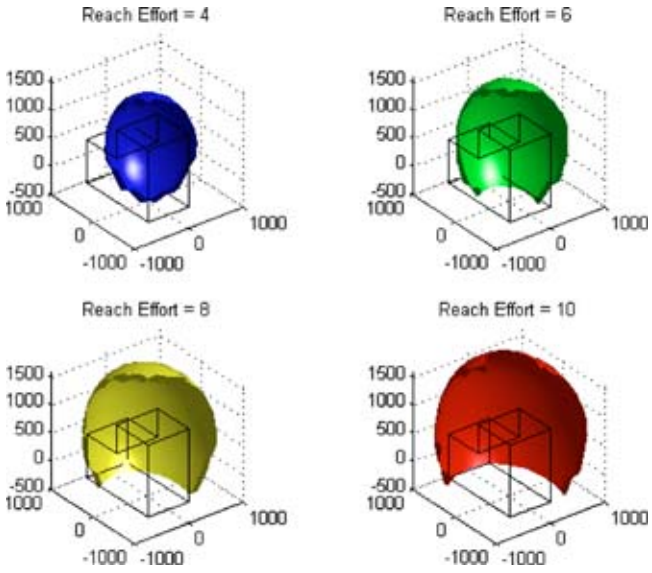

Fig. 6 Plots of "iso-reach difficulty." The wire box indicates the encroachment zone around the test subject. Distances are measured in mm.

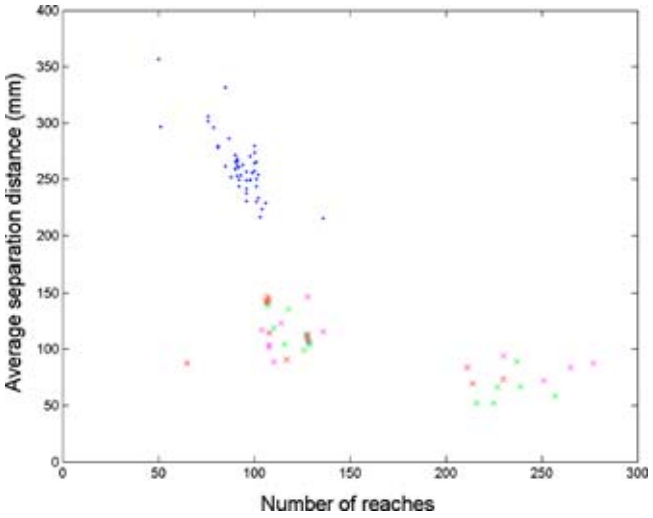

Fig. 7 Average separation distance of sample points for each test subject. Points and x's are results from adaptive and static experiments, respectively.

from each sample point to the nearest neighboring sample point was measured. Then, the average value of those separation distances was computed to give a measure of how spread out the data set was for a given subject. More complicated metrics such as discrepancy measures could also be used, but the trends from this comparison were adequately clear.

Table 2 shows the results averaged across all test subjects. The average separation distance for the adaptive test was significantly larger than that of the static test, meaning the adaptive DOE was more space-filling. The minimum separation distance results indicate that the two data points closest to each other were farther apart in adaptive experiments than in static ones. That means that data redundancies were less pronounced for the adaptive data because even the closest data pair were over 11 cm apart—twice as far as the static experiments.

It is important to note that the separation distance is highly influenced by the number of sample points in the data. The more sample points in a given space, the closer they must be to each other. Figure 7 shows the average separation distance for each test subject. The data from adaptive and static experiments are shown as points and x's, respectively.

The plot includes both kinds of static experiments: the short test (about 120 reaches) and the long test (about 240 reaches). Clearly the long tests resulted in a smaller average separation distance. Compared to the short test data sets however, the adaptive experiments provided a significantly more spread out experimental design. This means that, for data sets of comparable size, the adaptive designs were more space-filling and would provide a better model of the reach difficulty. This supports the claim that adaptive sampling has met the experimental goal 2a.

The next comparison looked at how well the experimental designs spread sample points into the three different zones defined in Table 1. Ideally, the data would have 33.3% of the sample points in each of the three zones. Table 3 shows the actual percentage of points in each zone and how far they deviated from a perfect balance (averaged across all subjects). The results from the adaptive testing appear to have a slightly better balanced design. The average deviation from ideal is 7.7% for the adaptive design, compared to 13.0% for the static design. Note that, on average, the adaptive sampling criterion did in fact succeed at producing 33.3% of the sample points in zone 1. There was however an uneven split between zones 2 and 3, with more points lying in zone 3. As was described in the previous section, the search tended to put more points in zone 3 than desired due to errors in the predicted reach value. If one could better avoid sampling along the predicted boundary between zones 2 and 3, a better balance may be achieved. The tolerance parameter included in the ISC was unable to do so as well as hoped. Perhaps a larger value of tol in Eq. (13) would result in a more conservative estimate of the reach zone and increase the zone prediction success rate, but further experimentation would be required to evaluate this hypothesis. Despite this shortcoming, superEGO still achieved a better reach zone distribution that static experiments, thereby satisfying the experimental goal 2b.

The final comparison measured how tightly the data set defined the maximum reach envelope. For each data set (either from adaptive or static sampling), a smoothed kriging model was made for predicting the maximum reach range. The distance between any sample point in zone 3 and the nearest location where $\hat{y}=10.5$ was computed. This produced an estimate of how far the sample point was from the predicted maximum reach envelope. For each subject, we computed the average, median, minimum and maximum of these distances for all points in zone 3. The lower the value, the closer the sample points were to the predicted maximum reach envelope, and therefore the more accurately the envelope was mapped.

From the results summarized in Fig. 8, three trends can be observed. First, the adaptive experiments provided sample points

Table 2 Separation distance of data set (in mm) averaged over all subjects of a given testing protocol.

	Adaptive	Static
Average separation distance	262.1	101.6
Minimum separation distance	112.5	24.7

Table 3 Percent of reaches in each zone averaged over all test subjects of a given testing protocol. Deviation from an ideal balance is shown in parentheses.

Zone	Adaptive (Δ)		Static (Δ)	
1	33.3%	(0%)	52.9%	(+19.6%)
2	21.7%	(−11.6%)	26.5%	(−6.8%)
3	44.9%	(+11.6%)	20.6%	(−12.7%)

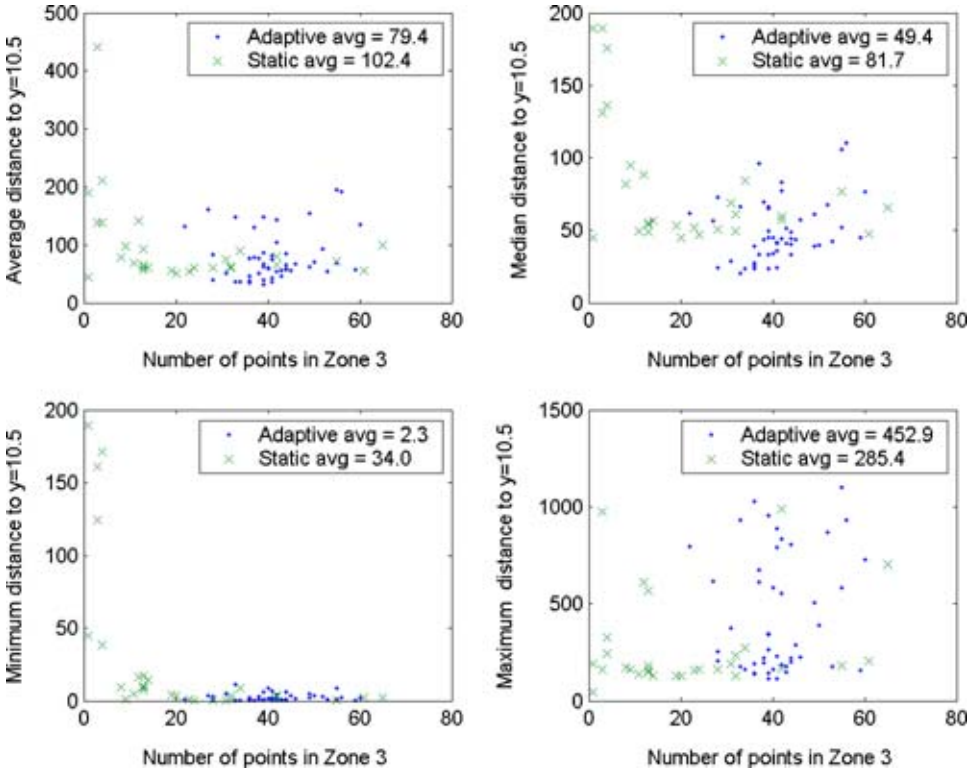

Fig. 8 Summary metrics of the distance (in mm) to the predicted maximum reach envelope for each test subject. Points and x's are from adaptive and static experiments, respectively.

that were, on average, significantly closer to the predicted maximum reach envelope (see the average and median distance plots in Fig. 8). Second, the adaptive data produced better results using a smaller data sample—half as many samples in some cases (see Fig. 7). Third, the adaptive experiments provided more outlier points that were far away from the maximum reach envelope (see the maximum distance plot in Fig. 8). One can attribute this to the first several reaches being quite far away from the subject, until enough data were gathered to create a more accurate model.

Figure 8 indicates that the data gathered by adaptive sampling are likely closer to the maximum reach range envelope than data from the static approach. Adaptive sampling has therefore provided the researchers with more accurate results and met experimental goal 2c.

Based on these results, it is clear that adaptive sampling has gathered a different data set for each subject, and therefore satisfied the first experimental goal, to "adapt to the test subject." During the testing, the inclusion of constraints in superEGO prevented sample points from being either within the encroachment zone or extremely far away. It therefore satisfied experimental goal 3. Finally, the testing was limited to 1 hour, satisfying the fourth experimental objective.

Conclusions and Future Work

The main goal of this research was to demonstrate that superEGO was well-suited to creating adaptive experimental designs. The generality and flexibility of the algorithm lent itself well to using problem-specific sampling criteria and constraints. It has been shown that the adaptive sampling successfully created experimental designs that adapted to the individual test subjects. The kriging models used to predict the reach difficulty of the subject enabled the protocol to incorporate information from previous data samples when deciding where to sample next.

The integration of constraints into the solution of the ISC maximization problem allowed for irregularly shaped design spaces that took subject behavior into account. While it may not always be necessary to avoid sampling infeasible points, it was safety critical for this experiment to do so. SuperEGO was able to satisfy this demand by evaluating the encroachment zone constraint directly at each iteration. In addition, the ability to create smoothed kriging models was shown to be an important aid to the case study.

The adaptive methods described here were quite efficient. The original methods used required 1.5–3 h of testing for each subject. By using adaptive sampling to select a more efficient DOE, that time has been reduced to 1 h, including the time superEGO spent calculating the next sample location. As described above, this time savings came without any sacrifice in the usefulness of the data generated. In fact, the data was shown to have a number of advantages over the static methods originally used. Overall, the sample size was reduced, but the quality of the data was improved.

When compared to static experiments, the adaptive designs provided data that was more spread out throughout the design space and led to a somewhat better balance of samples in the three reach zones. They also tended to place sample points closer to the predicted maximum reach envelope, thereby narrowing the uncertainty in its location.

Overall, superEGO performed very well at meeting the experimental goals. The shortcomings of using a static DOE have been quantified. It would be interesting future research to see how well other optimization algorithms can meet the challenges presented by this experiment. In addition, the impact of the initial sample set could be further investigated. Researching the number of initial sample points taken and their locations was beyond the scope of this paper.

Further work could also be done to enhance the abilities of the smoothed kriging models. Rather than having a uniformly smoothed model, one could incorporate a measurement error that is a function of the predicted value of the model. In this way, one could account for the fact that the subjects are more consistent when assigning very high or very low reach difficulty values than

they are in assigning values in the medium range. The kriging model could potentially include more smoothing in the region of higher intrasubject variability and remain more true to the data at the extremes.

Acknowledgments

This research has been partially supported by the Automotive Research Center at the University of Michigan, a U.S. Army Center of Excellence in Modeling and Simulation of Ground Vehicles, by the General Motors Collaborative Research Laboratory at the University of Michigan, and by the HUMOSIM Laboratory. This support is gratefully acknowledged. We also wish to thank the reviewers for their insightful and constructive feedback.

References

[1] Montgomery, D. C., 2001, *Design and Analysis of Experiments*, 5th ed., Wiley, New York.
[2] Atkinson, A., Hackl, P., and Müller, W., 2001, in *Proceedings of MODA 6: Model Oriented Data Analysis*, Physica Verlag, Heidelberg, Germany.
[3] Atkinson, A., Pronzato, L., and Wynn, H. P., 1998, in *Proceedings of MODA 5: Advances in Model-oriented Data Analysis and Experimental Design*, Physica Verlag, Heidelberg, Germany.
[4] Flournoy, N., Rosenberger, W. F., and Wong, W. K., 1997, *New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Research Conference*, Institute of Mathematical Statistics, Hayward, CA.
[5] Thompson, S. K., and Seber, G. A. F., 1996, *Adaptive Sampling*, Wiley, New York.
[6] Zacks, S., 1996, "Adaptive Designs for Parametric Models," *Handbook of Statistics*, Vol. 13, Ghosh, S. and Rao, C. R., eds., NY, Chap. 5.
[7] Hardwick, J. P., and Stout, Q. F., 1998, "Flexible Algorithms for Creating and Analyzing Adaptive Sampling Procedures," *New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Conference*, Flourny, N., Rosenberger, W. F., and Wong, W. K., eds., Institute of Mathematical Statistics, Hayward, CA, pp. 91–105.
[8] Jin, R., Chen, W., and Sudjianto, A., 2003, "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments," *ASME Des. Automation Conf.*, ASME Paper No. DETC-DAC48760.
[9] Jin, R., Chen, W., and Sudjianto, A., 2002, "On Sequential Sampling for Global Metamodeling in Engineering Design," *ASME Des. Automation Conf.*, ASME Paper No. DETC-DAC34092.
[10] Jones, D. R., 2001, "A Taxonomy of Global Optimization Methods Based on Response Surfaces," J. Global Optim., **21**, pp. 345–383.
[11] Pacheco, J. E., Amon, C. H., and Finger, S., 2003, "Bayesian Surrogates Applied to Conceptual Stages of the Engineering Design Process," J. Mech. Des., **125**(4), pp. 664–672.
[12] Prez, V. M., Renaud, J. E., and Watson, L. T., 2001, "Adaptive Experimental Design for Construction of Response Surface Approximations," AIAA J., **40**(12), pp. 2495–2503.
[13] Sasena, M. J., 2002, "Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations," Ph.D. thesis, University of Michigan.
[14] Simpson, T., Dennis, L., and Chen, W., 2002, "Sampling Strategies for Computer Experiments: Design and Analysis," Int. J. Reliab. Appl., **2**(3), pp. 209–240.
[15] Torczon, V., and Trosset, M. W., 1998, "Using Approximation to Accelerate Engineering Design Optimization," in *Proceedings of the 7th AAIA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Paper AIAA-98-4800.
[16] Sasena, M. J., Papalambros, P. Y., and Goovaerts, P., 2002, "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization," Eng. Optimiz., **34**(3), pp. 263–278.
[17] Sasena, M. J., Papalambros, P. Y., and Goovaerts, P., 2001, "The Use of Surrogate Modeling Algorithms to Exploit Disparities in Function Computation Time Within Simulation-based Optimization," *The Fourth World Congress of Structural and Multidisciplinary Optimization*, Dalian, China.
[18] Schonlau, M., 1997, "Computer Experiments and Global Optimization," Ph.D. thesis, University of Waterloo, Waterloo, Canada.
[19] Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.
[20] Jones, D. R., Perttunen, C. D., and Stuckman, B. E., 1993, "Lipschitzian Optimization Without the Lipschitz Constant," J. Optim. Theory Appl., **79**(1), pp. 157–181.
[21] Jones, D. R., 2001, "The DIRECT Global Optimization Algorithm," *Encyclopedia of Optimization*, Vol. 1, Floudas, C. A. and Pardalos, P. M., eds., pp. 431–440.
[22] Sacks, J., Schiller, S. B., and Welch, W. J., 1989, "Design for Computer Experiments," Technometrics, **31**, pp. 41–47.
[23] Sacks, J., Welch, W. J., Mitchell, W. J., and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Stat. Sci., **4**(4), pp. 409–435.
[24] Sasena, M. J., Parkinson, M., Reed, M., Papalambros, P. Y., and Goovaerts, P., 2002, "Adaptive Experimental Design Applied to an Ergonomics Testing Procedure," *ASME DETC Conf.*, ASME Paper No. DETC2002/DAC-34091.