A Stochastic Probing Algorithm for Global Optimization

PURUSHOTTAM W. LAUD
Northern Illinois University, Dekalb, Illinois, U.S.A.

and

L. MARK BERLINER and PREM K. GOEL
The Ohio State University, Columbus, Ohio, U.S.A.

(Received: 3 July 1991; accepted: 22 October 1991)

Abstract. Recently, simulated annealing methods have proven to be a valuable tool for global optimization. We propose a new stochastic method for locating the global optimum of a function. The proposed method begins with the subjective specification of a probing distribution. The objective function is evaluated at a few points sampled from this distribution, which is then updated using the collected information. The updating mechanism is based on the entropy of a move selecting distribution and is loosely connected to some notions in statistical thermodynamics. Examples of the use of the proposed method are presented. These indicate its superior performance as compared with simulated annealing. Preliminary considerations in applying the method to discrete problems are discussed.

Key words. Simulated annealing, Gibbs' distribution, entropy, Bayesian analysis.

1. Introduction

Stochastic approaches to global optimization have recently received considerable attention in various literatures. Since the pioneering work of Cerny (1985), Geman and Geman (1984), Kirkpatrick *et al.* (1983), and Pincus (1970), much of this attention has been devoted to the theory and application of simulated annealing as an optimization technique. Some useful references are Bélisle *et al.* (1990), Pronzato *et al.* (1984), Rinnooy Kan *et al.* (1985), and van Laarhoven and Aarts (1987). Of special interest to statisticians, Bohachevsky *et al.* (1986) and Haines (1987) have applied annealing to experimental design and Geman and Geman (1984) and Geman and McClure (1985) employed simulated annealing for Bayesian image restoration.

We will describe a new technique for stochastic search for optima. The method is related to simulated annealing, especially the "generalized" approach of Bohachevsky *et al.* (1986). The method attempts to incorporate problem specific information to direct the search. Preliminary results indicate that the method can be quite successful, and in some cases, more efficient than generalized simulated annealing.

Journal of Global Optimization **2**: 209–224, 1992.

1.1. STOCHASTIC SEARCH FOR OPTIMA

The problem considered here is as follows: Minimize a function ϕ over a "feasible region", $\mathscr{X}$. It is assumed that $\mathscr{X}$ is a subset of a Euclidean space.

Stochastic search is typically considered in cases when ϕ and/or $\mathscr{X}$ are "complex". The character of complexity for such problems can take on various forms which render standard differentiation based techniques infeasible. First, ϕ may not be differentiable and/or $\mathscr{X}$ may be discrete. Even when such obvious impediments do not arise, standard techniques may not be practical for cases in which ϕ has many local extrema and/or χ is high dimensional.

In problems of these types, it is typically the case that random searchers are more efficient than exhaustive, deterministic searches. For further discussions and references the reader is referred to Aarts and Korst (1989) and Ripley (1987), Chapter 7.

A slight over-simplification of stochastic searches is that such methods are constructed from two basic inputs: (i) Search Recipe and (ii) Stopping Rule. A search recipe is the method by which a sequence of random points or candidates is obtained in $\mathscr{X}$. The stopping rule is the rule which determines how the search is stopped and the global optima of ϕ is guessed.

1.2. SIMULATED ANNEALING

This technique is based on an analogy between the optimization problem and statistical physics. Only a brief description of the analogy is given here. First, imagine a physical system, for example, a large number of molecules contained in a box. Under thermoequilibrium, i.e., the box has been at a constant temperature, T, for a while, the distribution of states of the system is the Gibbs' distribution:

$$\Pr(\text{state}) = \frac{1}{Z} e^{-E(\text{state})/KT},$$

where K is Boltzman's constant, $E(\cdot)$ is the appropriate energy function, and Z is just the normalizer (or "partition function"). Note that Gibbs' distribution favors small energy levels. Next, the limiting distribution as T tends to zero is of interest. In particular, the limiting distribution concentrates all of its mass on those states which yield the (global) minimum energy. (This is the Third Law of Thermodynamics.) For the optimization problem, ϕ is equated with the energy E, points in $\mathscr{X}$ are considered to be "states", and T is viewed as a control parameter.

The standard simulated annealing algorithm is described in two basic components.

Component 1. Search. Generate a sequence of candidates in $\mathscr{X}$ as follows. Start at any point, say x in $\mathscr{X}$. Using a specified *generation distribution*, choose a random

local perturbation of x, say x^*. Note that x^* is a random point "near" x. What is actually meant by near depends on the structure of $\mathscr{X}$. Once nearness is quantified, say by a parameter, Δr, it is treated as a control parameter. The search then chooses to move to x^* or remain at x for the next iteration according to the *move selection distribution* described as:

$$\text{Accept } x^* \text{ with probability } \min[1, \exp\{-\beta(\phi(x^*) - \phi(x))\}]. \tag{1.1}$$

In this distribution β is a control parameter. Note that x^* is accepted if $\phi(x^*) \leq \phi(x)$. If $\phi(x^*) > \phi(x)$, x^* may be accepted with a probability related to the likelihood ratio comparison, under a Gibbs model, of x and x^*. Heuristically, the possibility of accepting a worse point may discourage the procedure from stopping in a local minimum. The search continues in this fashion until it becomes trapped at the same point for L consecutive iterations. Of course, L is yet another control parameter.

Component 2: Cooling Schedule. In the above search the parameter β is thought of as the inverse of the temperature in Gibbs' distribution. The search is repeated for an increasing sequence of β's. The choice of this sequence is known as the cooling schedule. The selection of the cooling schedule as well as the decision of when the process is sufficiently cooled requires some problem specific "trial and error" (van Laarhoven and Aarts, 1987, review work on semi-automatic schedules.)

1.3. GENERALIZED SIMULATED ANNEALING

Bohachevsky, Johnson and Stein (1986) (hereafter, BJS) considered a modification of the standard algorithm for continuous $\mathscr{X}$. The key to their suggestion is to modify the move selecting distribution in an attempt to accelerate the convergence to the minimum.

Suppose the global minimum value of ϕ, say $\phi_{\min}$, is known. (There are problems in which $\phi_{\min}$, but not its location, is known. Also, suppose one wishes to find the roots of a function, say f. One could then minimize $\phi = f^2$, for example, with known minimum $\phi_{\min} = 0$.) In this case BJS consider the annealing algorithm as above but with (1.1) replaced by

$$\min[1, \exp\{-\beta(\phi(x^*) - \phi_{\min})^g(\phi(x^*) - \phi(x))\}]. \tag{1.2}$$

Here g is an additional control parameter. In examples, BJS use $g = -1$.

For problems in which $\phi_{\min}$ is unknown, BJS suggest a simple approach in which one begins with an initial estimate of $\phi_{\min}$. The estimated value of $\phi_{\min}$ is then just decreased, in a problem specific fashion, whenever the search encounters smaller values.

2. A Stochastic Probing Algorithm

The motivation behind the algorithm presented is essentially Bayesian in nature. We suppose the user considers the incorporation of prior information concerning the location of the minimum of ϕ. This information is used to construct a *probing distribution* on the feasible region. Though this prober plays the role of a prior in a Bayesian approach, it is updated via an ad hoc scheme, rather than via Bayes Theorem, as the search proceeds. A thorough discussion of Bayesian approach to the development of numerical methods of global optimization of continuous multimodal functions and functions with "noise" is given in Mockus (1989). In this monograph, advantages and disadvantages of Bayesian approach (average case analysis), comparing it with more usual minimax approach; some theoretical and computational problems of probability related to the Bayesian approach to global optimization; and software for global optimization are also discussed.

2.1. OUTLINE OF THE ALGORITHM

In the following description we consider the continuous minimization problem and assume $\mathscr{X}$ is a subset of m-dimensional Euclidean space. We begin with the construction of a probing distribution, with density $p(\cdot \mid x_0, \sigma_0)$. It is assumed that p is a multivariate location-scale density, with location x_0, and scale σ_0, whose support contains $\mathscr{X}$. Through the choice of x_0 and σ_0, p can be chosen to reflect the prior information of the user. However, the generation distributions for the search are of the same family as p, so the ability to easily simulate from p is essential.

The search recipe of the algorithm involves the production of a sequence of pairs of location-scale parameters (x_n, σ_n), $n \geq 1$, as follows:

(1) *Generation Step:* At stage $n (n \geq 0)$, generate k independent, identically distributed points $x_{n1}, \ldots, x_{nk}$ from $p(\cdot \mid x_n, \sigma_n)$, subject to $x_{ni} \in X$. Let $x_{n0} = x_n$ and $\mathbf{x}^n = (x_{n0}, \ldots, x_{nk})$.

(2) *Move Selection*: x_{n+1} is chosen according to the following Gibbs-like distribution on $k+1$ points;

$$\Pr(x_{n+1} = x_{ni}) = \frac{1}{Z} e^{-B(\phi, i, \mathbf{x}^n)}, \quad i = 0, \ldots, k; \tag{2.1}$$

where $Z = \sum_{i=0}^{k} e^{-B(\phi, i, \mathbf{x}^n)}$. The choice of the functional B is left open for now.

(3) *Prober Scale Reduction*: The parameter σ_n controls the breadth of the search at stage n. Intuitively, σ_n should be relatively large for $n = 0$, but should decrease if the search is successful. The basic rule considered here is of the form

(i) $\sigma_{n+1} = \sigma_n$ if $\phi(x_{n+1}) \geq \phi(x_n)$

(ii) $\sigma_{n+1} = r(\sigma_n) < \sigma_n$ if $\phi(x_{n+1}) < \phi(x_n)$,

where $r(\cdot)$ is an appropriately chosen scale reducing function. Some choices are discussed below.

Finally, we consider stopping rules as follows: Continue the search until the volume of the $(1-\alpha)\times 100\%$ highest probability density region of $p(\cdot \mid x_n, \sigma_n)$ is less than ε. Both α and ε are chosen by the user based on the desired accuracy of the final guess (the best point ever seen during the search) at the minimizer.

To keep the rest of the discussion in perspective, note some of the similarities and differences between our algorithm and annealing. First, our search is a directed global search which only "localizes" once it is close to the minimum (hopefully). We believe this can offer a potential improvement over annealing in that we can achieve very fast convergence if (i) the initial prober is "on target" or (ii) the early search is simply "lucky". A more important advantage is that the method avoids the "cooling schedule" problem of annealing. Of course, nothing comes for free: We must choose a scale reduction function as discussed later.

2.2. SUGGESTIONS FOR COMPONENTS OF THE ALGORITHM

2.2.1. Move Selection. We assume, as do BJS initially, that the minimum $\phi_{\min}$ is known. A useful choice for $B(\cdot)$ in (2.1) is

$$B(\phi, i, \mathbf{x}^n) = \frac{\phi(x_{ni}) - \phi(x_{n0})}{\min(\phi(x_{ni}), \phi(x_{n0})) - \phi_{\min}} . \tag{2.2}$$

This choice is partially motivated by (i) the generalized approach of BJS and (ii) some computational experiments.

2.2.2. Scale Reduction Rule. As mentioned above, the prober is chosen from a convenient location, scale family. To start the search we suggest that σ_0 should be relatively large. For example, σ_0 could be chosen so that the $(1-\alpha)\times 100\%$ highest probability density region (HPD) of $p(\cdot \mid x_0, \sigma_0)$ just covers X. (Of course, since we generate sample points from the prober, it would then be possible to generate points outside of X. Such points are simply discarded. It is assumed that the simulation costs are negligible compared to costs in evaluating ϕ.) The scale reducing function considered here is as follows. (Recall that $\sigma_{n+1} = \sigma_n$ if $\phi(x_{n+1}) \geq \phi(x_n)$.) The key is to relate σ_n with v_n where v_n is the *volume* of the $(1-\alpha)$ 100% HPD region of $p(\cdot \mid x_n, \sigma_n)$. Specifically,

Reduction Rule. If $\phi(x_{n+1}) < \phi(x_n)$, then let σ_{n+1} be that value such that for $0 < w_n < 1$,

$$V_{n+1} = w_n V_n . \tag{2.3}$$

The choice of w_n should ideally allow w_n to reflect how close the algorithm is to the minimizer.

The ad hoc choice for w_n considered here is based on the entropy of the current move selecting distribution; namely

$$\text{Ent}(n) = -\sum_{i=0}^{k} \Pr(x_{n+1} = x_{ni}) \log \Pr(x_{n+1} = x_{ni}) ,$$

where $\Pr(\cdot)$ is given in moving selecting distribution, (2.1). In particular, consider

$$w_n = \frac{\text{Ent}(n)}{\log(k+1)} . \tag{2.4}$$

First, $\text{Ent}(n)$ is a measure of the variability of the move selection distribution. Since this distribution was tailored to favor small values of ϕ, a small $\text{Ent}(n)$ indicates, *very roughly*, some concentration toward the minimum. (Also, there is a phenomenological analogy to thermodynamics in that as the temperature goes to zero the entropy of Gibbs' distribution collapses to the logarithm of the number of states which yield the minimum energy. If there is only one such state, the limiting entropy is then zero. Of course, the analogy is very loose.)

If $\phi_{\min}$ is known, we can make use of this information in the choice of w_n. A potential concern arises if $\phi(x_{n+1})$ is small compared to the other k candidates at the nth step, but $\phi(x_{n+1})$ is still much larger than $\phi_{\min}$. To avoid the resulting artificial decrease in w_n, we employed the following modification w_n^* of (2.4). First, consider the two point distribution on $\phi(x_{n+1})$ and $\phi_{\min}$ where

$$\Pr^*(\phi(x_{n+1})) = \frac{1}{Z^*} e^{-(\phi(x_{n+1})-\phi_{min})} ,$$

$$\Pr^*(\phi_{\min}) = 1 - \Pr^*(\phi(x_{n+1})) ,$$

$$Z^* = 1 + e^{-(\phi(x_{n+1})-\phi_{\min})} .$$

Let Ent* be the entropy of this distribution, and define

$$w_n^* = \min\left[1, \frac{w_n}{(\text{Ent}^*/\log(2))}\right] . \tag{2.5}$$

2.2.3. Elliptically Contoured Probers. In order for the algorithm to be "efficient", the choice of the prober should generally be made based on flexibility as well as computational ease. As a general suggestion, the class of elliptically contoured distributions should be very useful. An excellent reference for this discussion is Johnson (1987), especially Chapter 6.

Assume that each x in $\mathcal{X}$ is a $p \times 1$ vector. Consider probers of the form

$$p(\cdot \mid \mu, \sigma) = c_p \mid \sigma \mid^{-1/2} g((x-\mu)'\sigma^{-1}(x-\mu)) , \tag{2.6}$$

where c_p is a normalizer, μ is a $p \times 1$ location vector and σ is a $p \times p$, non-singular scaling matrix. The choice of $g(\cdot)$ is left open for this discussion.

Such probers are quite attractive for our purposes. First, simulation from the prober is relatively easy; see Johnson (1987), p. 110. Second, such probers permit

a very simple rule for updating σ_n according to (2.3). The key observation is that HPD regions for (2.6) can be expressed in terms of level sets of the quadratic form $(x-\mu)^t\sigma^{-1}(x-\mu)$, whose distribution is independent of μ and σ. In particular, suppose σ_0 and v_0 have been specified. For an elliptically contoured prober, σ_n can be updated according to the reduction rule, (2.3) as follows.

LEMMA. *If* $0 \leq w_n \leq 1$ *and*

$$\sigma_{n+1} = (w_n)^{2/p}\sigma_n \tag{2.7}$$

then $v_{n+1} = w_n v_n$.

Proof. The proof follows directly from the above observations and is therefore omitted.

Note, that (2.7) is not the only update consistent with (2.3). In particular, (2.7) maintains the orientation of the prober.

3. Performance of the Probing Algorithm

Bohachevsky *et al.* (1986) considered three functions on $[-1, 1]^2$ to study the performance of their generalized simulated annealing algorithm. These functions are:

$$\phi_1(x, y) = ax^2 + by^2 - c\cos(\alpha x) - d\cos(\gamma y) + c + d \tag{3.1}$$

$$\phi_2(x, y) = ax^2 + by^2 - c\cos(\alpha x)\cos(\gamma y) + c \tag{3.2}$$

$$\phi_3(x, y) = ax^2 + by^2 - c\cos(\alpha x + \gamma y) + c \tag{3.3}$$

with $a = 2$, $b = 2$, $c = 0.3$, $d = 0.4$, $\alpha = 3\pi$, $\gamma = 4\pi$. Each of these functions has many local minima and a unique global minimum of 0 at $(0, 0)$. The following 3-dimensional version of ϕ_1 was also used in the performance comparisons discussed below:

$$\begin{aligned}\phi(x, y, z) = {} & x^2 + 2y^2 + 3z^2 - 0.3\cos(3\pi x) - 0.4\cos(4\pi y) \\ & - 0.5\cos(5\pi z) + 1.2 .\end{aligned} \tag{3.4}$$

A three-dimensional perspective plot of ϕ_1 as well as contour plots of ϕ_1, ϕ_2 and ϕ_3 can be seen in BJS.

3.1. AN IMPLEMENTATION OF THE STOCHASTIC PROBING ALGORITHM

The stochastic probing method was applied to each of the above functions. In each case, the prober used was Gaussian with independent components having common scale σ. The scale reduction rule given by (2.5) was employed. The initial value of σ was taken to be 0.7 while the stoping rule required this scale to reduce by a factor of $(0.0001)^{1/p}$ where p is the dimension of the prober. (This is

equivalent to reducing the volume of a fixed probability-content HPD region by a factor of 0.0001.)

Figure 1 depicts the progress of a typical instance of the stochastic probing algorithm applied to ϕ_1. The initial guess for the location of the prober was (1.0, 1.0) and the probing sample size was $k = 2$. Panel (a) shows a plot of the objective function value at each of the three points (the location of the prober marked by 'x' and the two points sampled from it marked by 'o') at each iteration of the algorithm. The horizontal axis is labelled by the number of function evaluations rather than the iteration number. The panel also shows how σ, the scale of the prober, reduces as the search proceeds. The right-hand vertical axis marks the scale for σ and the solid line connects the plotted values. Panel (b) of Figure 1 shows the history of the same probing run as viewed through the distance from the optimal point. Figure 2 shows the progress of a typical probing run for

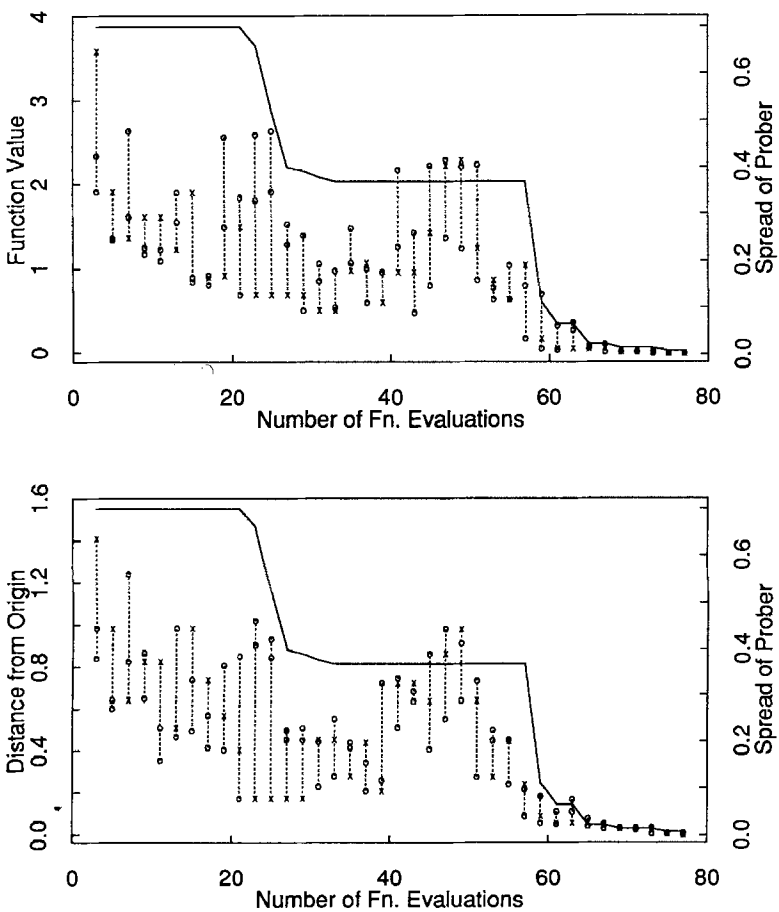

Fig. 1. Progress of a typical probing run for function phi-1 in 2 dimensions. x: value at current location of prober; o: value at sampled point.

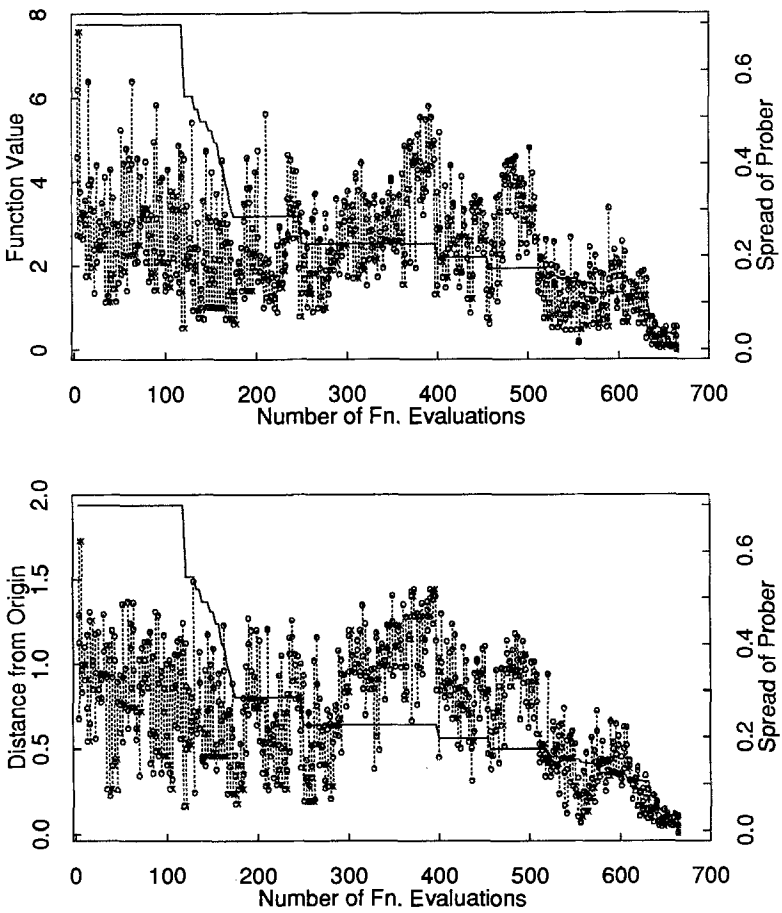

Fig. 2. Progress of a typical probing run for function phi-1 in 3 dimensions. x: value at current location of prober; o: value at sampled point.

the three-dimensional version of ϕ_1 given in (3.4) above. Here the initial location of the prober is (1.0, 1.0, 1.0), initial $\sigma = 0.7$ and the probing sample size is $k = 3$. It should be pointed out that these runs used only the input mentioned above and that, unlike annealing, no parameters had to be fine-tuned or empirically determined.

3.2. COMPARISON WITH GENERALIZED SIMULATED ANNEALING

We next describe a comparison of the generalized simulated annealing algorithm of BJS and our stochastic probing algorithm as described. Both methods were applied 75 times each to the functions in (3.1)–(3.4) above. The performance measures used were (i) number of evaluations of the objective function, (ii) objective function value, and (iii) distance from the optimal point when the

algorithm terminated. The first of these addresses speed and the other two reflect the quality of the solution found.

For each of the above objective functions, the 75 runs included three different initial guesses and 25 different seeds for the random number generator. For ϕ_1, ϕ_2, and ϕ_3 on the two-dimensional domain $[-1, 1]^2$, the starting points were: (1.0, 1.0), (0.6, 0.45)and (0.6, 0.0). For the 3-dimensional version of ϕ_1 the initial guesses (1.0, 1.0, 1.0), (0.6, 0.45, 0.25) and (0.6, 0.0, 0.0) were used. The results for the four objective functions are presented Figures 3, 4, 5 and 6. To be specific, we describe Figure 3 in detail, the others being similar. The figure contains six plots arranged in three columns and two rows. The first column shows, in two different ways, the comparison using the performance measure (i). Here we note the difficulty in counting the number of objective function evaluations for a run of the simulated annealing method: parameters such as β and Δr (see BJS for details) must be determined empirically. This clearly involves a substantial number of function evaluations and preliminary computations. In the comparisons pictured here, this cost was completely ignored in that the experimentally determined best values suggested in BJS were used for each function: $\Delta r = 0.5$; $\beta = 3.5$ for ϕ_1 and $\beta = 3.0$ for ϕ_2, ϕ_3.

The plot in the top row shows the box-plots for the two algorithms. As noted in the plot, only 60 observations out of the possible 75 were used for the simulated

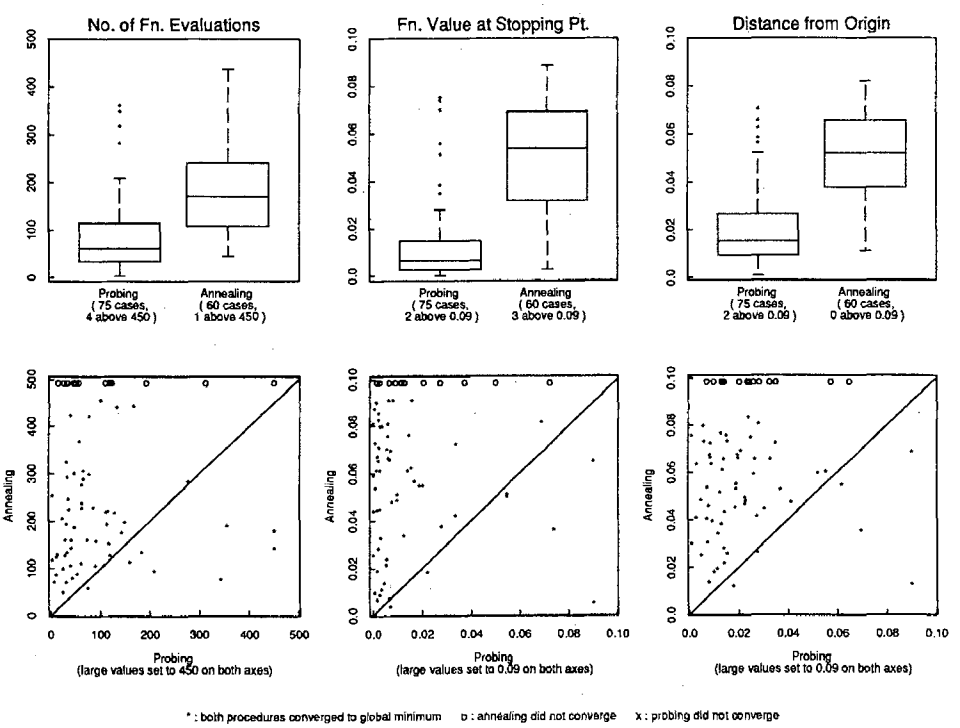

Fig. 3. BJS function phi-1 in 2 dimensions.

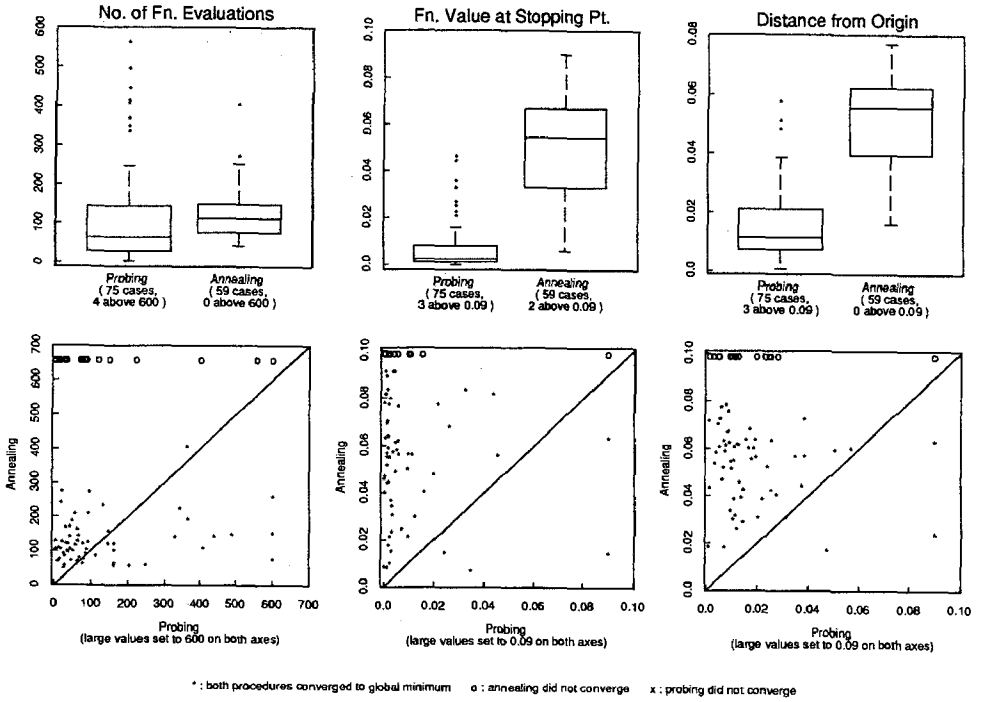

Fig. 4. BJS function phi-2 in 2 dimensions.

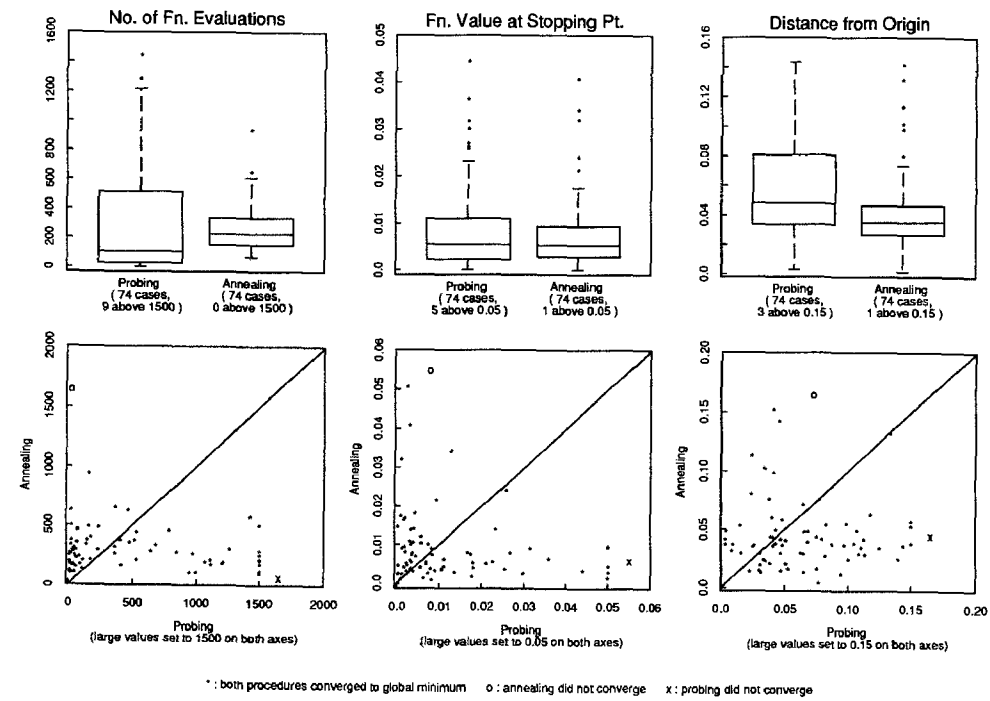

Fig. 5. BJS function phi-3 in 2 dimensions.

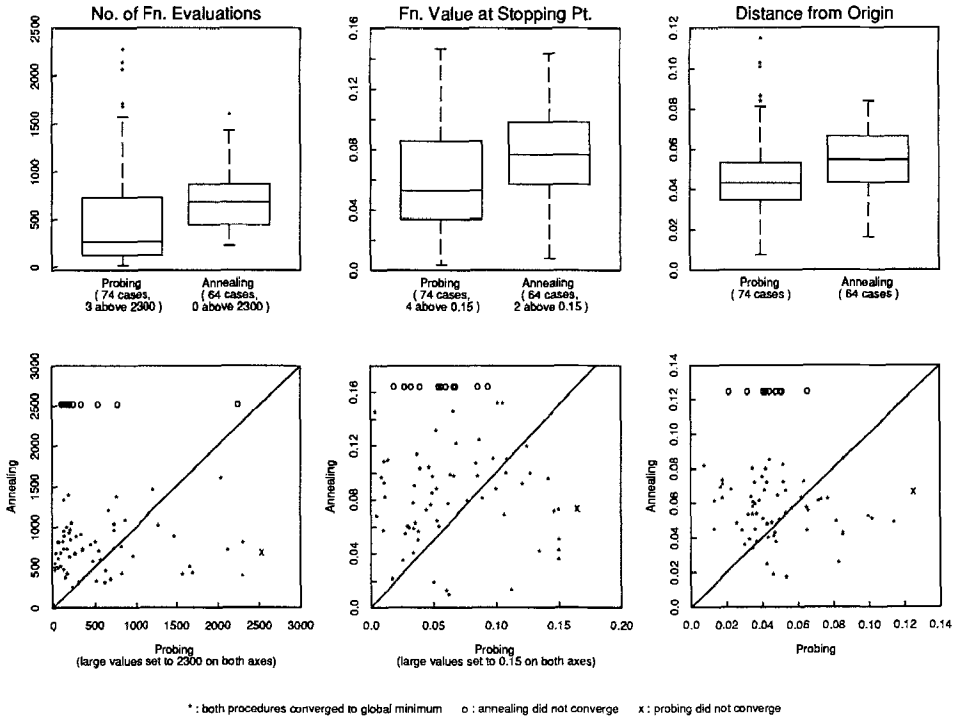

Fig. 6. BJS function phi-1 in 3 dimensions.

annealing box-plot. This was caused by the 15 cases in which the algorithm failed to converge to the global minimum. The stopping criteria for the annealing algorithm were: (i) $\phi < 0.0001$ or (ii) 35 successive trials fail to produce an acceptable move. Finally, the data were truncated at 450 for convenience of plotting scale. The number of truncations are noted in the plot. (The actual values were: 523 for annealing and 571,663,1255,3481 for probing.) Overall, the comparison favors the stochastic probing algorithm. When one notes the initial costs of experimentally determining the annealing constants, the stochastic probing algorithm appears to perform substantially better. Further, the number of failures of annealing to converge to the minimum is disturbing by comparison. Perhaps we should not make too much of this since we are running at only the recommended parameters and not allowing annealing to further "cool". It does make one wonder how the decision to stop the annealing algorithm in a given problem is to be made.

The plot in the second row is a scatter diagram of the 75 runs matched by initial guess and random number seed. Here matching by seeds is only for convenience. Lack of convergence of either algorithm is indicated by special symbols as noted at the bottom of the Figure. The 45° line is drawn for ease in comparing the two methods; points above it favor stochastic probing.

Plots in columns 2 and 3 of Figure 3 use performance measures (ii) and (iii) as noted in the column titles; in other respects these plots were made in the same manner as those in column 1. The entire Figure 3 provides a visual comparison of the two algorithms applied to ϕ_1. All three performance measures indicate that stochastic probing outperformed simulated annealing.

Figure 4 shows the comparison using the objective function ϕ_2. Again the conclusion is the same as with ϕ_1. Figure 5 shows that among the objective functions used, ϕ_3 brings the performance of the two methods closer together than any other. The comparison using the 3-dimensional objective function (3.4) is contained in Figure 6. Overall, the stochastic probing algorithm appears to outperform the generalized simulated annealing algorithm even when the latter is "pre-cooled". It should also be noted that we did not allow any of the probing runs to rely on one of the primary goals of stochastic probing; namely, the incorporation of prior information. In particular, all probing runs were initiated at poor initial choices for the location of the minimum. At least in these test cases, stochastic probing displayed a reasonable degree of robustness with respect to poor initialization.

3.3. VARIATION FROM RUN-TO-RUN

There is an appreciable variation in the number of function evaluations needed to reach the global minimum. This is apparent in the box-plots of Figures 3–6. To

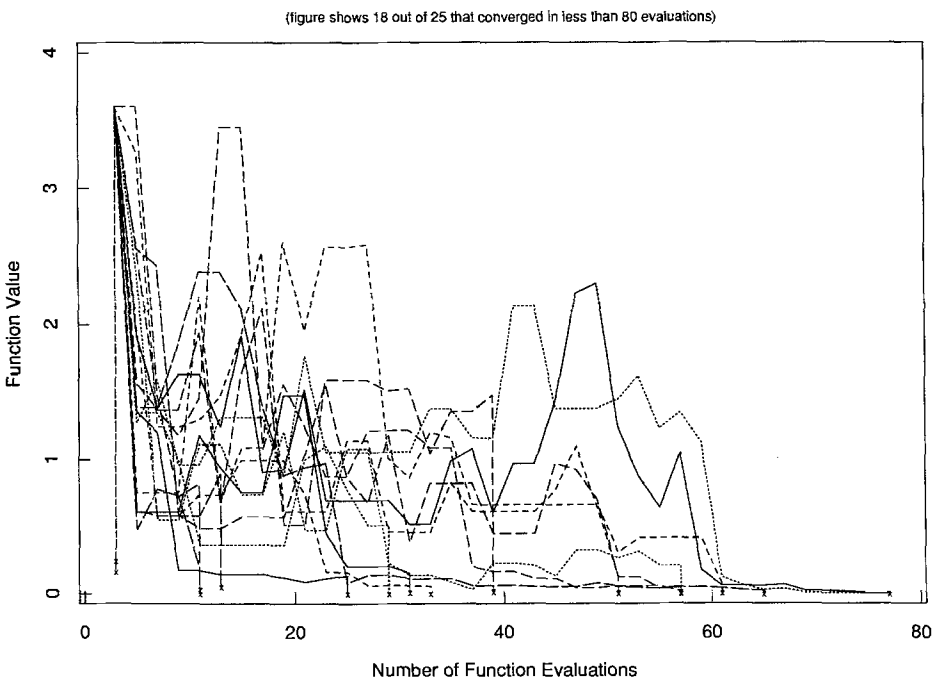

Fig. 7. Samples of probing runs, phi-1 in 2 dimensions, $k = 2$, initial guess = (1, 1).

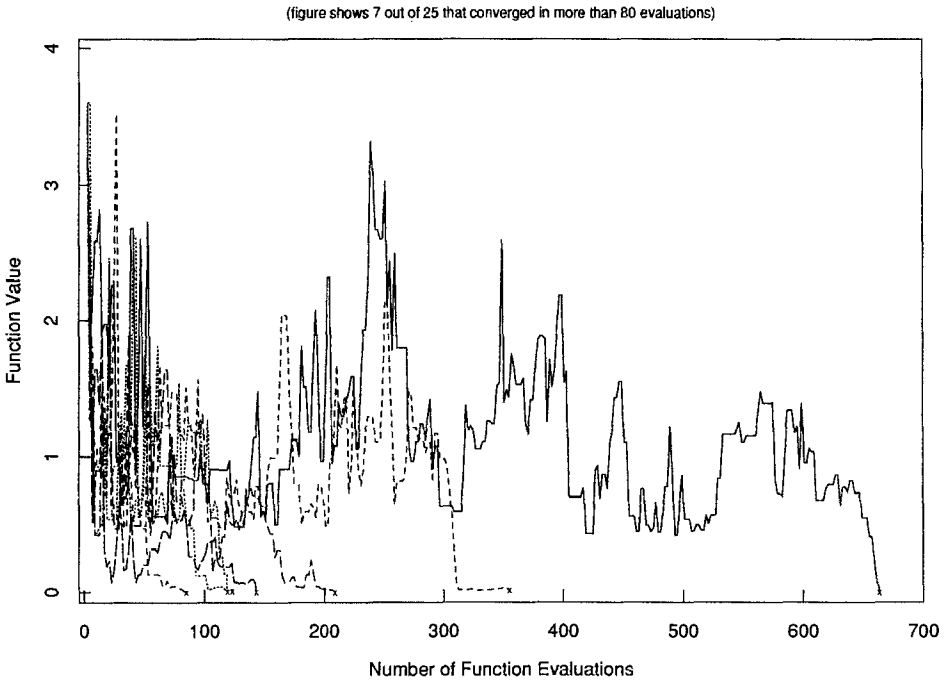

Fig. 8. Samples of probing runs, phi-1 in 2 dimensions, $k = 2$, initial guess = (1, 1).

visually display the run-to-run variation in the progress toward the global minimum, we plotted 25 such runs for ϕ_1 with the initial prober location (1.0, 1.0). (These are 25 of the 75 runs of Figure 3.) For reasons of good plotting scale and visual clarity, these runs are pictured in two figures: Figure 7 contains those runs that took less than 80 evaluations, Figure 8 contains the rest. It is interesting to note how the function value fluctuates while maintaining a downward trend. Many of the quick converges appear to involve a sudden drop in the function value. These are chance occurrences which the global approach of the stochastic probing method seems to exploit better than the local search approach of the simulated annealing method. The stochastic probing algorithm as implemented here performs reasonably well in recognizing these situations and reducing the scale of the prober accordingly. There are, of course, pitfalls that sometimes cannot be avoided: sudden drop to a local minimum is an example. Some limited experience with other objective functions suggests that restarting a search is sometimes helpful when a particular run appears to wander aimlessly with a reasonably small value for the scale of the prober. Parallel implementations could prove quite useful in this regard.

4. Discussion

Our goal in this paper was to describe the "bare bones" of our probing algorithm. Encouraged by what we perceive to be the positive results of Section 3, we plan to

pursue a number of theoretical and computational issues related to probing, as well as its applications to large scale problems.

Many of the general issues to be considered concern the compromise between generality of the probing algorithm and the modification of the method in problem specific directions. For example, an ad hoc updating procedure of BJS for estimating the unknown value of $\phi_{\min}$ could be improved upon, perhaps by a Bayesian learning model, in some cases. Another area for refinements of the algorithm is the class of discrete optimization problems such as the Traveling Salesman Problem; see Aarts and Korst (1989). We are confident that some success is possible in such cases. Finally, as indicated at the end of Section 3, parallel implementations of probing may be desirable. We are developing parallel and "vectorized" versions of the probing algorithm.

In view of the spirit of the previous paragraph, some readers may develop a negative impression of probing in relation to annealing. Indeed, some may argue that annealing is a general, easily implemented procedure whereas probing is too problem specific for general use. Without wishing to detract from annealing, we do suggest the following points. First, annealing is not really a completely automatic procedure. The many control parameters used in annealing as well as the problem of its cooling schedule force substantial computation time for a successful outcome. Second, probing can be implemented with very little user input in the annealing style; that is, by running several probes with different probing parameters. In fact, an appealing general procedure, easily implemented based on this paper, is an annealing-prober algorithm achieved by simply introducing a temperature parameter in (2.2) and cooling as in annealing. Finally, the potential for problem specific tailoring, including the use of prior information when available, of the probing algorithm should be perceived as a positive characteristic.

Acknowledgements

We have profited from discussions with George Casella, Jerry Sacks, Michael Steele and Robert Wolpert. We are grateful to the referees for their helpful comments. This research was supported in part by the National Science Foundation under grants DMS-8906787 and DMS-9008067.

References

Aarts, E. and J. Korst (1989), *Simulated Annealing and Boltzmann Machines*, New York: John Wiley.

Bélisle, C. J. P., H. E. Romeijn and R. L. Smith (1990), Hide-and-Seek: A Simulated Annealing Algorithm for Constrained Global Optimization, Technical Report 176, Department of Statistics, The University of Michigan, Ann Arbor, Michigan, U.S.A.

Bohachevsky, I., M. E. Johnson and M. L. Stein (1986), Generalized Simulated Annealing for Function Optimization, *Technometrics* **28**, 209–217.

Cerny, V. (1985), Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm, *Journal of Optimization Theory and Applications* **45**, 41–51.

Geman, S. and D. Geman (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Proc. Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geman S. and D. E. McClure (1985), Bayesian Image Analysis: An Application to Single Photon Emission Tomography, *Proc. American Statistical Association, Statistical Computing Section*, 12–18.

Haines, L. M. (1987), The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models, *Technometrics* **29**, 439–448.

Kirkpatrick S., C. D. Gelatt and M. P. Vecchi (1983), Optimization by Simulated Annealing, *Science* **220**, 671–680.

Johnson, M. E. (1987), *Multivariate Statistical Simulation*, New York: John Wiley.

Laarhoven, P. J. M. van, and E. H. L. Aarts (1987), *Simulated Annealing: Theory and Applications*, Dordrecht: Reidel.

Mockus, J. (1989), *Bayesian Approach to Global Optimization*, Dordrecht: Kluwer Academic Publishers.

Pincus M. (1970), A Monte-Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems, *Operations Research* **18**, 1225–1228.

Pronzato, L., E. Walter, A. Zenot and J. F. LeBruchec (1984), General Purpose Global Optimizer, *Mathematics and Computers in Simulation* **26**, 412–422.

Rinnoy Kan, A. H. G., C. G. E. Boender and G. Th. Timmer (1985), A Stochastic Approach to Global Optimization, in *Computational Mathematical Programming*, ed. K. Schittkowski, Berlin: Springer-Verlag, 281–308.

Ripley, B. D. (1987), *Stochastic Simulation*, New York: John Wiley.