Bayesian method for global optimization

Prasana K. Venkatesh,[1,2] Morrel H. Cohen,[2] Robert W. Carr,[1] and Anthony M. Dean[2]
[1]*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455*
[2]*Corporate Science Laboratories, Exxon Research and Engineering Co., Annandale, New Jersey 08801*
(Received 19 September 1996)

We introduce a Bayesian approach for obtaining the global optimum of multimodal functions. The set of observed minima of a multimodal function is viewed as a sample from a multinomial distribution, whose cells correspond to the basins of attraction of the local optima. We then derive the posterior distribution of the number of local optima. This posterior information is obtained from a Bayesian analysis and is used to construct a stopping criterion for a *sequential* random search method which finds the optimal tradeoff between reliability and computational effort. The computational complexity of this global optimization method is a strong function of the total number of local optima and a weak function of the dimensions of the configuration space. Application to four classical problems from the global optimization literature, a bifunctional catalytic reactor problem, and the conformation problem of Lennard-Jones microclusters is demonstrated. Comparisons with the Bayesian method of Boender and Rinooy Kan [Math. Program. **37**, 59 (1987)] and the simulated annealing method of Dekkers and Aarts [Math. Program. **50**, 367 (1991)] are provided and, in each case, the computational complexity of our method is shown to be smaller than that of these methods. [S1063-651X(97)13205-6]

PACS number(s): 02.70.Lq, 02.60.Pn, 36.40.−c

I. INTRODUCTION

Since the advent in 1983 of simulated annealing [1,2], a *Monte Carlo* method for stochastically tracing the convergence of the Gibbsian distribution of a multimodal function to a Dirac δ function over its global minimum, there has been a resurgence of interest in the global optimization of functions. Significant developments include genetic algorithms [3,4] based on heuristic principles of natural selection and the GOP method [5,6], which is a deterministic method of global optimization, for certain restricted classes of mathematical programming problems, based on decomposition principles relating to the theory of duality [7] of mathematical programs. Each of these methods possesses particular advantages and disadvantages.

Bayesian methods for global optimization have been proposed for some time now [8], and those relating to the statistical structure of multiextremal problems [9–12] are not only interesting in their own right but could prove to be computationally promising. In the Bayesian approach, one expresses beliefs regarding some unknown relevant parameters of the function in the form of a prior distribution. Experimental or computational information obtained about the function is then used to transform this prior distribution into a posterior distribution through the utility of *Bayes*' theorem. The latter distribution is representative of the manner in which prior notions are affected by the outcome of the experiments. Thereafter, a decision on whether to stop the search can be taken based on a criterion that quantifies a loss if the search is stopped prior to the location of all of the optima. In [10] Boender and Rinooy Kan provided a rigorous framework for the development of optimal stopping rules for a *multistart* method for global optimization wherein the configuration space is repeatedly sampled at batches of uniformly distributed points and local searches initiated therefrom. The probability of the aggregate of events resulting from their multistart method is given by a generalized multinomial distribution [10]. Assuming that the number of optima of the function is equiprobable along the positive half of the integral real axis and that the relative sizes of the regions of attraction are uniformly distributed, they derived posterior expected quantities using Bayesian estimation theory [13]. They used the information generated by these quantities to construct single-step sequential Bayesian stopping rules for their multistart global optimization method. Their method, however, does not make full use of the information gleaned during the local searches to construct posterior notional quantities that are more effective.

In this paper we develop and analyze a Bayesian method for global optimization. We illustrate the performance of our method on classical multimodal test functions from the global optimization literature; as well as on an optimal control problem in the design of a bifunctional catalytic reactor [14] and a multimodal atomic-microcluster conformation problem [15]. The latter problems are particularly complex in terms of the number of variables in the configuration space and the number of optima supported by the objective functions. The computational complexity of our method increases with the total number of local optima supported by the objective function, and is only weakly dependent on the dimensions of the configuration space, in contrast to the complexity of simulated annealing, which increases exponentially with the dimension of the configuration space of the objective function. Simulated annealing assumes and requires no initial knowledge of the function. Similarly, when there is no initial knowledge of the function, the Bayesian prior distribution is taken as uniform. Accordingly, we compare the performance of our Bayesian method only with the most efficient of the stopping rules of Boender and Rinooy Kan [10] and with the most efficient of the algorithms thus far proposed for simulated annealing [2]. The measure of performance used is the total number of function evaluations required to satisfy the

stopping rule of each method and in each case the computational complexity of our method is shown to be smaller than that of these methods.

II. PRELIMINARIES

The objective is to find the global optimum x^* of a real-valued multimodal function $f(x)$ defined over a compact set S. The design of our scheme is based on a statistical analysis of a sequential-*search* method. In this sequential-search method a point is drawn randomly from the *uniform* probability distribution over S and a local search initiated from this point thereafter. This sequence of drawing a uniformly distributed point, and conducting a local search, is then reiterated until a suitable probabilistic guarantee is obtained of having obtained all of the optima of $f(x)$. In order to construct a formalism for obtaining such a guarantee, we perform a statistical analysis based on Bayes' theorem.

We begin by assuming that the function $f(x)$ can have a specific set of minima of number N, with associated basins of attraction of volume fraction θ_k, $k=1,\ldots,N$,

$$\sum_{k=1}^{N} \theta_k = 1. \tag{2.1}$$

In practice, N and θ_k, $k=1,\ldots,N$ are always unknown. Suppose that we conduct M successive searches for local minima and their associated basins of attraction via our sequential-search method. Suppose also that we find ν distinct basins. We are then faced with the question of whether ν equals or is less than the actual number of minima N present in $f(x)$. Since we now know the values of $f(x)$ only at a set of points of zero measure in S, we can answer that question only probabilistically. Given that we know we have at least ν minima, we can estimate the probability that N exceeds ν, $P(N>\nu/M)$. When $P(N>\nu/M)$ becomes acceptably small, i.e., less than some preset $\varepsilon>0$ at some M as M is increased, the search is terminated. The more prior information fed into the analysis, the more accurately $P(N>\nu/M)$ is estimated, and we do know more than just ν. More than one of the M starting points must be found to lie in a single basin of attraction when $M>\nu$. Label the basins found by an index k, $k=1,\ldots,\nu$. We thus know μ_k, the number of starting points found to lie in each basin k,

$$\sum_{k=1}^{\nu} \mu_k = M. \tag{2.2}$$

Let $P(\mu_1,\ldots,\mu_\nu/M)$ be the probability of finding such a set of M starting points in ν distinct basins of attraction. Let $P(\mu_1,\ldots,\mu_N/M,N)$ be the conditional probability of finding the number of starting points $\{\mu_1,\ldots,\mu_N\}$ lying in their appropriate basins after M random searches on the function $f(x)$ of which it is known only that it has $N\geqslant\nu$ minima such that

$$\mu_k = 0, k=\nu+1,\nu+2,\ldots,N. \tag{2.3}$$

Then $P(\mu_1,\ldots,\mu_\nu/M)$ can be expressed as

$$P(\mu_1,\ldots,\mu_\nu/M) = \sum_{N=\nu}^{\infty} P(\mu_1,\ldots,\mu_N/M,N)\varphi(N), \tag{2.4}$$

where $\varphi(N)$ is the *prior* probability that the function $f(x)$ has N minima. The Bayesian stopping rule of our sequential-search method consists in terminating the search method consists in terminating the search at that value of M at which the *posterior* probability

$$P(N>\nu/\mu_1,\mu_2,\ldots,\mu_\nu;M) = \sum_{N=\nu+1}^{\infty} P(N/\mu_1,\mu_2,\ldots,\mu_\nu;M) \leqslant \delta = \frac{\varepsilon}{1+\varepsilon} \tag{2.5}$$

becomes acceptably small, or, equivalently,

$$P(N=\nu/\mu_1,\mu_2,\ldots,\mu_\nu;M) \geqslant 1-\delta = \frac{1}{1+\varepsilon} \tag{2.6}$$

becomes close to unity. $P(N>\nu/\mu_1,\mu_2,\ldots,\mu_\nu;M)$ is then sharp about $N=\nu$ because

$$P(N<\nu/\mu_1,\mu_2,\ldots,\mu_\nu;M) = 0. \tag{2.7}$$

Under the circumstances denoted by Eqs. (2.1)–(2.7), it is reasonable to employ Bayes' theorem to estimate $P(N/\mu_1,\mu_2,\ldots,\mu_\nu;M)$. Since we have found $\{\mu_1,\ldots,\mu_\nu\}$, but have no prior information about N other than that it exceeds ν, $P(N/\mu_1,\mu_2,\ldots,\mu_\nu;M)$ is, according to Bayes' theorem, simply proportional to $P(\mu_1,\ldots,\mu_N/M,N)$ itself

$$P(N/\mu_1,\mu_2,\ldots,\mu_\nu;M) = \frac{P(\mu_1,\ldots,\mu_N/M,N)}{\sum_{N=\nu}^{\infty} P(\mu_1,\ldots,\mu_N/M,N)}. \tag{2.8}$$

III. APPROXIMATING ABOUT THE MONTE CARLO ESTIMATE OF VOLUME FRACTIONS

Equation (2.8) impels an explicit analysis of $P(\mu_1,\ldots,\mu_N/M,N)$ to obtain a reliable estimate of $P(N/\mu_1,\mu_2,\ldots,\mu_\nu;M)$.

A. Estimating the conditional probability $P(\mu_1,\ldots,\mu_N/M,N)$.

Our development thus far parallels that of [9]. Where we differ is in the estimation of $P(\mu_1,\ldots,\mu_N/M,N)$. Let $P(\theta_1,\ldots,\theta_N/N)$ be the probability that the N basins of attraction have the set of N volume fractions $\{\theta_1,\ldots,\theta_N\}$. Each θ_k lies in the open interval (0,1). If any of the θ_k were 0 or 1, there would be $N-1$ minima or 1 minimum, respectively, contradicting the prior notion that there be N minima. In addition the θ_k sum to unity as given by Eq. (2.1).

At this point in the argument, we deliberately ignore any prior knowledge of the θ_k which might be inferred from the local searches. For now, we suppose the θ_k to be uniformly distributed on the hyperplane given by Eq. (2.1) within the open simplex $\theta_k\in(0,1)$, as Boender and Rinooy Kan have done [10,12]. We introduce such prior knowledge in Sec. V below.

The M successive starting points are independently and randomly chosen. The conditional probability of obtaining the set of starting points $\{\mu_1,\ldots,\mu_N\}$, given the values $\{\theta_1,\ldots,\theta_N\}$, of the volume fractions, is the same as that of the probability of sorting M objects into N boxes,

$$P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)=M!\prod_{k=1}^{\nu}\frac{\theta_k^{\mu_k}}{\mu_k!}. \quad (3.1)$$

The full probability $P(\mu_1,\ldots,\mu_N/M,N)$ is then

$$P(\mu_1,\ldots,\mu_N/M,N)=\int\cdots\int P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)\times P(\theta_1,\ldots,\theta_N/N)d\theta_1,\ldots,d\theta_N. \quad (3.2)$$

B. Evaluating $P(\mu_1,\ldots,\mu_N/M,N)$ for moderately large M

For values of $\{\mu_k\}$ of order 10 or larger, $P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)$ has a single maximum for $N=\nu$ and a single sharp supremum at $\{\theta_1^*,\ldots,\theta_N^*\}$ for $N>\nu$; this is shown in Appendix A. The value of $\{\theta_1^*,\ldots,\theta_N^*\}$ is

$$\theta_k^*=\frac{\mu_k}{M},\quad k=1,\ldots,N \quad (3.3)$$

in both cases, Eq. (3.3) implying that $\theta_k^*=0$ for $k=\nu+1,\ldots,N$, when $N>\nu$. The maximum value of $P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)$ is

$$P^*(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)=M!\prod_{k=1}^{\nu}\frac{(\mu_k/M)^{\mu_k}}{\mu_k!} \quad (3.4)$$

in both cases and is sharp if $\mu_k\gg 1\,\forall k\in[1,\nu]$ and $M\gg\nu$. Exponentiating the θ_k in Eq. (3.1) and expanding the θ_k dependence of the right-hand side about $\{\theta_k^*\}$ yields

$$P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)=P^*(\,)\exp\left(-\frac{1}{2}\sum_{k=1}^{\nu}\frac{M^2}{\mu_k}\left\{\theta_k-\frac{\mu_k}{M}\right\}^2\right), \quad (3.5)$$

for $P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)$ near $\{\theta_k^*\}$ in the case $N=\nu$. For the case, $N>\nu$, the same expansion yields

$$P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)=P^*(\,)\exp\left[M\sum_{k=1}^{\nu}\left(\theta_k-\frac{\mu_k}{M}\right)-\frac{1}{2}\sum_{k=1}^{\nu}\frac{M^2}{\mu_k}\left(\theta_k-\frac{\mu_k}{M}\right)^2\right]. \quad (3.6)$$

In view of Eqs. (2.1) and (2.2), Eq. (3.6) can be rewritten as

$$P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)=P^*(\,)\exp\left[-M\sum_{k=\nu+1}^{N}\theta_k-\frac{1}{2}\sum_{k=1}^{\nu}\frac{M^2}{\mu_k}\left(\theta_k-\frac{\mu_k}{M}\right)^2\right]. \quad (3.7)$$

From Eq. (3.6) or, especially, Eq. (3.7), one sees explicitly that the maximum at $\{\theta_k^*\}$ for $N=\nu$ is replaced by a supremum for $N>\nu$. The supremum is sharp around the nonzero θ_k^*, provided M and all the nonzero μ_k individually are large.

For $M\gg N$, $N>\nu$, and $\mu_k\gg 1\,\forall k$, the θ_k^* become the values obtained by the Monte Carlo measurement of the volumes of the basins of attraction. In such a measurement, one is concerned with the probability density $P(\theta_1,\ldots,\theta_N/\mu_1,\ldots,\mu_N,M,N)$, which, according to Bayes' rule, is proportional to $P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)$, which is given by Eq. (3.1). Equation (3.1) can be replaced by Eq. (3.5) in the limit

$$\lim_{\{\mu_1,\ldots,\mu_N\}\to\infty}P(\theta_1,\ldots,\theta_N/\mu_1,\ldots,\mu_N,M,N)=\prod_{k=1}^{N}\delta\left(\theta_k-\frac{\mu_k}{M}\right). \quad (3.8)$$

This limit forms the basis for the Monte Carlo measurement of volumes of the basins of attraction. We note here that $\delta(.)$ is the Dirac δ function.

The expressions (3.5) and (3.6) can be used to evaluate $P(\mu_1,\ldots,\mu_N/M,N)$ via Eq. (3.2) when M is large enough so that the vicinity of $\{\theta_k^*\}$ within which each is valid dominates the integrals in Eq. (3.2). The first step is to introduce the δ function $\delta(\Sigma_{k=1}^N\theta_k-1)$ as a factor into the integrand of Eq. (3.2) so that the integration over the $\{\theta_k\}$ can be extended to the entire open simplex $\theta_k\in(0,1)$. The next step is to introduce the Fourier representation of the δ function,

$$\delta\left(1-\sum_{k=1}^{N}\theta_k\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(i\tau\left\{1-\sum_{k=1}^{N}\theta_k\right\}\right)d\tau, \quad (3.9)$$

and invert the order of the integration. The limits on the integrations over the $\{\theta_k\}$ are extended to the domain $(-\infty,\infty)$ for $k\leqslant\nu$ because of the presumption of the sharpness of the integrand. The form resulting for the integrals in Eq. (3.2) is then

$$P(\mu_1,\ldots,\mu_N/M,N)=P^*(\,)\,\frac{1}{2\pi}\sum_{k=1}^{\nu}\left(\frac{2\pi\mu_k}{M^2}\right)^{1/2}\int_{-\infty}^{\infty}\left\{\frac{\exp(-i\tau)-1}{-i\tau}\right\}^{N-\nu}\times\exp\left(-\frac{1}{2M}\{\tau+iM\}^2\right)d\tau, \quad (3.10)$$

which simplifies to

$$P(\mu_1,\ldots,\mu_N/M,N) = P^*(\,)\,\frac{1}{2\pi}\prod_{k=1}^{\nu}\left(\frac{2\pi\mu_k}{M^2}\right)^{1/2} \times\int_{-\infty}^{\infty}\exp\left(-i\,\frac{N-\nu}{2}\,\tau\right)\left\{\frac{2\sin(\tau/2)}{\tau}\right\}^{N-\nu} \times\exp\left(-\frac{1}{2M}\{\tau+iM\}^2\right)d\tau, \tag{3.11}$$

and one may thus obtain the following upper bound for $P(\mu_1,\ldots,\mu_N/M,N)$,

$$P(\mu_1,\ldots,\mu_N/M,N)\leqslant P^*(\,)\exp(M/2)\left(\frac{M}{2\pi}\right)^{1/2} \times\exp\left(-\frac{1}{2}\left\{1+\frac{N-\nu}{2}\right\}^2 M\right) \times\prod_{k=1}^{\nu}\left(\frac{2\pi\mu_k}{M^2}\right)^{1/2}. \tag{3.12}$$

The condition of equality in Eq. (3.12) holds for $N=\nu$.

C. A stopping rule

The probability that the function $f(x)$ has exactly ν minima is given by

$$P(\nu/\mu_1,\mu_2,\ldots,\mu_\nu;M)=\frac{P(\mu_1,\ldots,\mu_\nu/M,\nu)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)+\sum_{\omega=1}^{\infty}P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}=\frac{1}{1+\varepsilon}. \tag{3.13}$$

As $M\to\infty$, the rate of convergence of the quantity ε, defined by

$$\varepsilon=\sum_{\omega=1}^{\infty}\frac{P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)}, \tag{3.14}$$

to zero provides a stopping criterion for this sequential-search method. Because

$$\frac{P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)}\leqslant\exp\left(-\left\{\frac{\omega}{2}+\frac{\omega^2}{8}\right\}M\right), \tag{3.15}$$

we have

$$\varepsilon\leqslant\sum_{\omega=1}^{\infty}\exp\left(-\left\{\frac{\omega}{2}+\frac{\omega^2}{8}\right\}M\right). \tag{3.16}$$

It may be seen that $\varepsilon\cong 0.002$ for $M\cong 10$ with $\omega=1$ dominating, a result which holds true *independent* of the dimension of the search space of the function $f(x)$. Recall that the limits on the integrations over the $\{\theta_k\}$ in Eq. (3.2) were extended to the domain $(-\infty,\infty)$ for $k\leqslant\nu$ because of the presumption of the sharpness of the integrand. This could lead to an overestimate of ε when the individual μ_k are not sufficiently large. Note however that the convergence quantifying parameter ε is now a function only of M, the total number of local searches conducted, but not of ν the total number of discovered optima at any stage of the search process and this could lead to an underestimate of ε when the total number of minima N, of the function is very large. Clearly, a more general analysis is needed to incorporate the latter into the stopping criterion.

IV. A MORE GENERAL ANALYSIS

The preceding analysis was based on an expansion about the Monte Carlo estimate of the volumes of the basins of attraction which approaches the true volumes of the basins in the limit $\{\mu_1,\mu_2,\ldots,\mu_N\}\to\infty$. Instead, a more general analysis is possible. It is possible to compute

$$P(\mu_1,\ldots,\mu_N/M,N)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(i\tau) \times\left(\frac{\exp(-i\tau)-1}{-i\tau}\right)^{N-\nu}M!\prod_{k=1}^{\nu}\frac{1}{\mu_k!} \times\int_0^1\exp(i\tau\theta)\,\theta^{\mu_k}d\theta\,d\tau. \tag{4.1}$$

explicitly by integrating over the entire space of the volume fractions, the N-dimensional unit simplex, again with no presumption of prior knowledge of the $\{\theta_k\}$. Equation (4.1) is retained in the following form after simplification:

$$P(\mu_1,\ldots,\mu_N/M,N)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(i\tau)\exp\left(-i\,\frac{N-\nu}{2}\,\tau\right) \times\left(\frac{2\sin(\tau/2)}{\tau}\right)^{N-\nu}M!\prod_{k=1}^{\nu}\frac{1}{\mu_k!}\,\frac{d^{\mu_k}}{dx^{\mu_k}} \times\left(\frac{\exp(x)-1}{x}\right)\Bigg|_{x=-i\tau}d\tau. \tag{4.2}$$

It is then readily verified that

$$P(\mu_1,\ldots,\mu_N/M,N) \leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(i\left\{1-\frac{N-\nu}{2}\right\}\tau\right)M! \times\prod_{k=1}^{\nu}\frac{1}{\mu_k!}\frac{d^{\mu_k}}{dx^{\mu_k}}\left(\frac{\exp(x)-1}{x}\right)\Bigg|_{x=-i\tau} d\tau. \quad (4.3)$$

The parameter ε quantifying convergence requires the computation of

$$\frac{P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)},$$

which is now given by

$$\frac{P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)}=\frac{\int_{-\infty}^{\infty}\exp(i\tau)\exp\left(-i\,\frac{\omega}{2}\,\tau\right)\left(\frac{2\sin(\tau/2)}{\tau}\right)^{\omega}\prod_{k=1}^{\nu}\frac{d^{\mu_k}}{dx^{\mu_k}}\left(\frac{\exp(x)-1}{x}\right)\Big|_{x=-i\tau} d\tau}{\int_{-\infty}^{\infty}\exp(i\tau)\prod_{k=1}^{\nu}\frac{d^{\mu_k}}{dx^{\mu_k}}\left(\frac{\exp(x)-1}{x}\right)\Big|_{x=-i\tau} d\tau}, \quad (4.4)$$

and thereby

$$\varepsilon=\frac{\int_{-\infty}^{\infty}\exp(i\tau)\chi_g(i\tau)\prod_{k=1}^{\nu}\frac{d^{\mu_k}}{dx^{\mu_k}}\left(\frac{\exp(x)-1}{x}\right)\Big|_{x=-i\tau} d\tau}{\int_{-\infty}^{\infty}\exp(i\tau)\prod_{k=1}^{\nu}\frac{d^{\mu_k}}{dx^{\mu_k}}\left(\frac{\exp(x)-1}{x}\right)\Big|_{x=-i\tau} d\tau}, \quad (4.5)$$

where

$$\chi_g(i\tau)=\frac{1}{\exp(i\tau/2)-2\sin(\tau/2)/\tau-1}. \quad (4.6)$$

Equations (4.4) and (4.5) require the specification of ω, ν, M, and $\{\mu_k\}$ where $\Sigma_{k=1}^{\nu}\mu_k=M$, $\mu_k\geqslant 1$, $k\leqslant\nu$, and $M\geqslant\nu$. The integrals in Eq. (4.5) can readily be evaluated by the Gauss-Hermite quadrature [16]. Equation (4.5) has been derived from the most general analysis of sequential random local searches possible in the absence of prior information about the θ_k. Note that the convergence quantifying parameter ε will now be a function of ν, the total number of local optima discovered at any given stage of the search process, as well as M, the total number of local searches conducted.

V. AN ANALYSIS BASED ON DIRECT COMPUTATION OF VOLUME FRACTIONS

The analysis of the preceding section presupposes that we do not have any knowledge of the volume fractions of the basins of attraction derived from the process of local searches on the topography of the function. If information about the volume fractions of the discovered basins of attraction $\{\theta_k\}$ were to be generated, a sharper stopping criterion than that due to Eq. (4.4) could be derived. Numerical estimation of the actual volume fractions $\{\theta_k\}$ can only be possible up to a certain accuracy and is not easy. In fact the problem of numerically computing the volume of a convex body, of locating the global optimum of a multimodal function, and of determining the separatrices between the basins of attraction for a given function all form an equivalence class of *NP-complete* problems [17,18].

Typically it will only be possible to obtain lower and upper bounds, $\theta_k^{\min}$ and $\theta_k^{\max}$, on the volume fractions such that $\{\theta_k\in[\theta_k^{\min},\theta_k^{\max}]\}$. Treating the $\{\theta_k\}$ as independent random variables distributed on the hyperplane (2.1) over the open simplex $\theta_k\in(0,1)$ subject to these bounds, the joint probability density for $\{\theta_k\}$ is given by

$$P(\theta_1,\ldots,\theta_N)=\prod_{k=1}^{N}p_k(\theta_k). \quad (5.1)$$

TABLE I. Performance of the method based on the more general analysis.

Test Problem	Number of variables	Number of minima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
Goldstein-Price	2	4	17	396	1.0×10^{-6}	2.4×10^{-5}
Branin	2	3	16	376	1.0×10^{-6}	4.5×10^{-5}
Hartman-3	3	3	16	352	1.0×10^{-6}	4.5×10^{-5}
Hartman-6	6	3	12	253	1.0×10^{-6}	5.5×10^{-5}

Here for the ν discovered basins of attraction one may assume a uniform probability density for the volume fractions, within the numerically computed estimates of $\theta_k^{\min}$ and $\theta_k^{\max}$,

$$p_k(\theta_k)=\frac{1}{\theta_k^{\max}-\theta_k^{\min}},\quad \theta_k\in(\theta_k^{\min},\theta_k^{\max}),\quad \forall k\in[1,\nu], \tag{5.2}$$

and zero outside. Let θ_u be given by

$$\theta_u=1-\sum_{l=1}^{\nu}\theta_l^{\min}. \tag{5.3}$$

Then, for the $N-\nu$ undiscovered basins of attraction, the following most conservative uniform probability density function may be adopted for the relative volume fractions:

$$\begin{aligned} p_k(\theta_k)&=\frac{1}{\theta_u},\quad \theta_k\in(0,\theta_u)\quad \forall k\in[\nu+1,N]\\ &=0,\quad \theta_k\in(\theta_u,1)\quad \forall k\in[\nu+1,N]. \end{aligned} \tag{5.4}$$

Incorporating Eq. (5.1)–(5.4) in the analysis of the quantifying parameter ε, for the convergence of our *sequential-search* method, Eq. (4.5) now reads

$$\varepsilon=\frac{\displaystyle\int_{-\infty}^{\infty}\exp(i\tau)\chi_b(i\tau\theta_u)\prod_{k=1}^{\nu}\int_{\theta_k^{\min}}^{\theta_k^{\max}}\exp(-i\tau\theta)\,\frac{\theta^{\mu_k}}{\theta_k^{\max}-\theta_k^{\min}}\,d\theta\,d\tau}{\displaystyle\int_{-\infty}^{\infty}\exp(i\tau)\prod_{k=1}^{\nu}\int_{\theta_k^{\min}}^{\theta_k^{\max}}\exp(-i\tau\theta)\,\frac{\theta^{\mu_k}}{\theta_k^{\max}-\theta_k^{\min}}\,d\theta\,d\tau}. \tag{5.5}$$

We note that

$$\chi_b(i\tau\theta_u)=\frac{i\tau\theta_u}{1-\exp(-i\tau\theta_u)}. \tag{5.6}$$

All of the integrals over infinite domains can be evaluated numerically using the Gauss-Hermite quadrature while those over finite domains can be evaluated analytically [19] or numerically using the Gauss-Chebyshev quadrature.

Ellipsoidal lower bounds on the volume fractions of the basins of attraction

Procedures for computing bounds on the volume fractions of the basins with reasonable computational complexity can only be heuristic in design. The problem of computing the upper bound on the volume fraction of a basin is tractable only via a crude Monte Carlo estimation which itself would require too many objective function evaluations. For the lack of a better procedure we thus advocate setting the upper bounds $\theta_k^{\max}$ to θ_u. However, it is fairly straightforward to estimate lower bounds of the basin volumes.

Let us assume that the configuration space is of dimension n. Leading to a local minimum x^* will be a set of points $\{x_1,x_2,\ldots,x_m\}$ obtained during the process of conducting local searches from a uniformly distributed sample of points in the configuration space. This set includes points considered at each step in the location of the *Cauchy* point along a *Cauchy* arc of descent. Let the function take the values $\{f_1,f_2,\ldots,f_m\}$ at the points $\{x_1,x_2,\ldots,x_m\}$. Furthermore, let its value and Hessian at the local minimum x^* be, respectively, denoted by f^* and H^{*-1}. Let $\overline{f}$ denote the smallest, next to smallest or median of the values in the set $\{f_1,f_2,\ldots,f_m\}$. Choose the ellipsoid

$$E_n=\{x|(x-x^*)^T H^{*-1}(x-x^*)\leqslant\overline{f}-f^*\} \tag{5.7}$$

centered such that it just contains the maximum number of feasible points $\{x_1,x_2,\ldots,x_m\}$ generated during the local search. Its volume could serve as an estimate for the lower bound of the basin of attraction pertaining to the local minimum x^*. The volume of this ellipsoid is computed by considering an ellipsoid as an affine transformation of a hypersphere.

Denote by

$$\Omega_n=\{x|x^T x\leqslant 1\} \tag{5.8}$$

the hypersphere in n dimensions. Its volume is given by the well-known formula

$$V(\Omega_n)=\frac{\pi^{n/2}}{\Gamma\left(\frac{n+2}{2}\right)}. \tag{5.9}$$

We are now ready to write down the formula for the volume of the ellipsoid given by Eq. (5.7). It is given by

$$V(E_n)=V(\Omega_n)|\det(Q)|(\overline{f}-f^*)^{n/2}, \tag{5.10}$$

where the matrix Q is obtained from the relation

$$H^*=QQ^T. \tag{5.11}$$

Alternatively, the volume of the convex hull encompassing the points obtained during the local searches could have been used as an estimate of the lower bound of the basin volume fraction, but the method of ellipsoids is preferred over using convex hulls because it delivers a greater lower bound on the basin volume fraction.

TABLE II. Performance of the method with ellipsoidal estimation of the lower bounds of volume fractions.

Test Problem	Number of variables	Number of minima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
Goldstein-Price	2	4	17	396	1.0×10^{-6}	2.4×10^{-5}
Branin	2	3	16	376	1.0×10^{-6}	4.5×10^{-5}
Hartman-3	3	3	13	278	1.0×10^{-6}	2.9×10^{-4}
Hartman-6	6	3	12	253	1.0×10^{-6}	5.5×10^{-4}

VI. THE PROBABILITY OF MISSING THE GLOBAL MINIMUM

An important issue is the probability of missing the global minimum. Let $P(\omega/\mu_1,\ldots,\mu_\nu,M)$ be the probability that there are ω undiscovered minima after ν minima have been discovered in M local searches on the function $f(x)$. Then the conditional probability that one of the ω undiscovered minima is the global minimum is given by $\omega/(\nu+\omega)$. Let $P(\text{miss}/\mu_1,\ldots,\mu_\nu,M)$ denote the probability of missing the global minima having discovered ν minima in M searches. Then we may write

$$P(\text{miss}/\mu_1,\ldots,\mu_\nu,M)=\sum_{\omega=1}^{\infty}\frac{\omega}{\omega+\nu}P(\omega/\mu_1,\ldots,\mu_\nu,M)$$
$$=\sum_{\omega=1}^{\infty}\frac{\omega}{\omega+\nu}\times\frac{P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\nu+\omega)}{P(\mu_1,\ldots,\mu_\nu/M,\nu)}\frac{\varepsilon}{1+\varepsilon}. \tag{6.1}$$

But, from Bayes' theorem $P(\omega/\mu_1,\ldots,\mu_\nu,M)$ is proportional to $P(\mu_1,\ldots,\mu_{\nu+\omega}/M,\ \nu+\omega)$ and we may write generically

$$P(\omega/\mu_1,\ldots,\mu_\nu,M)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(i\tau)\left(\frac{\exp(-i\tau)-1}{-i\tau}\right)^{\omega}\times F(\tau;\{\mu_1,\ldots,\mu_\nu\},\{\theta_1,\ldots,\theta_\nu\})d\tau. \tag{6.2}$$

Here the functional form of $F(\tau;\{\mu_1,\ldots\mu_\nu\},\{\theta_1,\ldots,\theta_\nu\})$ depends on the method of computing the volume fractions of the basins of attraction. We thus have for the probability of missing the global minimum

$$P(\text{miss}/\mu_1,\ldots,\mu_\nu,M)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(i\tau)S(\tau;\nu)\times F(\tau;\{\mu_1,\ldots,\mu_\nu\},\{\theta_1,\ldots,\theta_\nu\})d\tau, \tag{6.3}$$

where $S(\tau;\nu)$ is given by

$$S(\tau;\nu)=\sum_{\omega=1}^{\infty}\frac{\omega}{\omega+\nu}\left(\frac{\exp(-i\tau)-1}{-i\tau}\right)^{\omega}. \tag{6.4}$$

It is straightforward to show that

$$S(\tau;\nu)=\frac{y}{1-y}+\frac{\nu}{y^\nu}\ln(1-y)+\frac{\nu}{y^\nu}\sum_{\omega=1}^{\nu}\frac{y^\omega}{\omega}, \tag{6.5}$$

where $y=[1-\exp(-i\tau)]/i\tau$, and thus Eq. (6.3) may be readily evaluated via the Gauss-Hermite quadrature. From Eqs. (6.1) and (3.13) it follows that

$$\frac{1}{(1+\nu)}\frac{\varepsilon}{(1+\varepsilon)}<P(\text{miss}/\mu_1,\ldots,\mu_\nu,M)<\frac{\varepsilon}{(1+\varepsilon)}<\varepsilon \tag{6.6}$$

and $P(\text{miss}/\mu_1,\ldots,\mu_\nu,M)<\varepsilon$ provides a less conservative stopping rule than ε itself. Accordingly, we shall use ε to set the stopping rule in the following.

The development in this section can be used to assess the likelihood of missing the global minimum provided the func-

TABLE III. Performance of simulated annealing and the Bayesian method of Boender and Rinooy Kan.

Test Problem	Number of variables	Number of minima	Number of function evaluations required by simulated annealing	Number of function evaluations required by the method of Boender and Rinooy Kan
Goldstein-Price	2	4	563	721
Branin	2	3	505	683
Hartman-3	3	3	1459	633
Hartman-6	6	3	4648	318

TABLE IV. The four best optimal solutions for the bifunctional catalyst design problem.

	No. 1	No. 2	No. 3	No. 4
$u(0)$	0.6661	0.6651	0.6637	0.9
$u(1)$	0.6734	0.6721	0.90	0.6724
$u(2)$	0.6764	0.9	0.9	0.6755
$u(3)$	0.9	0.9	0.9	0.9
$u(4)$	0.9	0.9	0.9	0.9
$u(5)$	0.9	0.9	0.9	0.9
$u(6)$	0.9	0.9	0.9	0.9
$u(7)$	0.9	0.9	0.9	0.9
$u(8)$	0.9	0.9	0.9	0.9
$u(9)$	0.9	0.9	0.9	0.9
objective function	10.0942×10^{-3}	10.0527×10^{-3}	9.9047×10^{-3}	9.8805×10^{-3}

tion does not exhibit the pathology of a global minimum with zero volume fraction such as the Dirac δ function.

VII. PERFORMANCE AND COMPARATIVE ANALYSIS ON TEST PROBLEMS

In this section we present the performance of our sequential random search method on a variety of test problems. Comparisons with the Bayesian multistart method of Boender and Rinooy Kan [10] and the simulated annealing method of Dekkers and Aarts [2] are also presented. The test problems considered range in complexity from simple unconstrained optimization of algebraic functions, to the optimal control of a catalytic reactor and to the conformation problem of an atomic microcluster. The yardstick for performance of each method is the total number of function computations required during the execution of the method. The local minimizations were performed using a conjugate-gradient method [20]. In implementing the Bayesian multistart method of Boender and Rinooy Kan [10], we used the stopping criterion that the expected volume fraction of the undiscovered regions of attraction of the function $\langle\Theta\rangle$ be no larger than 0.005. Appendix B provides a sketch of the relevant details of their method. The details of the simulated annealing method implemented by us are identical to those of the one due to Dekkers and Aarts [2]. Appendix C provides a sketch of this annealing method. For the purposes of differentiating among the various stopping criteria of our method we will denote by $\varepsilon_{\text{Monte Carlo}}$, $\varepsilon_{\text{general}}$, and $\varepsilon_{\text{ellipsoidal}}$, the stopping criteria due to the three different analyses presented earlier. It was found that the stopping criterion due to $\varepsilon_{\text{Monte Carlo}}$ is unreliable in general, and in implementations of our method we imposed the stopping criterion that the probability of undiscovered minima be no larger than 1.0×10^{-6}, i.e., $\varepsilon_{\text{general}}$ or $\varepsilon_{\text{ellipsoidal}}\leqslant1.0\times10^{-6}$.

In our analysis, $\langle\Theta\rangle$ is given by

$$\langle\Theta\rangle=\sum_{\omega=1}^{\infty}\int_0^1\Theta P(\Theta/\omega)P(\omega/\mu_1,\ldots,\mu_\nu,M)d\Theta, \tag{7.1}$$

where Θ is the unknown undiscovered volume fraction and $P(\Theta/\omega)$ is its conditional probability density given ω undiscovered basins. Since $\Theta\in(0,1)$ holds, Eq. (7.1) can be replaced by the inequality

$$\langle\Theta\rangle\leqslant\frac{\varepsilon}{1+\varepsilon}. \tag{7.2}$$

Accordingly, choosing $\langle\Theta\rangle\leqslant0.005$ is equivalent to choosing $\varepsilon_{\text{general}}$ or $\varepsilon_{\text{ellipsoidal}}>0.005$ as a stopping criterion. Thus our test of the method of Boender and Rinooy Kan is much less stringent than that of our own methods.

A. Four classical test problems from the global optimization literature

We consider four standard test functions [21–23] for global optimization. The first problem is the Goldstein-Price problem. It is given by

$$\begin{aligned}f(x_1,x_2)=&[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2\\&+6x_1x_2+3x_2^2)][30+(2x_1-3x_2)^2\\&\times(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)]\end{aligned} \tag{7.3}$$

TABLE V. Performance of the method based on the more general analysis.

Test Problem	Number of variables	Number of minima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
Bifunctional Catalyst Design Problem	10	25	117	2942	1.0×10^{-6}	0.0

TABLE VI. Performance of the method with ellipsoidal estimation of the lower bounds of volume fractions.

Test Problem	Number of variables	Number of minima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
Bifunctional catalyst Design problem	10	25	109	2691	1.0×10^{-6}	0.0

and possesses four minima in the region $\{-2\leqslant x_1,\ x_2\leqslant 2\}$ over which it is to be minimized. The global minimum occurs at $(0,-1)$ and has the value of 3.

The second problem is the Branin problem. It is given by

$$f(x_1,x_2)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2 +10\left(1-\frac{1}{8\pi}\right)\cos x_1+10 \tag{7.4}$$

and possesses three minima all of which are global in the region $\{-5\leqslant x_1\leqslant 10,0\leqslant x_2\leqslant 15\}$ over which it is to be minimized.

The last two functions belong to the Hartman family of problems and are denoted as $H3$ and $H6$. The Hartman family of functions is given in general by

$$f(x_1,x_2,\ldots,x_n)=-\sum_{i=1}^{l} c_i\exp\left(-\sum_{j=1}^{n} a_{ij}(x_j-p_{ij})^2\right). \tag{7.5}$$

In the $H3$ test function $l=4$ and $n=3$. In the $H6$ test function $l=4$ and $n=6$. The parameters $\{c_i\}$, $\{a_{ij}\}$, and $\{p_{ij}\}$ for each of these problems are given in Appendix D. Both the three-dimensional $H3$ function and the six-dimensional $H6$ function have four minima in the region $\{x\in R^n|0\leqslant x_j\leqslant 1,\ l\leqslant j\leqslant n\}$ over which they have to be minimized. The location of the ith local minimum is approximately p_{ii} and its value is approximately $-c_i$. One may also note that a_{ij} is proportional to the jth eigenvalue of the Hessian at the ith local minima.

Tables I and II exhibit the performance of our method with the stopping criteria, respectively, derived from the more general analysis of Sec. IV and the analysis of Sec. V with explicit incorporation of lower bounds into the formulation using volumes of ellipsoids. Table III depicts the performance of the Bayesian method of Boender and Rinooy Kan and the simulated annealing method of Dekkers and Aarts. Also presented for each test problem is the value of $\varepsilon_{\text{Monte Carlo}}$ corresponding to the stopping set of parameters obtained for $\varepsilon_{\text{general}}$ and $\varepsilon_{\text{ellipsoidal}}$.

B. An optimal control problem in the design of bifunctional catalysts

We now consider a problem [14] in the conceptual design of a bifunctional catalytic reactor for the conversion of methylcyclopentane to benzene. It concerns the optimization over a trajectory where the dynamics are described by a set of coupled nonlinear ordinary differential equations. The function to be maximized represents the concentration of benzene at the end of a tubular reactor and is described as follows:

$$\text{minimize}\ \ J(u(t))=-x_7(t_f),$$

subject to the constraints

$$\begin{aligned}
&\frac{d}{dt}x_1(t)=-k_1x_1(t),\quad x_1(t)=1.0,\\
&\frac{d}{dt}x_2(t)=k_1x_1(t)-(k_2+k_3)x_2(t)+k_4x_5(t),\quad x_2(t)=0.0,\\
&\frac{d}{dt}x_3(t)=k_2x_2(t),\quad x_3(t)=0.0,\\
&\frac{d}{dt}x_4(t)=-k_6x_4(t)+k_5x_5(t),\quad x_4(t)=0.0,\\
&\frac{d}{dt}x_5(t)=k_3x_2(t)+k_6x_4(t)-(k_4+k_5+k_8+k_9)x_5(t)\\
&\qquad +k_7x_6(t)+k_{10}x_7(t),\quad x_5(t)=0.0,\\
&\frac{d}{dt}x_6(t)=k_8x_5(t)-k_7x_6(t),\quad x_6(t)=0.0,\\
&\frac{d}{dt}x_7(t)=k_9x_5(t)-k_{10}x_7(t),\quad x_7(t)=0.0,
\end{aligned} \tag{7.6}$$

TABLE VII. Performance of simulated annealing and the Bayesian method of Boender and Rinooy Kan.

Test Problem	Number of variables	Number of minima	Number of function evaluations required by simulated annealing	Number of function evaluations required by the method of Boender and Rinooy Kan
Bifunctional catalyst Design problem	10	25	26 471	3856

where

$$k_i=\sum_{j=1}^{4} c_{ij}[u(t)]^{j-1}, \quad i=1,2,\ldots,10,$$

$$0\leqslant t\leqslant t_f=2000 \text{ gm h/mol},$$

and

$$0.60\leqslant u(t)\leqslant 0.90.$$

Here $x_1(t)$ denotes moles of methylcyclopentane, the reactant, while $x_7(t)$ denotes moles of benzene, the desired product. Additionally, we may note that

$$u=\frac{\text{(mass of hydrogenation component)}}{\text{(total mass of catalyst)}},$$

$$t=\frac{\text{(total mass of catalyst up to a given section of the reactor)}}{\text{(molar flow rate of methylcyclopentane into the reactor)}},$$

and

$$t_f=\frac{\text{(total mass of the catalyst in the reactor)}}{\text{(molar flow rate of methylcyclopentane into the reactor)}}.$$

The coefficients $\{c_{ij}\}$ are given in Appendix E. To obtain the solution of this optimal control problem we convert it to an optimal parameter estimation problem by dividing the time interval into ten equal sections each of length 200 gm h/mol and seek the ten piecewise constant controls $[u(0),u(1),\ldots,u(9)]$ that maximize the benzene concentration at the reactor outlet. This ten variable parameter estimation problem possesses 25 local optima [14], one of which is the global optimum. The globally optimal solution and its three nearest suboptimal solutions are shown in Table IV.

Tables V and VI exhibit the performance of our method with the stopping criteria, respectively, derived from the

TABLE VIII. Lennard-Jones atomic-microcluster problem.

Number of particles	Number of local optima	Globally maximum binding energy
3	1	3.0
4	1	6.0
5	1	9.104
6	2	12.712
7	4	16.505
8	8	19.821
9	14	24.113
10	50	28.422
11	132	32.766
12	451	37.968
13	988	44.327
14	2497	47.845

TABLE IX. Performance of the method based on the more general analysis.

Number of particles in the Lennard-Jones Cluster	Number of variables	Number of optima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
3	6	1	20	611	1.0×10^{-6}	3.7×10^{-6}
4	9	1	20	557	1.0×10^{-6}	3.7×10^{-6}
5	12	1	20	548	1.0×10^{-6}	3.7×10^{-6}
6	15	2	31	1031	1.0×10^{-6}	3.8×10^{-9}
7	18	4	62	1871	1.0×10^{-6}	1.4×10^{-17}
8	21	8	114	2178	1.0×10^{-6}	0.0
9	24	14	197	5332	1.0×10^{-6}	0.0
10	27	50	702	21 859	1.0×10^{-6}	0.0
11	30	132	1839	42 524	1.0×10^{-6}	0.0
12	33	451	6339	19 677 9	1.0×10^{-6}	0.0
13	36	988	13 618	43 046 1	1.0×10^{-6}	0.0
14	39	2497	34 901	12 216 11	1.0×10^{-6}	0.0

TABLE X. Performance of the method with ellipsoidal estimation of the lower bounds of volume fractions.

Number of particles in the Lennard-Jones cluster	Number of variables	Number of optima	Number of searches	Number of function evaluations	Probability of missing undiscovered minima	$\varepsilon_{\text{Monte Carlo}}$
3	6	1	20	611	1.0×10^{-6}	3.7×10^{-6}
4	9	1	20	557	1.0×10^{-6}	3.7×10^{-6}
5	12	1	20	548	1.0×10^{-6}	3.7×10^{-6}
6	15	2	31	1031	1.0×10^{-6}	3.8×10^{-9}
7	18	4	59	1759	1.0×10^{-6}	9.6×10^{-17}
8	21	8	111	2116	1.0×10^{-6}	0.0
9	24	14	190	5137	1.0×10^{-6}	0.0
10	27	50	689	21 462	1.0×10^{-6}	0.0
11	30	132	1814	41 928	1.0×10^{-6}	0.0
12	33	451	6302	19 567 7	1.0×10^{-6}	0.0
13	36	988	13 584	42 933 6	1.0×10^{-6}	0.0
14	39	2497	34 828	12 188 93	1.0×10^{-6}	0.0

analysis about the more general analysis and the analysis with explicit incorporation of lower bounds into the formulation using volumes of ellipsoids. Table VII depicts the performance of the Bayesian method of Boender and Rinooy Kan and the simulated annealing method of Dekkers and Aarts. Also presented for each test problem is the value of $\varepsilon_{\text{Monte Carlo}}$ corresponding to the stopping set of parameters for $\varepsilon_{\text{general}}$ and $\varepsilon_{\text{ellipsoidal}}$.

C. An atomic-microcluster conformation problem

We consider the restricted problem of describing the ground state of a classical N-particle system interacting through the Lennard-Jones pair potential. The potential is given conventionally in reduced units

$$\nu_{LJ}(r)=r^{-12}-2r^{-6}, \tag{7.7}$$

where $\nu_{LJ}(r)$ is the potential in units of the well depth, and r is the interparticle distance in units of the distance at the potential minimum. The objective is then to find the global minimum of the potential energy hypersurface defined by

$$V_{LJ}(r^N)=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\nu_{LJ}(|r_i-r_j|) \tag{7.8}$$

subject to translational and rotational invariance of the cluster. This problem has $3N-6$ degrees of freedom. Hoare and McInnes [15] have shown that the number of local minima of the potential energy surface of a Lennard-Jones cluster becomes large growing faster than linearly in N. The Bayesian method of Boender and Rinooy Kan has been considered earlier for the location of the globally maximum binding energy of such clusters [24]. We consider the performance of our method on twelve clusters ranging from 3 to 14 particles. Table VIII shows the global minimum values for each of these clusters.

Tables IX and X exhibit the performance of our method with the stopping criteria, respectively, derived from the

TABLE XI. Performance of simulated annealing and the Bayesian method of Boender and Rinooy Kan.

Number of particles in the Lennard-Jones cluster	Number of variables	Number of optima	Number of function evaluations required by simulated annealing	Number of function evaluations required by the method of Boender and Rinooy Kan
3	6	1	19 855	651
4	9	1	28 247	588
5	12	1	33 261	546
6	15	2	47 459	1188
7	18	4	61 073	1921
8	21	8	69 719	2299
9	24	14	84 536	5564
10	27	50	91 329	22 268
11	30	132	10 354 4	43 334
12	33	451	30 114 3	19 821 9
13	36	988	58 726 1	43 702 6
14	39	2497	70 149 36	12 362 35

TABLE XII. Parameters for the $H3$ function.

i	a_{i1}	a_{i2}	a_{i3}	c_i	p_{i1}	p_{i2}	p_{i3}
1	3	10	30	1	0.368 9	0.1170	0.2673
2	0.1	10	35	1.2	0.469 9	0.4387	0.7470
3	3	10	30	3	0.109 1	0.8732	0.5547
4	0.1	10	35	3.2	0.038 15	0.5743	0.8828

more general analysis and the analysis with explicit incorporation of lower bounds into the formulation using volumes of ellipsoids, respectively. Table XI depicts the performance of the Bayesian method of Boender and Rinooy Kan and the simulated annealing method of Dekkers and Aarts. Also presented for each test problem is the value of $\varepsilon_{\text{Monte Carlo}}$ corresponding to the stopping set of parameters for $\varepsilon_{\text{general}}$ and $\varepsilon_{\text{ellipsoidal}}$.

VIII. CONCLUSIONS: FUTURE WORK

Our Bayesian sequential random search method has been shown to be a reliable tool for locating the global minimum of multimodal functions of continuous variables. Within the framework of the general analysis and the ellipsoidal estimation of basin lower bounds its performance compares well with those of the Bayesian method of Boender and Rinooy Kan [10] and the simulated annealing method of Dekkers and Aarts [2].

In studying the results of the computations described in Sec. VII, certain definitive patterns can be identified. Due to the combinatorial nature of the transitions in the search space, simulated annealing possesses a computational complexity which is exponential in the dimensions of the search space independent of the total number of minima supported by the objective function. On the other hand, both the Bayesian method described in Sec. III–VI, and the Bayesian method of Boender and Rinooy Kan possess computational complexities that are strong functions of the total number of minima supported by the objective function while growing algebraically, and hence weakly, as a function of the dimensions of the search space. Recalling the identities of Eqs. (7.1) and (7.2), we note that the stopping criterion used in the implementations of the Bayesian method described in this paper is much more stringent than that used in the implementation of the Bayesian method of Boender and Rinooy Kan. Tables I–XI attest to the fact that the latter method requires more function evaluations than the former to locate the global minimum with sufficient assurance. Furthermore, the use of identical stopping criteria for both methods would only serve to amplify the difference between the number of function evaluations required by the latter and that required by the former.

TABLE XIII. Parameters for the $H6$ function.

i	a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}	a_{i6}	c_i
1	10	3	17	3.5	1.7	8	1
2	0.05	10	17	0.1	8	14	1.2
3	3	3.5	1.7	10	17	8	3
4	17	8	0.05	10	0.1	14	3.2

TABLE XIV. Parameters for the $H6$ function (continued).

i	p_{i1}	p_{i2}	p_{i3}	p_{i4}	p_{i5}	p_{i6}
1	0.1312	0.1696	0.5569	0.0124	0.8383	0.5886
2	0.2329	0.4135	0.8307	0.3736	0.1004	0.9991
3	0.2348	0.1451	0.3522	0.2883	0.3047	0.6650
4	0.4047	0.8828	0.8732	0.5743	0.1091	0.0381

With respect to the Bayesian method described in this paper, two stopping criteria have been considered: that due to the general analysis of Sec. IV and that due to the ellipsoidal estimation of basin volume fractions described in Sec. V. As seen in Tables I–III for the four classical problems from the global optimization literature, both stopping criteria require an almost equal number of function evaluations for termination within an allowance of fluctuations in the number of local searches required. However, when one examines the results of the bifunctional catalyst design problem in Tables IV–VII and of the conformation problem of the Lennard-Jones clusters in Tables VIII–XI it is evident that knowledge of the basin volume fractions, although heuristic, serves to diminish the number of function evaluations required for the termination of the Bayesian algorithm. Thus, from the results of the various computations, it cannot be overemphasized that more accurate, rapid estimate of lower and upper bounds on basin volume fractions are required to obtain a sharper stopping criterion for the method.

Finally, we note that the Monte Carlo stopping criterion of Eq. (3.16) is clearly inadequate. It underestimates the total number of undiscovered minima when the total number of minima supported by the function is very large and when the total number of minima supported by the function is small, it overestimates the number of undiscovered minima of the function if the visitation numbers to the individual minima are not sufficiently large.

ACKNOWLEDGMENTS

Financial support for this work was provided by the Exxon Research and Engineering Co., Annandale, NJ, the Department of Chemical Engineering and Materials Science at the University of Minnesota, Minneapolis, MN, and the National Science Foundation under Grant No. NSF/CTS-9504827. Computational support was provided by the Minnesota Supercomputing Institute.

APPENDIX A: EVALUATING THE SUPREMUM OF $P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N)$

Consider the multinomial distribution given by

$$P(\mu_1,\ldots,\mu_N/\theta_1,\ldots,\theta_N,M,N) = M!\prod_{k=1}^{v}\frac{\theta_k^{\mu_k}}{\mu_k!} = \frac{M!}{\Pi_{k=1}^{v}\mu_k!}\exp\left(\sum_{k=1}^{v}\mu_k \ln \theta_k\right), \tag{A1}$$

TABLE XV. Coefficients for the rate constants.

i	c_{i1}	c_{i2}	c_{i3}	c_{i4}
1	0.002 918 487	−0.008 045 787	0.006 749 947	−0.001 416 647
2	9.509 97	−35.009 94	42.833 29	−17.333 3
3	26.820 93	−95.560 79	113.039 8	−44.299 97
4	208.724 1	−719.805 2	827.746 6	−316.665 5
5	1.350 005	−6.850 027	12.166 71	−6.666 689
6	0.0192 199 5	−0.079 453 20	0.110 566 6	−0.050 333 33
7	0.132 359 6	−0.469 625 5	0.553 932 3	−0.216 666 4
8	7.339 981	−25.273 28	29.933 29	−11.999 99
9	−0.395 053 4	1.679 353	−1.777 829	0.497 498 7
10	−0.000 025 046 65	0.010 058 54	−0.019 866 96	0.009 833 470

where $0 \leqslant \theta_k \leqslant 1$, $\forall k$. Introducing a Lagrange multiplier λ, this supremum is the sumpremum of

$$\sum_{k=1}^{v} (\mu_k \ln \theta_k - \lambda \theta_k) - \lambda \sum_{k=v+1}^{N} \theta_k, \tag{A2}$$

which has negative curvature with respect to θ_k, $1 \leqslant k \leqslant v$. Therefore (A2) is a maximum with respect to θ_k, $1 \leqslant k \leqslant v$ when

$$\theta_k = \theta_k^* = \frac{\mu_k}{\lambda}, 1 \leqslant k \leqslant v, \tag{A3}$$

and is a supremum when $\theta_k = 0$, $v+1 \leqslant k \leqslant N$ because $\lambda = M$ is positive.

APPENDIX B: THE BAYESIAN METHOD OF BOENDER AND RINOOY KAN

In Boender and Rinooy Kan's *multistart* method for global optimization [10] the configuration space is repeatedly sampled at batches of uniformly distributed points and local searches initiated therefrom. The probability of the aggregate of events resulting from their multistart method is given by a generalized multinomial distribution. Given a uniform prior distribution of local minima and a generalized multinomial distribution of the outcome $\{N,w\} = \{N_1, N_2, \ldots, N_w\}$, the posterior expected value of the total relative volume of observed regions of attraction, denoted by Ω, is given by

$$E(\Omega/\{N,w\}) = \frac{(N-w-1)(N+w)}{N(N-1)}, \quad N \geqslant w+2. \tag{B1}$$

Here w is the total number of local minima discovered in N local searches and $\{N_i, 1 \leqslant i \leqslant w\}$ the frequency of visits to the minima. The proof of this statement can be referred to in Boender and Rinooy Kan [10]. Following Maier [24], we set the stopping criterion for the Bayesian method of Boender and Rinooy Kan in our implementations such that $E(\Omega/\{N,w\}) \geqslant 0.995$.

APPENDIX C: THE SIMULATED ANNEALING METHOD OF DEKKERS AND AARTS

The algorithm of Dekkers and Aarts, as applied to a multivariate function $f(x)$, $x \in S \subseteq R^n$, is stated in text below; we may note here that the parameter T is the control parameter notionally equivalent to temperature in statistical physics and L is the length of the Markov chain which is fixed for a given instance of the problem.

In the simulated annealing method, an initial choice of the control parameter T and of a configuration point x is taken. Thereupon, L steps of the following Markovian process are undergone: A new point y in the configuration space is sampled; a transition $x \leftarrow y$ is made if the objective function $f(.)$ satisfies the relation $f(y) \leqslant f(x)$ or if the functional $\exp\{-[f(y)-f(x)]/T\}$ is greater than a random number drawn from a uniform distribution over the interval [0,1). After undergoing this Markovian process, a criterion is tested for lowering the cooling parameter T. If it is necessary to cool the system further, the Markovian process is reiterated with a new value of the cooling parameter. If it is determined that there is no further need to cool, the global optimum has been located.

For the cooling schedule we used the parameters prescribed by Dekkers and Aarts [2]: $\gamma_0 = 0.9$, $d = 0.1$, $e_s = 1.0 \times 10^{-4}$, and $L_0 = 10$. Initially, the temperature must be sufficiently large such that almost all transitions are accepted. This is achieved by generating a number of trials, say t_0, and requiring that the initial acceptance ratio γ_0 be close to unity, where the acceptance ratio is defined as the ratio of the number of accepted transitions to the number of proposed transitions. The initial value of the temperature is computed by the following expression:

$$T_0 = \langle \Delta f^+ \rangle \left(\frac{t_+}{t_+ \gamma_0 + (1-\gamma_0) t} \right)^{-1} \tag{C1}$$

where t_- and t_+, respectively, denote the number of trials with $[\Delta f_{yx} = f(y) - f(x)] \leqslant 0$ and $\Delta f_{yx} > 0$. (Δf^+) denotes the average of those values of Δf_{yx} for which $\Delta f_{yx} > 0$. The temperature is decreased according to

$$T \leftarrow T \left(1 + T \frac{\ln(1+d)}{3\sigma(T)} \right)^{-1}, \tag{C2}$$

where $\sigma(T)$ denotes the standard deviation of the values of the objective function of all the points of the Markov chain at temperature T. The constant d is called the distance parameter and determines the rate of decrement of the temperature. The simulated annealing algorithm is terminated when

$$\left| \frac{\partial}{\partial T} \langle f_s(x;T) \rangle \frac{T}{\langle f(x;T_0) \rangle} \right| < e_s . \tag{C3}$$

Here $\langle f(x;T_0) \rangle$ is the mean value of the objective function at the points found in the initial Markov chain, $\langle f_s(x;T) \rangle$ is the smoothed value of $\langle f(x;T) \rangle$ over a number of chains in order to reduce the fluctuations of $\langle f(x;T) \rangle$, and e_s is the stopping parameter. The length of the Markov chain is chosen to be

$$L = L_0 \dim(S) = L_0 n \tag{C4}$$

and a new point y is generated from a current point x according to

$$\begin{aligned} y &= LS(x), \quad \text{if } w > 0.75 \\ &= U(S), \quad \text{if } w \leqslant 0.75, \end{aligned} \tag{C5}$$

where w is a random number uniformly distributed in [0,1). $LS(x)$ is a local search procedure that generates a point y, not necessarily a local minimum, along a descent direction of x and $U(S)$ is a point drawn from the uniform distribution over the compact set S.

APPENDIX D: PARAMETERS FOR THE HARTMAN FAMILY OF FUNCTIONS

Table XII gives the parameters for the $H3$ function while Tables XIII and XIV give the parameters for the $H6$ function.

APPENDIX E: PARAMETERS FOR THE BIFUNCTIONAL CATALYST PROBLEM

The parameters for the rate constants for the bifunctional catalyst design problem are given in Table XV.

[1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[2] A. Dekkers and E. Aarts, Math. Program. **50**, 367 (1991).
[3] J. H. Holland, Sci. Am. July Issue, 66 (1992).
[4] D. A. Coley, Contemp. Phys. **37** (2), 145 (1996).
[5] C. A. Floudas and V. Visweswaran, Comput. Chem. Eng. **14**, 1397 (1990).
[6] V. Visweswaran and C. A. Floudas, Comput. Chem. Eng. **14**, 1419 (1990).
[7] J. F. Benders, Num. Math. **4**, 238 (1962).
[8] J. Mockus, *Bayesian Approach to Global Optimization* (Kluwer, Boston, 1989).
[9] R. Zielinski, Math. Program. **21**, 348 (1981).
[10] C. G. E. Boender and A. H. G. Rinooy Kan, Math. Program. **37**, 59 (1987).
[11] M. Piccioni and A. Ramponi, Optimization **21**, 697 (1990).
[12] C. G. E. Boender and A. H. G. Rinooy Kan, J. Global Optimization **1**, 331 (1991).
[13] M. H. De Groot, *Optimal Statistical Decisions* (McGraw-Hill, New York, 1970).
[14] R. Luus, J. Dittrich, and F. J. Keil, Can. J. Chem. Eng. **70**, 780 (1992).
[15] M. R. Hoare and J. McInnes, Adv. Phys. **32**, 791 (1983).
[16] M. J. D. Powell, *Approximation Theory and Methods* (Cambridge University Press, Cambridge, England, 1981).
[17] G. Elekes, Discrete Comput. Geometry **1**, 289 (1986).
[18] I. Barany and Z. Furedi, *Computing the Volume is Difficult*, Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing (ACM, New York, 1986), p. 442.
[19] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1952).
[20] S. G. Nash, SIAM J. Numer. Anal. **21** (4), 770 (1984).
[21] L. C. W. Dixon and G. P. Szego, *Towards Global Optimization* (North-Holland, Amsterdam, 1975), Vol. I.
[22] L. C. W. Dixon and G. P. Szego, *Towards Global Optimization* (North-Holland, Amsterdam, 1978), Vol. II.
[23] A. Torn and A. Zilniskas, *Global Optimization* (Springer-Verlag, Berlin 1989).
[24] R. S. Maier, Army High Performance Computing Research Center, University of Minnesota, Report No. 92-072, 1992 (unpublished).