Knowledge-Based Interval Modeling Method for Efficient Global Optimization

Dongzhe Yang, Kourosh Danai, and Sundar Krishnamurty*

Department of Mechanical and Industrial Engineering University of Massachusetts
Amherst, Massachusetts

Received: November 2, 2000; Accepted: March 3, 2001

Abstract: A Knowledge-Based Interval Modeling (KBIM) Method is introduced for empirical search of the global optimum. The method takes advantage of the *a priori* knowledge of the system in terms of the linear/nonlinear, monotonic/nonmonotonic, sensitivity information between the objective function/constraints and the system variables and incorporates it in an interval model. This enables the KBIM Method to uniquely interleave model-building and model-refinement in the optimal search process, which makes it ideally suited to cases where exact input-output relationship is not defined. The efficiency of the KBIM Method is enhanced by an on-line learning scheme which improves the accuracy of the interval model after each search iteration by comparing the estimated range of its outputs with the actual outputs. The KBIM method has several advantages over conventional empirical search methods: (1) the interval model provides a generic form of representation for linear and nonlinear problems alike; therefore, there is no need for selecting the form of the empirical model through trial and error, (2) the use of *a priori* knowledge in modeling eliminates the need for initial trials to construct an empirical model, so from the beginning a plausible region can be identified within the input-space as the basis of search for the global optimum, (3) the use of intervals relaxes the need for precise information, so there is less demand for exploration within the input-space, and (4) the identification of a plausible region early on focuses the search within the plausible region, leading to a more complete model of this region by using the input/output data from the search for learning.

Key Words: Knowledge-Based • Global Optimization • *a priori* knowledge • interval model.

I. INTRODUCTION

Efficient search of the global optimum is a continuing objective in optimization. Conventional optimization methods, such as Genetic Algorithms, Simulated Annealing, or Random Search may be used to seek the global optimum [1-4]. These methods rely exclusively on evaluation of the objective function and constraints based on a well-established input-output system model, and guide the search based on certain heuristic concepts, such as evolution in Genetic Algorithms, thermodynamics in Simulated Annealing, or probability in Random Search. The Bounding Method, another method used in global optimization, searches for the optimum by partitioning the search space into smaller subregions and producing a bound for the range of values of the objective function/constraints over each subregion. It then rejects those subregions which definitely cannot contain a global optimum [5]. Although this method is guaranteed to find the global optimum given sufficient computing time and memory, it typically relies on the availability of the analytical models

* Corresponding author.

1058-8337/01/$.50

of the objective function/constraints for its search. For many optimization problems, however, input-output system models representing objective function/constraints either cannot be defined or are computationally prohibitive. Although high fidelity computational models (i.e., FEA codes) can be employed in place of analytical models in many disciplines, the high cost of computation associated with them prevents them from being used in global optimization procedures [6].

In the absence of analytical input-output system models, empirical models of the objective function/constraints may be developed off-line using statistical methods such as design of experiment (DOE) methods [7-9]. The main drawbacks are the high cost of experimentation and the difficulty in determining an appropriate form for the model. In cases where the degree of nonlinearity of the problem is unknown, an inappropriate model may lead to misrepresentation and inefficiency of the search. In order to relax the need for global modeling of the search space, model management or move limit strategies may be employed to develop an empirical model within a trust region around the current search point [6,10]. Although local modeling of the trust region requires a smaller number of experiments than global modeling, the total number of experiments has been shown to be similar due to the need for modeling several regions [6]. Empirical modeling may also be performed on-line using the data that become available during the search [11]. But, it has the same limitations as off-line modeling.

In this paper, the Knowledge-Based Interval Modeling (KBIM) Method is presented to enhance the efficiency of empirical search in several ways. The interval model that represents the objective function/constraints has a generic form, so there is no need for trials to determine the form of the model. Furthermore, this model provides the framework to incorporate linear/nonlinear and monotonic/nonmonotonic sensitivity information between the objective function/constraints and the system variables in its initial definition. As such, it requires less training. The KBIM Method simultaneously interleaves modeling and search, and continually searches for optimum within the plausible region estimated by the model and provides data for model refinement as well. This leads to a more concentrated modeling effort within the search space and, therefore, a more efficient search of the optimum. In this method, the basic concepts of interval arithmetic [12, 13] are extended to estimate the bounds for the objective function/constraints as the basis for defining the plausible region. Due to the inherently conservative nature of the estimated bounds, the estimated plausible region is likely to continually encompass the actual plausible region of the global optimum such that the search is less likely to be trapped at local optima.

II. THE KBIM METHOD

The *a priori* knowledge of a system may include causal relationships between the inputs and outputs. This knowledge may also contain the level of each input's effect on various outputs. In this work, an interval model is introduced to represent such knowledge in the form of an interval model as follows [12]:

$$\Delta \vec{y}_j(k) = \vec{C}_{1j}(k)\Delta x_1(k) + \vec{C}_{2j}(k)\Delta x_2(k) + \cdots + \vec{C}_{nj}(k)\Delta x_n(k), \quad j = 1, \cdots, m \qquad (1)$$

where each coefficient denoted with a left-right arrow '$\leftrightarrow$' is defined as an interval,

$$\vec{C}_{ij}(k) = \left[\vec{C}_{Lij}(k), \vec{C}_{Uij}(k)\right], \quad i = 1, \cdots, n$$

In the above model, $C_{Lij}(k)$ and $C_{Uij}(k)$ represent the current values of the lower and upper bounds of the sensitivity functions between each input $Dx_i(k)$ and output $\Delta \vec{y}_j(k)$. The interval $\Delta \vec{y}_j(k)$ denotes the estimated range of change of the *j*th output resulted from the change to the current inputs [$Dx_1(k)$, ..., $Dx_n(k)$], and k denotes the current search iteration number. The coefficient intervals defined above provide a convenient mechanism for incorporating a variety of qualitative relationships. For example, a monotonically increasing relationship can be represented by the coefficient interval [0,+A], or a monotonically decreasing relationship can be represented by the coefficient interval [–A,0], where A is an unknown positive number. A coefficient interval of [–A, A] would indicate an unknown causal relationship, and the coefficient interval [0,0] would indicate a null effect. Furthermore, if the relative effect of two inputs is known, then their upper limits can be defined relative to each other. For example, if the effect of one input on an output is known to be at least twice the effect of another input, then the coefficient interval of this input can be set as [0, +2A], defining its upper limit as twice the upper limit of another output having a coefficient interval [0,+A]. While the above coefficient intervals can only define qualitative relationships between the inputs and outputs, they do not exclude quantitative knowledge. For example, if the effect of a design variable such as cross sectional area on a design objective such as deflection is between

–1 and –2 (mm/mm²), the coefficient interval for the cross section area can be set as [–1,–2].

The fit provided by the interval model for a wildly nonlinear input/ouput relationship is illustrated in Figure 1, when the output range is estimated relative to one reference input. According to Eq. 1, the estimated range of the output becomes larger, and therefore less accurate, as the potential input is selected farther from the current input. This potential drawback of the interval model is eliminated after several inputs are explored and output ranges are estimated. Note that although the output range is quite large at points far from the current input in Figure 1, it still encompasses the actual output, mainly due to its compliance with the actual input-output relationship and its conservatively defined coefficient interval.

The range of outputs resulting from changes to the present inputs can be estimated using the interval model, according to the following interval analysis computation rules [12]:

$$\tilde{C}_{ij}\,\Delta x_1 = \left[\min\left\{C_{Lij}\Delta x_i, C_{Uij}\Delta x_i\right\}, \max\left\{C_{Lij}\Delta x_i, C_{Uij}\Delta x_i\right\}\right] \quad (2)$$

$$\tilde{C}_{ij}\,\Delta x_1 + R = \left[\min\left\{C_{Lij}\Delta x_i, C_{Uij}\Delta x_i\right\} + R, \max\left\{C_{Lij}\Delta x_i, C_{Uij}\Delta x_i\right\} + R\right] \quad (3)$$

where R denotes a constant, and

$$\tilde{C}_{1j}\,\Delta x_1 + \tilde{C}_{2j}\Delta x_2 = \left[\min\left\{C_{L1j}\Delta x_1, C_{U1j}\Delta x_1\right\} + \left[\min\left\{C_{L2j}\Delta x_2, C_{U2j}\Delta x_2\right\}, \max\left\{C_{L1j}\Delta x_1, C_{U1j}\Delta x_1\right\} + \max\left\{C_{L2j}\Delta x_2, C_{U2j}\Delta x_2\right\}\right]\right. \quad (4)$$

Using these computation rules, the estimated interval of the output, $\tilde{y}_j(k)$, or any set of inputs $[x_1, x_2,..., x_n]$ can be computed relative to the explored input $[x_1(k), x_2(k),..., x_n(k)]$ and its output $y_j(k)$ as

$$\begin{aligned}
\tilde{y}_j(k) &= y_j(k) + \Delta\tilde{y}_j(k) \\
&= y_j(k) + \tilde{C}_{1j}(k)\Delta x_1(k) + \cdots + \tilde{C}_{nj}\,\Delta x_n(k) \\
&= \left[y_j(k) + \min\left\{C_{L1j}\Delta x_1(k), C_{U1j}\Delta x_1(k)\right\} + \cdots\right. \\
&\quad + \min\left\{C_{Lnj}\Delta x_n(k), C_{Unj}\Delta x_n(k)\right\}, \\
&\quad y_j(k) + \max\left\{C_{L1j}\Delta x_1(k), C_{U1j}\Delta x_1(k)\right\} + \cdots \\
&\quad \left. + \max\left\{C_{Lnj}\Delta x_n(k), C_{Unj}\Delta x_n(k)\right\}\right]
\end{aligned} \quad (5)$$

where

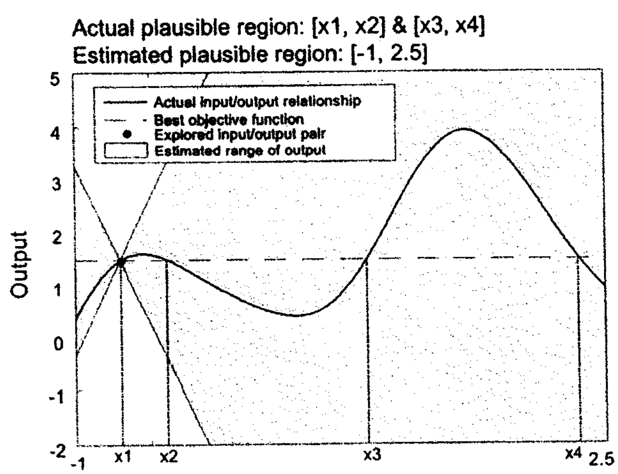

Figure 1. Estimated range of output by the interval model using one input-output pair.

$$\Delta x_i(k)=x_i-x_i(k),\ i=1,\cdots,n \quad (6)$$

The estimated output $\breve{y}_j(k)$ at a candidate input xi may be computed relative to any set of previously explored inputs, yielding different estimates of $\breve{y}_j(k)$ (due to different values of $\Delta x_i(k)$). In order to cope with the multiplicity of estimates, $\breve{y}_j(k)$ is defined here as the common range among all of the $\breve{y}_j(k)$ estimates. The estimation of $\breve{y}_j(k)$ using this commonality rule is illustrated in Figure 2 when $\breve{y}_j(k)$ is estimated relative to three explored inputs. It is clear from Figure 2 that using this estimation approach for $\breve{y}_j(k)$ enables representation of the system nonlinearities in a piecewise fashion. This representation capability of the interval model is illustrated in Figure 3 when several input-output data have been available for estimation purposes. Of course, it should be noted that this representation capability of the interval model will be successful when the model complies with the overall nonlinear input-output relationship. A case where such a compliance does not exist is illustrated in Figure 4, where the lack of commonality between the estimated ranges of output causes a part of the input-output relationship to not be accurately represented by the interval model. Any lack of compliance between the interval model and the input-output relationship will be corrected by adaptation of the coefficient intervals through learning.

A. Learning

Although the *a priori* knowledge may provide a good initial estimate of the plausible region, it may not be able to carry the search process to the end unless it is refined through learning. In the proposed method, the coefficient intervals of the interval model are updated according to the input-output data acquired during search by considering new values for each of the upper and lower limits of individual coefficient intervals. The adaptation of the coefficient intervals may be performed in two steps: enlargement and shrinkage. First, the output ranges for all the search iterations are estimated according to the current interval model. If the estimated output ranges do not include the actual outputs, the upper and lower limits of the first coefficient intervals will be enlarged in small steps to yield an enlarged coefficient interval, as

$$\left[C_{Lij}(k)-\beta, C_{Uij}(k)+\beta\right] \quad (7)$$

where the small steps β can be defined proportional to the current interval length, as

$$\beta=\gamma\left[C_{Uij}(k)-C_{Lij}(k)\right] \quad (8)$$

with the parameterg within the range $\left[\frac{1}{50},\frac{1}{20}\right]$. After each enlargement, the output ranges are re-estimated using the enlarged coefficient. If the newly estimated outputs do not include the actual outputs, the next coefficient will be enlarged, and enlargement of the coefficient intervals will be repeated until the estimated output ranges include all of the actual outputs.

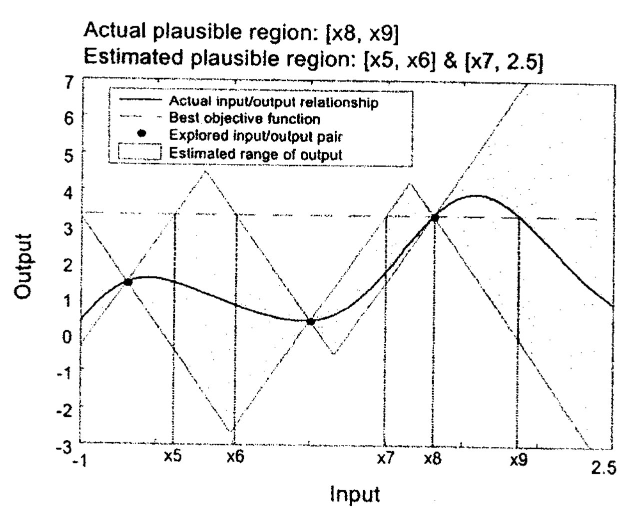

Figure 2. Estimated range of output by the interval model when three input-output pairs are used.

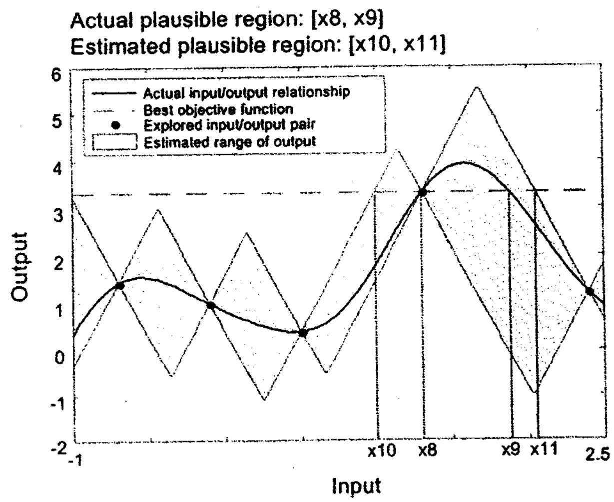

Figure 3. Estimated range of output by the interval model using five input-output pairs.

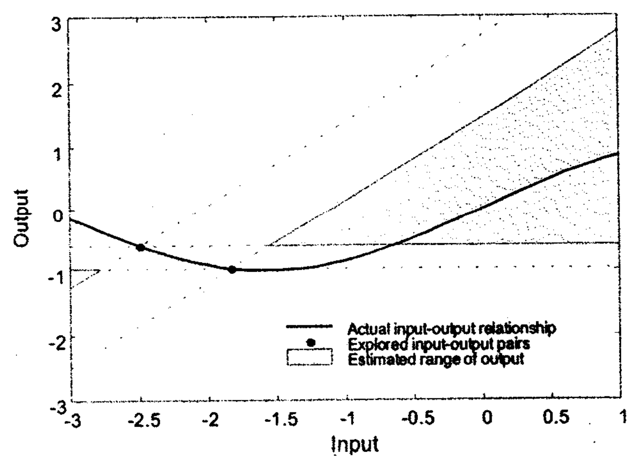

Figure 4. Estimated range of output by a monotonic interval model for a nonmonotonic input-output relationship.

Although the newly updated interval model will fit the input-output data after the enlargement phase, some of the coefficient intervals may be larger than they need to be and may provide largely over-estimated output ranges. In order to rectify this situation, each coefficient interval will be shrunk individually by selecting two new candidates respectively for its upper and lower limits in proportion to the current interval length. The two new candidate coefficient intervals will have the form

$$\left[C_{Lij}(k)+\beta, C_{Uij}(k)\right] \text{and} \left[C_{Lij}(k), +C_{Uij}(k)-\beta\right] \tag{9}$$

and their suitability will be tested by how well they include the actual outputs within the newly estimated ranges. The goodness of fit of the outputs for each candidate interval is quantified by the evaluation function

$$E=\sum_{j=1}^{m}\sum_{q=1}^{k}\left\{\left[y_j(q)-y_{jl}(q)\right]^2+\left[y_{ju}(q)-y_j(q)\right]^2\right\} \tag{10}$$

where m represents the number of outputs, k represents the current number of search iterations, $y_{jl}(q)$ and $y_{ju}(q)$ represent the lower and upper limits of the estimated output ranges, and $y_j(q)$ denotes the actual output. A smaller evaluation function E indicates a more centered set of outputs within the estimated ranges, which is more desirable; therefore, the candidate interval with a smaller value of E is selected. This procedure of shrinking coefficients will be repeated for every coefficient interval in the model. The univariate shrinkage of coefficient intervals will reduce the value of E, however it may not minimize it. The alternative is to perform an optimization of E, but that will demand prohibitively long searches. This shrinkage procedure is repeated in a cyclic fashion until either the estimated output ranges begin to exclude the actual outputs or the intervals violate the minimum length constraint defined as:

$$minL=\left\{C_{Uij}(0)-C_{Lij}(0)\right\}(1-\alpha)^k \tag{11}$$

which is provided to prevent drastic shrinkage of the coefficient intervals. In the above equation, the parameter $\alpha \in [0,1]$ determines how fast the coefficient will be shrunk, and the parameter k denotes the current search iteration. The coefficient interval cannot be shrunk when $\alpha = 0$ and can be shrunk without limits when $\alpha = 1$. With α equal to 0.1, a coefficient interval can be shrunk to 35% of its original length at the 10th iteration and to 12% at the 20th iteration. For conservatively defined initial coefficient intervals, the parameter α can be set to a large value (e.g., between 0.15 and 0.3) in order to have a fast shrinkage rate. On the other hand, for a system with several inputs, the parameter a can be set to a small value (e.g., between 0.05 and 0.15), since a slow shrinkage rate is preferred due to the larger number of data needed for training. For a highly nonlinear system which also needs extensive training, a small value of α is preferred as well. In general, the value of α only affects learning at the beginning of search because once enough data are available, learning is driven more by the data than the value of α. In order to demonstrate the learning algorithm, the interval model used for the estimation in Figure 4 is updated according to the two explored input-output data and the estimated outputs by the updated interval model are shown in Figure 5. The results indicate the effectiveness of the learning algorithm in correcting initially ill-defined models.

B. Search

The search process is guided by selecting the inputs for the next search iteration within the plausible region, which comprises all of the inputs that will produce better objective functions than those at explored inputs during previous search iterations. In order to estimate the plausible region, the estimated ranges $\tilde{y}_j(k)$ of the objective function at N randomly tested inputs within the search space are compared with the best value of the objective function at the previously explored inputs so as to determine whether the inputs belong to the plausible region. The number N is an arbitrary large number that corresponds to the size of the search space. In the case of minimization, when the lower limit of the estimated range of the objective function is smaller than the lowest value of the explored objective functions in the previous search iterations, the corresponding input is considered to be in the plausible region. Similarly for maximization, the upper limit of the estimated range of the objective function needs to be larger than the highest value of the explored objective functions so that the corresponding inputs be considered in the plausible region. Estimation of the plausible region using the upper estimate of the objective function is illustrated in Figures 1–3. As shown in these figures, overestimation of the objective function leads to estimated plausible regions that are larger than the actual plausible region. However, these estimated plausible regions are expected to always encompass the actual plausible region due to the conservative

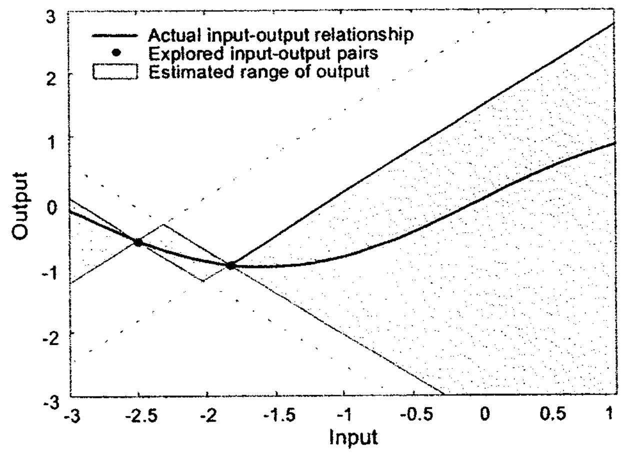

Figure 5. Estimated range of output by the updated interval model for a nonmonotonic input-output relationship.

definition of the interval model. In Figure 3, for example, the estimated plausible region $[x_{10}, x_{11}]$ includes the actual plausible region $[x_8, x_9]$. The results in these figures also indicate that the estimated plausible regions are usually shrunk as more input-output pairs are explored, even when nonplausible inputs are explored. For example, compared to the estimated plausible region in Figure 2, the two additionally explored nonplausible input-output pairs in Figure 3 shrink the estimated plausible region to a more appropriate region in terms of optimization, that is, a region around the global maximum of the function. By only exploring the smaller estimated plausible region, fewer input-output data will be needed for modeling, so the search efficiency will be improved.

For constrained optimization problems, the plausible region should be restricted inside the feasible region. Without loss of generality, all the constraints are assumed to be '≤' inequality constraints with the right hand side set to 0. Therefore, in order to estimate the feasible region, the estimated ranges of each constraint at the N randomly tested inputs are compared with 0 so as to see whether the corresponding inputs belong to the feasible region. If the lower limit of the estimated range of the constraint is smaller than 0, the corresponding input is considered to be in the feasible region. Otherwise, it is not. The above procedure of estimating the feasible region based on a single constraint is then extended to multiple constraints by forming the conjunction of the estimated feasible region from each constraint. For constrained optimization problems, the conjunction of the estimated feasible region and plausible region will be used as the estimated plausible region, with the expectation that the inputs within the estimated plausible region will produce better values of the objective function while satisfying the constraints.

The suitability of the inputs within the plausible region cannot be compared, since only the ranges of the corresponding objective function and constraints can be estimated. Therefore, an input is randomly selected among the estimated plausible inputs for the next search iteration. When no plausible inputs are found within the N random trials, it is decided that the plausible region is too small for any further search to yield a better input. Therefore, the search is considered to have converged and the best input among all the explored inputs is considered as the optimum. The search may also be considered exhausted when a maximal number of search iterations is reached.

III. THE KBIM METHOD

A schematic of the optimization procedure used by the KBIM Method is shown in Figure 6. First, the *a priori* knowledge is used to formulate the interval model representing the objective function and constraints. Then, an initial input is randomly selected within the search space, and its objective function and constraints are evaluated through experiments. Based on the explored input, a plausible region is estimated by the interval model as the basis for selecting the inputs for the second search iteration. Next, the estimated ranges and actual values of the objective function and constraints at the selected inputs are compared for refining the interval models. Again, the updated interval model is used to estimate the plausible region. The above procedure is repeated until the termination criteria is satisfied, i.e., no plausible region is found or the maximal number of search iterations is reached.

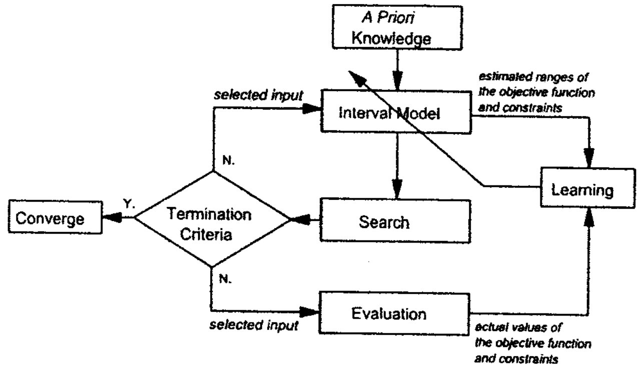

Figure 6. Optimization procedure used by the KBIM Method.

IV. EVALUATION OF THE KBIM METHOD

The performance of the KBIM Method was evaluated in two benchmark problems. The first benchmark is an unconstrained optimization problem with multiple local optima, and the second benchmark represents a constrained optimization problem.

A. Maximization of the Peak Function

The Peak Function has been used as a benchmark for Genetic Algorithms (Jang et al., 1996) [14]. It is a two-input function defined as

$$f(x,y) = -10\left(\frac{x}{5} - x^3 - y^5\right)e^{-x^2-y^2} + 3(1-x)^2 e^{-x^2-(y+1)^2} - \frac{1}{3}e^{-(x+1)^2-y^2} \qquad (12)$$

The surface plot of this function, shown in Figure 7, has three local maxima and two local minima. The objective is to find the global maximum.

The input-space was confined to a square of [-3,3] × [−3,3]. In order to obtain *a priori* knowledge, in the absence of process knowledge conventionally provided by the expert, the objective function was evaluated at ten different inputs within the input-space. The coefficients of the interval model were then defined according to the sensitivity ranges obtained at these inputs. Since ten sample points are found to be adequate in providing an overall representation of a two-input problem, the parameter α in Eq. 11 was set to 1 so that the coefficient intervals could be shrunk without limits during learning. The search number N for each search iteration was set to 20, 000, and altogether 100 search iterations were run.

The Peak Function is known to have a global maximum at $x^* = (-0.0077, 1.5823)$ with the function value $f(x^*) = 8.1062$. The best inputs selected at the 100th iteration by the KBIM Method for five different initial models, which were produced by different sets of initial inputs, are shown in Table 1. The results in Table 1indicate that the inputs are quite close to the optimal inputs, and that the search was successful independent of the initial coefficient intervals.

The inputs selected by the KBIM Method during one of these trials are shown in Figure 8, which include the ten inputs used for defining the *a priori* knowledge. The variations of the inputs at the beginning of the search are due to the plausible region being relatively large. However, as more input-output data becomes available to better estimate the plausible region, the estimated plausible region becomes smaller and the variations of the inputs decrease. After the 55th iteration, both inputs stay within a small region around the global maximum, indicating that the estimated plausible region has converged around the global maximum. The actual values and upper estimates of the Peak Function at these selected inputs are shown in Figure 9, verifying the convergence of the inputs towards the optimum after the 55th iteration. The results in Figure 9 also indicate that the upper estimates are mostly larger than the actual values. Given that the plausible region used for search by the KBIM Method is defined based on these upper estimates, these overestimated Peak Function values translate to overestimated plausible regions during search iterations, providing safeguards against entrappment in local maxima (minima).

To evaluate the significance of various learning parameters, the two coefficient intervals were initially set to [−20, 20], with the starting inputs set at (−2.4657, 2.7325). Altogether 50 iterations were run, and in order to eliminate the randomness of the search,

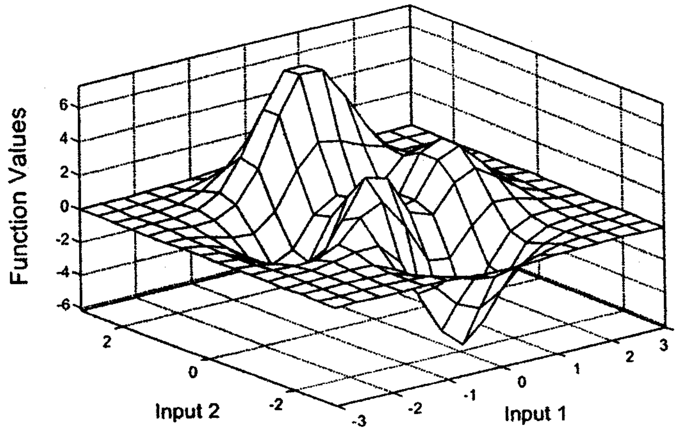

Figure 7. Surface plot of the Peak Function.

Table 1. The optimal inputs and best Peak Function values obtained by using different initial coefficient intervals.

Run No.	No. of Iterations	Best Inputs	Peak Function Value
1	100	(-0.005, 1.584)	8.1059
2	100	(0.005, 1.575)	8.1039
3	100	(-0.012, 1.579)	8.1061
4	100	(-0.009, 1.588)	8.1054
5	100	(-0.019, 1.583)	8.1054

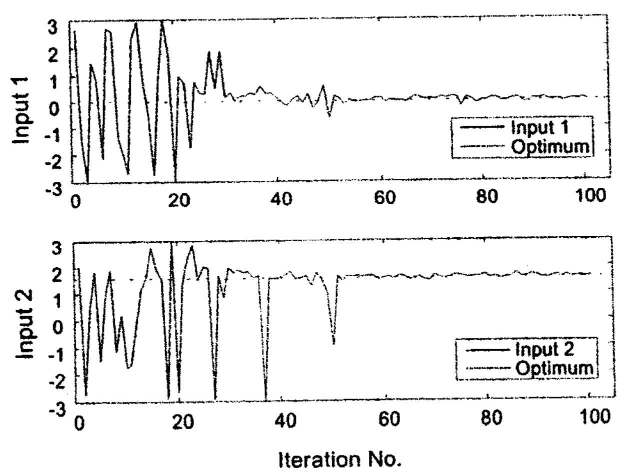

Figure 8. The selected inputs by the KBIM Method in the Peak Function maximization.

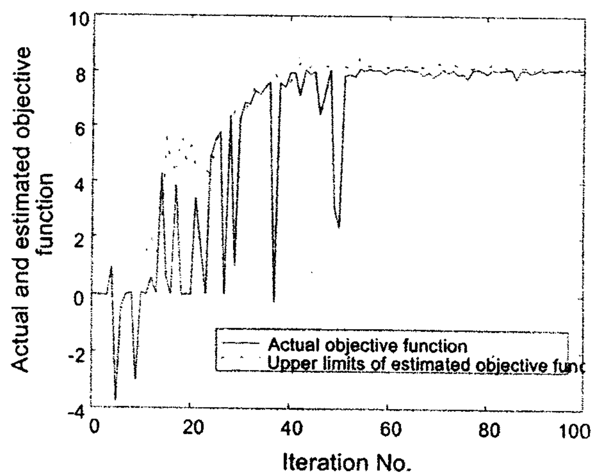

Figure 9. Actual and estimated Peak Function values at the selected inputs.

the search procedure was repeated ten times with different α values, and the Peak Function values obtained from these runs were averaged at each iteration. The averaged values of the Peak Function obtained with four different a values are shown in Figure 10. For the case with a equal to 0.01, the coefficient intervals have the slowest shrinkage rate. As the result, the coefficient intervals are kept largerthan-necessary and the plausible region is kept overestimated throughout the search. The overestimated plausible region resulted from this small α value hinders convergence of the inputs toward the global optimum within the 50 iterations (see Figure 10). However, for the case with α equal to 0.1, 0.2, or 0.3, the results in Figure 10 show that much better Peak Function values were found, indicating better convergence. The similar performance of the search with these different α values indicates the relative robustness of the KBIM Method to α values within this range. To further provide an indication of the effect of learning on the interval model, a sample of the two coefficient intervals at three different stages of the search are shown in Table 2. When the coefficients are defined according to only ten input-output data they have a relatively small range, but as they are adapted based on a larger set of input-output data they contain a larger range to reflect the nonlinearity of the objective function as reflected by a more encompassing input-output data.

The speed of convergence of the KBIM Method for the Peak Function can be best evaluated relative to a Genetic Algorithm (GA) [14]. In the GA implemen-tation, the same input-space was used and each generation contained 20 inputs. Each input's fitness was defined as the value of the Peak Function minus the minimum function value across the population. This guaranteed that all fitness values were nonnegative. A simple one-point crossover scheme was used with the crossover rate equal to 1.0. Furthermore, uniform mutation was utilized with the mutation rate equal to 0.01, and the best two inputs across current generation were kept for the next generation. Altogether, 30 generations were run, which would be equivalent to 600 iterations of the KBIM method. Based on the number of evaluations of the objective function and constraints (see Table 3), the KBIM Method used much less number of evaluations of the objective function and found a better set of inputs than the Genetic Algorithm.

B. Barnes Problem

The Barnes problem is a constrained optimization problem [15], so it represents a different format for testing the KBIM Method. This problem has two bounded variables (x_1, x_2) and five inequality constraints as follows,

$$
\begin{aligned}
&Minimize\, f(x) = -75.1964 + 3.81x_1 - 0.1269x_1^2 \\
&+2.0568\times10^{-3}x_1^3 - 1.0345\times10^{-5}x_1^4 + 6.831x_2 - 0.0302x_1x_2 \\
&+1.281^{-3}x_2x_1^2 - 3.5256\times10^{-5}x_2x_1^3 + 2.266\times10^{-7}x_2x_1^4 \\
&-0.25646x_2^2 + 3.4604\times10^{-3}x_2^3 - 1.3514\times10^{-5}x_2^4 + \frac{28.106}{x_2+1} \\
&+5.2375\times10^{-6}x_1^2x_2^2 - 7\times10^{-10}x_1^3x_2^3 + 6.3\times10^{-9}x_1^3x_2^2 \\
&-3.40546\times10^{-4}x_1x_2^2 + 1.6638\times10^{-6}x_1x_2^3 + 2.8673e^{0.0005x_1x_2}
\end{aligned}
\tag{13}
$$

subject to:

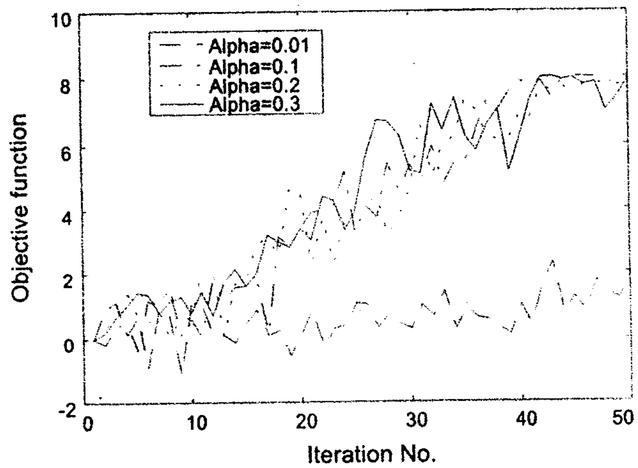

Figure 10. Averaged Peak Function values obtained by using different α values.

Table 2. The coefficient intervals of the Peak Function at different search iterations.

Iteration	$\overleftrightarrow{C}_1$	$\overleftrightarrow{C}_2$
initial	[-1.8790,1.7567]	[-2.5137,1.2152]
50th	[-3.2565,6.2086]	[-5.2254, 6.9820]
100th	[-3.1545,5.8953]	[-5.3293,6.9486]

Table 3. Comparison of the application results of the KBIM Method and Genetic Algorithm in the Peak Function maximization.

	KBIM	GA	Global Optimum
Evaluation number	100	600 (20 × 30)	
Optimal input	(-0.0139,1.5842)	(0.005,1.525)	(-0.008,1.582)
Objective function	8.1059	8.0512	8.1062

$$700 - x_1 x_2 \le 0$$
$$5x_1^2 - 625x_2 \le 0$$
$$5(x_1 - 55) - (x2 - 50)^2 \le 0$$
$$20(x1 - 45) - 30(x_2 - 45) \le 0$$
$$40(x2 - 40) - 25(x_1 - 35) \le 0$$

where $0 \le x_1 \le 75$ and $0 \le x_2 \le 65$.

Six interval models are required to represent the objective function and its five constraints. In order to define the knowledge necessary for defining the initial model, the objective function was evaluated at ten different inputs within the input-space. The coefficient intervals were then defined according to the sensitivity ranges obtained at these inputs. Since ten sample points are adequate for providing a comprehensive representation of a two-input problem, the parameter α in Eq. 11 was set to 1 so that the coefficient intervals could be shrunk without limit. The number N was set to 20,000.

The Barnes problem has a global minimum at x* = (75,65), f(x*) = –58.7279, with the fourth and fifth constraints active at the optimum. The KBIM Method was tested in search of the optimum using different initial models, generated by different samples. Each trial was ended when no more plausible inputs could be found. The best inputs from five of those trials are shown in Table 4. The results indicate that the KBIM Method completed the search in at most 40 iterations, and that it found solutions very close to the actual solution. The inputs during one of these trials are shown in Figure 11, with the actual and lower estimates of the objective function shown in Figure 12. In Figure 12, '*' indicates feasibility, and 'o' signifies infeasibility. Both the inputs and objective function values indicate the convergence tendency of the method. It is also shown in Figure 12 that several input sets selected during the search were infeasible, and that the lower estimates of the objective function were mostly smaller than the actual values. Given that the plausible region used for search by the KBIM Method is defined based on these lower estimates, the underestimated objective function values translate to overestimated plausible regions during search iterations. These overestimated plausible regions are the cause of infeasible inputs selected during the search.

Table 4. The optimal inputs and best objective function values obtained by using different initial coefficient intervals.

Run No.	No. of Iterations	Best Inputs	Obj. Function Value
1	30	(74.409, 64.624)	-58.1963
2	34	(74.230, 64.483)	-58.1589
3	40	(74.799, 64.868)	-58.5497
4	34	(74.420, 64.630)	-58.2067
5	38	(74.486, 64.671)	-58.2669

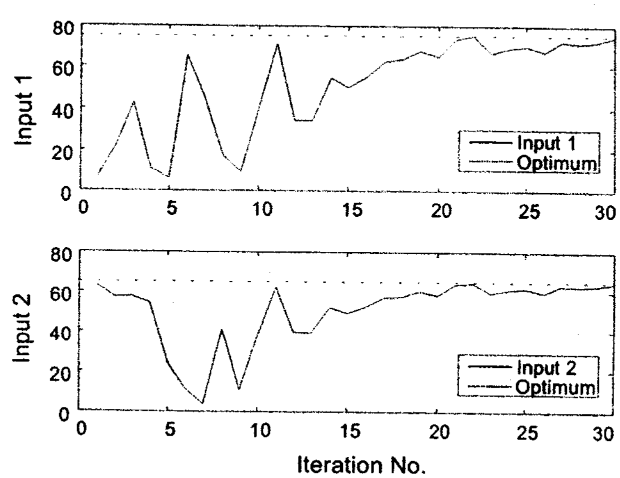

Figure 11. The selected inputs by the KBIM Method in the Barnes problem.

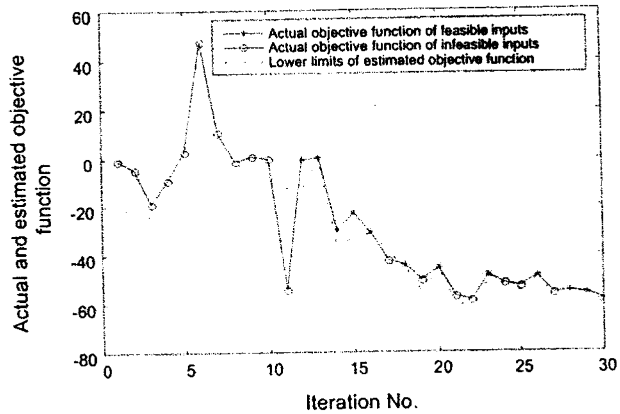

Figure 12. The actual and estimated objective function at the selected inputs in the Barnes problem.

The application results of the KBIM Method in the two benchmark problems indicate that the interval model could effectively estimate the objective function and constraints to guide the search toward the global optimum. The efficiency of the KBIM Method is attributed to two-factors: (1) the dynamic nature of the method which interleaves modeling and search, and (2) incorporation of learning that improves the interval model. Furthermore, an estimated plausible region larger than the actual plausible region ensures convergence to the global optimum, and provides safeguard against entrapment at a local optimum.

SUMMARY

A novel Knowledge-Based Interval Modeling (KBIM) Method is introduced for global optimization. It incorporates the *a priori* knowledge of a system in the form of an interval model, estimates bounds for the objective function and constraints, and defines a plausible region based on the estimated bounds for seeking the global optimum. Learning is also utilized to adapt the interval model after each search iteration. The performance of this method is tested in two benchmark problems. The results indicate the overall effectiveness of the method, and its robustness and computational efficiency.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant No. DMI-9700288).

REFERENCES

1. Aarts, E. H. L., and Korst, J., 1989, *Simulated Annealing and Boltzmann Machines : a Stochastic Approach to Combinatorial Optimization and Neural Computing*, Wiley, New York.
2. Rao, S. S., 1996, *Engineering Optimization: Theory and Practice*, John Wiley & Sons, Inc., 3rd ed.
3. Gen, M., and Cheng, R., 2000, *Genetic algorithms and engineering optimization*, Wiley, New York.
4. Mitchell, T. M., 1997, *Machine Learning*, McGraw-Hill, Boston, MA.
5. Leclerc, A. P., 1992, "Efficient and Reliable Global Optimization," Ph.D. thesis, Ohio State Univeristy.
6. Rodriguez, J. F., Renaud, J. E., and Watson, L. T., 1998, "Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization," ASME J. of Mechanical Design, **120**, pp. 58—66.
7. Neter, J., Kutner, H. M., Nachtsheim, C. J., and Wasserman, W., 1996, *Applied Linear Statistical Models*, IRWIN, 4th ed.
8. Montgomery, D. C., 1997, *Design and Analysis of Experiments*, New York: Wiley.
9. Atkinson, A. C., and Donev, A. N., 1992, *Optimum Experimental Designs*, Clarendon Press, Oxford.
10. Wujek, B. A., Renaud, J. E., and Batill, S. M., 1997, "A Concurrent Engineering Approach for Multidisciplinary Design in a Distributed Computing Environment," *Multidisciplinary Design Optimization: State-of-the-Art*, N. Alexamdrov and M. Y. Hussaini, ed., SIAM, Philadelphia.
11. Ivester, R., and Danai, K., 1998, *Tuning and Automatic Regulation of Injection Molding by the Virtual Search Method, ASME J. of Manufacturing Science and Eng.*, **120**, pp. 323—329.
12. Moore, R. E., 1979, *Methods and Applications of Interval Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.
13. Neumaier, A., 1990, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge.
14. Jang, J. R., Sun, C. T., and Mizutani, E., 1996, *Neuro-Fuzzy and Soft Computing: a Computational Approach*

to Learning and Machine Intelligence, Prentice Hall, Upper Saddel River, NJ.
15. Himmelblau, D. M., 1972, *Applied nonlinear programming*, McGraw-Hill, New York.

BIOGRAPHIES

Dongzhe Yang is currently working in the Artificial Intelligence Group at iNautix Technologies Inc., a CSFB company. He joined the company as an engineer in 2000. His research interests include machine learning, knowledge-based systems, data mining, empirical modeling, global optimization, and intelligent control. Dr. Yang received his BS (1992) and MS (1995) from Xi'an Jiaotong University, China, and his PhD (2001) in Mechanical Engineering from the University of Massachusetts, Amherst.

Kourosh Danai joined the faculty of Mechanical Engineering at the University of Massachusetts, Amherst in 1987. He was promoted to the rank of associate professor with tenure in Sept. 1993, and to professor in Sept. 1998. Dr. Danai's research is in manufacturing automation and fault diagnosis. Dr. Danai has been the recipient of three innovation awards from NASA. He spent the summer of 1990 at Sikorsky Aircraft Company and the fall of 1994 at the United Technologies Research Center, and organized two symposia on 'control of manufacturing processes' at the 1990 and 1991 ASME Winter Annual Meeting, and a symposium on 'applied fault diagnosis' at the 1994 ACC. Dr. Danai has been an associate editor of the *ASME Journal of Manufacturing Science and Engineering*, has chaired the Manufacturing Systems Panel of the ASME Dynamic Systems and Control Division, and was program chair of the ASME Dynamic Systems and Control Division at the 1996 Int'l Mechanical Engineering Congress and Exposition. He received his BS (1980), MS (1982), and Ph.D. (1986) all in mechanical engineering from the Univ. of Michigan, Ann Arbor.

Sundar Krishnamurty is currently the Associate Department Head of Mechanical and Industrial Engineering at the University of Massachusetts, Amherst. He joined the faculty as an assistant professor in 1989 and was promoted to the rank of associate professor with tenure in 1995. His research is in the area of mechanical design and focused on topics related to decision-based design, design optimization, visualization and spatial reasoning in kinematics, and design of assistive technology devices. Dr. Krishnamurty received his BS (1982) from IIT-Kanpur; MS (1984) from the University of Pennsylvania; and PhD (1989) in Mechanical Engineering from the University of Wisconsin.