A New Large-Scale Global Optimization Method and its Application to Lennard-Jones Problems

Richard H. Byrd, Elizabeth Eskow & Robert B. Schnabel*

CU-CS-630-92 November 1992

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

* This research was supported by NSF grant CCR-9101795, NSF grant CDA-8922510, AFOSR grant AFOSR-90-0109 and ARO grant DAAL03-91-G-0151

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **NOV 1992**	2. REPORT TYPE	3. DATES COVERED **00-00-1992 to 00-00-1992**
4. TITLE AND SUBTITLE **A New Large-Scale Global Optimization Method and its Application to Lennard-Jones Problems**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Computer Science,University of Colorado,Boulder,CO,80309**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**21**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR(S) AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE AGENCIES NAMED IN THE ACKNOWLEDGMENTS SECTION.

Abstract

We describe a new stochastic global optimization algorithm that is oriented towards solving large scale problems, and present the results of applying it to a class of problems in molecular chemistry. Our new algorithm incorporates some full-dimensional random sampling and local minimizations as in existing stochastic methods, but the keys to its success are two new phases that concentrate on selected small dimensional subproblems of the overall problem. These phases constitute a major portion of the computational effort of the new method, and represent a significant departure from existing stochastic methods. Computational results on Lennard-Jones problems show that the new method is considerably more successful than any other method that has tried to solve these problems without using prior knowledge of the solution structure in its algorithm, and that our method finds the presumptive global minimizer in all cases with up to 90 variables as well as in some larger cases. On the other hand, on most problems with over 100 variables, our method does not find as good a solution as has been found by the best special purpose methods for Lennard-Jones problems. This illustrates the inherent difficulty of large scale global optimization.

1 Introduction

We describe a new stochastic global optimization algorithm, and its application to a class of problems in molecular chemistry. The global optimization problem is to find the lowest minimizer of a nonlinear function f in a domain $D \in R^n$, where D is defined by upper and lower bounds on each variable. It is assumed that the global minimizer lies in the interior of D, and that the function f is twice continuously differentiable. The application considered in this paper is determining the structure of clusters of atoms whose potential energy is given by the sum of the pairwise interactions between atoms using the Lennard-Jones potential energy function. Problems in determining molecular structure lead to global optimization problems because the naturally occurring structure minimizes, or nearly minimizes, the potential energy of the system, and because typically such functions have very many local minimizers.

Because global optimization is a fairly young field, there exist many approaches to solving global optimization problems, but generally these approaches only have been used to solve small nonlinear problems, usually with 2-6 variables. Problems in molecular chemistry that are of practical interest may involve hundreds or thousands of atoms in three dimensional space; therefore it is important to develop methods that are capable of solving large dimensional problems. We will present results for problems with up to 165 variables, which is at least an order of magnitude more than the number of parameters generally used in the global optimization literature. More importantly, the techniques we develop, which combine computations in the full parameter space with phases that concentrate on small dimensional subproblems within the overall problem, are designed to be applicable to a wide variety of structured large scale global optimization problems.

The new algorithm is related to the stochastic methods of [24]. These methods combine random sampling over the domain space and local minimizations from selected sample points. Stochastic methods appear to be at least as efficient as other general global optimization methods in the comparisons reported in the literature, for example in [25], and also possess some probabilistic guarantee of success. We found that they were not able to solve the problems considered in this paper, however, nor even find reasonable approximations to the global minimizer, even after very large amounts of computation. This was basically due to the difficulty in dealing with a very large dimensional variable space and very many local minimizers.

The key new approach that we use in our method is to incorporate phases of the algorithm that focus on small dimensional subproblems within the overall large dimensional problem. In the context of the Lennard-Jones problems, these phases consist of series of steps that improve the positions of single atoms while the remainder of the molecular configuration remains fixed. This approach first is used to significantly reduce the function values of selected sample points before choosing start points for local minimizations. Second, this approach is used to generate improved local minimizers from the initial set of local minimizers. This latter phase constitutes a major portion of the computational effort of the calculations and represents a significant departure from previous stochastic methods. Similar techniques could be applied to any partially separable global optimization problem, although the criteria for selecting which subset of variables to choose for the small dimensional subproblems would be application dependent.

The remainder of the paper describes the new method, the molecular chemistry application, and the computational results obtained with the new method on Lennard-Jones problems. Section 2 discusses the Lennard-Jones molecular configuration problem. Section 3 briefly describes other global optimization approaches. Section 4 first gives a brief description of current stochastic methods, and explains why they proved unable to solve the large scale problems we consider. It then describes the new method, focusing on the new phases that concentrate on small dimensional subproblems. Section 5 presents and interprets the test results of this new method on a set of Lennard-Jones problems. Section 6 provides a brief summary and conclusions.

2 The Lennard-Jones Molecular Configuration Problem

The molecular configuration problem is the problem of finding the structure or relative positions of a cluster of atoms that minimizes the potential energy of the structure. In the experiments described in this paper, we consider clusters of N atoms in three dimensional space acting under a Lennard-Jones potential. These are problems that have often been used as model problems by the molecular chemistry community. They model the interaction of identical, spherically symmetric particles such as inert gases, and more importantly, their key features are indicative of certain interactions in more complex molecules. For a more detailed discussion of the Lennard-Jones problem and its applications, see e.g. [15].

Lennard-Jones problems assume that the potential energy of the molecule is given by the sum of the pairwise interactions between atoms, with these interactions being Van der Waals forces given by the Lennard-Jones 6-12 potential. That is, if we define the position of the molecule by

$$x = (x_1, x_2, ..., x_N)$$

where x_i is a three dimensional vector denoting the coordinates of the ith atom, then the potential energy function is

$$f(x) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} \left[\frac{1}{{d_{ij}}^{12}} - 2 * \frac{1}{{d_{ij}}^{6}} \right] \tag{2.1}$$

where d_{ij} is the Euclidean distance between x_i and x_j. In this formulation, the pairwise equilibrium distance is scaled to one and the pairwise minimum energy is scaled to -1.

The potential energy function (2.1) has numerous local minima, including many whose energy value is near that of the global minimum. For example, when $N = 13$, it is known that the number of local minima is 988, and Hoare has conjectured that the general number of local minimizers for an N-atom Lennard-Jones cluster is $O(e^{N^2})$ ([15]). Thus finding the minimum potential energy configuration of a Lennard-Jones cluster is a challenging global optimization problem.

In [23], upper bounds on the minimum potential energy for Lennard-Jones clusters of size $13 \leq N \leq 147$ have been presented. The method used to find these structures makes heavy use of the known approximate solution structure of optimal Lennard-Jones clusters, by utilizing a discrete lattice based search over multilayer icosahedral structures to generate starting configurations for local minimizations. [20] has reproduced the

Northby results, and found upper bounds for clusters of up to $N = 1000$, using a parallel version of the same lattice based search algorithm. The prior knowledge of the solution structure that both these methods utilize heavily is not expected to be available for many other large scale problems, including many other molecular configuration problems. In contrast, our method does not employ any knowledge about the solution to the problem in its algorithm. Instead, we introduce a general purpose large scale global optimization method that only makes use of the partially separable nature of the problem, a characteristic that is present in many large scale problems. This approach may be particularly useful in cases where no a priori knowledge of the solution is available.

3 Global Optimization Methods

Many approaches have been investigated for solving global optimization problems. For the general case where $f(x)$ can be any continuously differentiable nonlinear function, approaches that have been developed include trajectory methods (e.g. [?, 4]), deflation and tunneling methods (e.g. [11, 19]), interval arithmetic methods (e.g. [13]), piecewise approximation methods (e.g. [28]), genetic algorithms (e.g. [3, 16]), simulating annealing methods (e.g. [18]), and stochastic methods (e.g. [24]). Our approach is related to stochastic methods, which are described in Section 4.1. An excellent survey of global optimization methods for general nonlinear functions is contained in [25].

Recently, several global optimization techniques have been applied to various Lennard-Jones problems in two and three dimensional space. These include a new packet annealing method described in [27], a discrete descent technique discussed in [8], the tunneling method [2], and a convex decomposition method [21]. In addition, [17] conducted a comparison of simulated annealing, genetic algorithms and the Nelder Mead simplex method on two-dimensional Lennard-Jones problems, and suggested that for larger problems, a combination of simulated annealing and genetic algorithms may be effective. To our knowledge, none of these global optimization techniques have succeeded in consistently finding structures with energies as low as those found by the methods mentioned in the preceding section that utilize geometric, problem specific knowledge of the solution as a crucial component of the algorithm, even for problems with 10-20 atoms. Thus, it appears that the challenge of adapting a general purpose global optimization method to solve these problems as successfully as the methods that take advantage of the known geometries solution is a serious one. This is the challenge addressed in the remainder of this paper.

4 The New Large-Scale Stochastic Global Optimization Method

Our new method incorporates the basic features, sampling and local minimizations, of existing stochastic methods, but it is only able to solve large-scale problems because it also incorporates new phases that focus on small dimensional subproblems. The new phases constitute a significant departure from previous stochastic methods and account for a significant portion of the computational effort of the new method. In Section

4.1 we review existing stochastic methods and indicate why they proved incapable of finding solutions, or even near-solutions, to Lennard-Jones problems after large amounts of computation. Section 4.2 describes our new method in the context of Lennard-Jones problems and also explains how it can be applied to other large scale, partially separable problems.

4.1 Existing Stochastic Methods and their Limitations

The global optimization approach described in this paper is related to the stochastic methods of [24] and the parallel variants of [6] and [30, 29], so it is useful to briefly review those here. The basic structure of these stochastic methods is an iteration that consists of three phases. First random sample points are generated in the problem domain, then these are compared with each other and a subset of them are selected as start points for local minimizations, and finally local minimizations are performed from these start points. At the end of this three step iteration, a decision is made whether to terminate the algorithm or perform another iteration, based on the stopping criteria. As the algorithm proceeds, sample points in the current iteration are added to the sample points from all previous iterations, so that the start point selection phase has more information to base its selection upon. When the algorithm terminates, the global minimizer is considered to be the local minimizer with the lowest function value. Rinnooy Kan and Timmer have shown that if start points are chosen according to rules mentioned below, and the local minimization process has certain reasonable properties, then with probability one this type of procedure will find the global minimizer after doing a finite (though possibly very large) amount of work [24].

Several lower level features of these methods are important to our new method as well. In the sampling phase, a "cutoff level" may be employed such that any sample points generated whose function value falls above the cutoff level are discarded and not considered as possible start points for local minimizations. This does not affect the theoretical properties of the algorithm, but significantly improves the efficiency of the start point selection phase in practice. In the start point selection phase, a sample point is selected as a start point for a local minimization if it has the lowest function value of all other sample points or previously found local minimizers within the "critical distance" from it. The critical distance is a function of the number of sample points generated each iteration, the iteration number, and the Lebesque measure of the feasible region, and is crucially related to the probabilistic properties of the method. Finally, in our implementations of global optimization methods, the local minimizations are performed using the BFGS method of the UNCMIN package [26].

It is useful to include a brief outline of this procedure for the subsequent discussion and for comparison with our new method. This is presented in Algorithm 4.1 below.

Over the last several years, we have developed parallel variants of stochastic global optimization methods that have been implemented on networks of workstations used as an MIMD computer, as well as an Intel iPSC/2 hypercube ([6]). In order to parallelize the work performed by these methods, the feasible domain is divided up among the processors and sampling is done separately in each subdomain for phase 1. Phase 2 is also done within separate subdomains on the individual processors, but potential start

Algorithm 4.1 – Framework of a Standard Stochastic Global Optimization Algorithm

Given $f : R^n \rightarrow R$, feasible region D

For each iteration:

1. **Sampling :** Generate the coordinates of the new random sample points in the feasible region, and evaluate $f(x)$ at each new sample point. Discard all sample points whose function value is below the global "cutoff level".

2. **Start Point Selection :** Select a subset of the sample points to be start points for local minimizations. (A sample point is selected to be a start point if it has the lowest function value of all sample points within the "critical distance" from it.)

3. **Local Minimizations :** Perform a local minimization from each start point selected in step 2.

4. **Decide whether to stop :** If the stopping criteria has been met, regard the lowest local minimizer found as the global minimizer, otherwise repeat steps 1-4.

points that are close to the border of a subdomain must be handled differently to insure that neighboring subregions will not end up selecting start points which would lead to duplicate local searches. Start points for local searches are then collected centrally and assigned to processors as they become available. Similar techniques will be utilized in certain portions of the parallel implementation of our new method.

Recently, we have also developed sequential and parallel versions of stochastic methods that utilize adaptive decisions to target sampling and local minimizations to regions of the parameter space that appear to be the most productive in terms of producing points with relatively low function values ([30, 29]). The parallel versions of these methods also incorporate asynchronous features and new scheduling algorithms. The parallel adaptive global optimization algorithm will be utilized as a subroutine in the small subproblem global optimization phase of our new method.

The sequential and parallel methods based upon Algorithm 4.1 have proven to be effective on standard test problems where the number of parameters are very small, usually two to six. In attempting to use these methods to solve Lennard-Jones problems with more than about 25 parameters (and hundreds or thousands of local minimizers), however, it became obvious that they would not be at all successful and that additional techniques would be necessary. The first reason was that we found that even if very large numbers of random sample points were generated, these sample points did not lead to good start points for local minimizations. That is, virtually all the sample points had relatively high function values, and the local minimizations from the start points selected from these sample points were very expensive and not likely to lead to the lowest local

minimizers. This problem is due in part to the inherent difficulty of sampling in a large dimensional variable space, which makes it likely that any sample point will be far from a local minimizer and a local minimization from it will be expensive. It is also due to the fact that for most molecular configuration problems and many others, the function values at most random sample points are considerably higher than the function value at low minimizers, and are not indicative of whether the sample points are close to the lowest minimizers.

Secondly, even after the problem of finding good sample and start points was addressed by incorporating the first set of small dimensional subproblem techniques that we describe in Section 4.2, and some good local minimizers were found, it became clear that the orientation of stochastic methods towards finding all the low local minimizers would not lead to a computationally efficient method for locating the global minimizer for problems with so many local minimizers. Instead, it seemed desirable to be able to progress efficiently from low local minimizers to even lower local minimizers, rather than to do the work of finding better minimizers by complete local minimizations from more random sample points. Section 4.2 describes how the new method addresses both of the challenges we have just described. Note that at least some of the problem characteristics that lead to these challenges, namely the large number of variables, the very large number of local minimizers, and the variation in function values over the domain, will be present in any large scale global optimization problem.

4.2 New Features for Partially Separable Problems

The approach we have taken to generate a large-scale global optimization algorithm for Lennard-Jones problems that will be applicable to many other large problems as well is to use the "partially separable" property of the Lennard-Jones potential energy function (2.1) at several key places in the algorithm. A partially separable function is a function that is the sum of many terms, each of which involves only a small subset of parameters. Partially separable problems have been studied extensively in local optimization (see e.g. [12]), and it is felt that many large scale optimization problems have this form. For this type of function, it is reasonable to expect that one could lower the function value effectively by working globally with a small subset of variables at once, particularly if one has first identified this subset of the variables as one that is contributing significantly to the current function value. The techniques we now describe use this small subproblem approach. We will present them in terms of the Lennard-Jones problem, but they can be applied to a wide range of problems, especially to any large scale partially separable function, although as we discuss at the end of this section, the identification of the appropriate small subsets of the variables is application dependent.

Our new algorithm continues to incorporate some random sampling and local minimizations in the full dimensional variable space as parts of the method. It also contains two new phases that work on small dimensional subproblems. The first of these is used to efficiently generate improved sample points from an initial set of random sample points, before selecting the start points for full dimensional local minimizations. The second new phase, which constitutes a major portion of the computational effort of the new method and represents a significantly different approach from existing stochastic methods, effi-

ciently generates improved local minimizers from an initial set of local minimizers.

The algorithm starts by generating a moderate sized random sample of the function over the full domain. Then the first new phase, the generation of improved sample points through one-atom sampling, is applied. First, all but a small fraction of the sample points with the lowest function values are discarded. Then each remaining sample point whose function value is above some threshold value is then perturbed in the following manner. The contribution of each atom in the sample point to the total function value of that point is calculated, and the atom that contributes the most to the overall function value is determined. This atom is then randomly resampled some number of times, and for each new sample value of this atom, the function value of this molecular configuration with the remaining atoms unchanged is calculated. The coordinates for the resampled atom that result in the lowest function value are retained as the new values of these variables within this sample point. If the function value of the modified sample point is still above the threshold value, the process is repeated, i.e. a (new) single atom is selected, resampled, and moved, until the threshold has been reached. The threshold value may be varied slightly for each sample point, to ensure variation within the sample. From these perturbed sample points, start points for local minimizations are selected as in phase 2 of Algorithm 4.1.

In our experience, the modified sampling procedure not only leads to a quick, large reduction in function values of the sample points (for example for 55 atoms, from function values of 43 to 909 to function values of -20 to 0 in two to eight one-atom moves), but the perturbed sample points also lead to much better start points for local minimizations than the original sample points would have. That is, the local minimizations from the start points produced from the perturbed sample require fewer iterations, and generally result in minimizers with lower function values, than minimizations from start points from the original sample. These lower minimizers in turn generally decrease the cost and improve the results of subsequent steps of the algorithm. In addition, the cost of the entire sample modification process is very low and is easily outweighed by the savings. For the Lennard-Jones function (2.1). the cost of sampling on one atom is $\frac{1}{N}$ times the cost of sampling in the full space, and the cost of re-evaluating the overall energy of the configuration when one atom is perturbed is $\frac{2}{N}$ times the cost of evaluating (2.1) in general. (The selection of the atom that contributes most to the energy just requires a simple bookkeeping operation, and its cost is insignificant because it is performed so infrequently.) The benefit of this phase, and the second small subproblem phase, to the efficiency and success of the algorithm appear to increase significantly as the number of variables increases.

The second small dimensional subproblem phase that we have added improves the function values of already low local minimizers through one-atom global optimizations. A set of local minimizers is initially produced by full dimensional local minimizations from the start points selected from the perturbed sample points. The k minimizers with the lowest function values are then passed on to this new phase of the algorithm, which proceeds as follows. The best local minimizer that has not yet been used in this phase is chosen, and the energy contribution of each atom in the configuration is computed. The atom that contributes the most to the total function value is chosen for improvement, as was done in the sample point improvement phase. (The atom that

is second worst may sometimes be selected instead.) The key difference between this phase and the sample point improvement phase is that then, rather than just *sampling* on the selected atom, a *complete global optimization algorithm* is applied to find the best new positions for the selected atom within the existing configuration of the remainder of the molecule. Generally these new positions for this atom lead to configurations that are close to, but not exactly, local minimizers for the entire molecule. Therefore a full dimensional local minimization algorithm is then applied to "polish" the few best of these new configurations. (We chose the few best rather than just the best because we have found that sometimes the best polished solution does not come from the best unpolished solution.) These new full-dimensional local minimizers are then merged with those found previously, and the process just described – select lowest minimizer, select atom from it, one-atom global optimization, full-dimensional local minimization polishes – is repeated a fixed number of times.

This procedure has been very effective in locating improved local minimizers, in cases where simply doing more sampling and local minimizations in the full dimensional variable space was not productive. The cost of this phase is also low in relation to its benefit. Even though a global optimization routine is used as a subproblem, it is not very expensive because there are only three variables in the subproblem, and because each function (and gradient) evaluation costs only $\frac{2}{N}$ of a full function (or gradient) evaluation as explained above. Furthermore, we utilize the adaptive global optimization routine of [30] to make this the solution of this global optimization subproblem even more efficient. The subsequent local minimizations in the full parameter space generally require very few iterations and so they, too, are inexpensive.

The framework for the new global optimization algorithm for solving Lennard-Jones problems is outlined in Algorithm 4.2 below. We note that the reason that we say that the local minimizer improvement phase of Step 5 represents a departure in philosophy from standard stochastic methods is that it is deterministic in nature, at least in its high level formulation. In fact it solves the small subproblem by a global optimization routine that in our implementation is stochastic, but for such small problems the results of this routine may realistically be hoped to be deterministic, i.e. all the lowest minimizers should be found. To retain the probabilistic guarantees of success of standard stochastic methods, we would have to iterate on the entire Step 1-5 sequence, but we do not do this in our current implementation because the number of local minimizers is so large that satisfying the conditions for such guarantees would be impractical.

Algorithm 4.2 has been described as a sequential method, although in practice we have run it in parallel on a small network of Sun workstations for added efficiency. The details of the parallelization are only marginally relevant to this paper. Briefly, we have parallelized steps 1-4 in a fairly routine manner, subdividing the domain to parallelize steps 1-3 and collecting and allocating the local minimizations to parallelize step 4. In addition, we have used the parallel adaptive global optimization routine of [30] to parallelize the small dimensional global optimization in step 5, but for now we have run the remainder of step 5 sequentially.

Finally, we note that the approach to global optimization that we have described in this section will generalize to any partially separable function (although we do not claim that it will necessarily be effective in all cases). The determination of which subset of

Algorithm 4.2 – Framework of the Large-Scale Global Optimization Algorithm for Lennard-Jones Problems

Given $f : R^n \rightarrow R$, feasible region D

1. **Sampling in Full Domain :** Randomly generate the coordinates of the sample points in the region D, and evaluate $f(x)$ at each new sample point. Discard all sample points whose function value is below a global "cutoff level".

2. **One-atom Sampling Improvement :** For each remaining sample point : While the energy of the sample point is above the threshold value, Repeat:

   - Select the atom that contributes most to the energy function value
   - Randomly sample on the location of the selected atom
   - Replace this atom in the sample point with the new sample coordinates that give the lowest energy value.

3. **Start Point Selection :** Select a subset of the improved sample points from step 2 to be start points for local minimizations.

4. **Full-Dimensional Local Minimizations :** Perform a local minimization from each start point selected in step 3. Collect some number of the best of these minimizers for improvement in Step 5.

5. **Improvement of Local Minimizers :** For some number of iterations:

   - From the list of full-dimensional local minimizers, select the local minimizer with the lowest energy value that has not yet been modified in this phase.
   - Select the atom that contributes the most (or second most) to the energy function value, and was not moved in the previous iteration.
   - Apply a global optimization algorithm to this configuration with only this atom as a variable.
   - Apply a local minimization procedure, using all the atoms as variables, to the lowest configuration(s) that resulted from the one-atom global optimization.
   - Merge the new lowest configurations into the existing list of local minimizers.

variables should be selected for the small dimensional subproblems will be application dependent, however. In particular, note that for the Lennard-Jones problem, we do not select one variable, but an entire atom consisting of three variables, for the small dimensional subproblems. This is because an atom is the natural unit of granularity for this problem, since each term in the energy function that involves one coordinate of a given atom involves the other two coordinates as well. While it might be possible for a code to automatically determine the correct subproblem granularity for general partially separable functions, we do not pursue this issue for now, assuming instead that the code would be modified to incorporate the correct small subproblem granularity structure for the particular class of applications. We briefly mention another instance of this choice of small subproblem granularity in Section 6.

5 Test Results

We have applied Algorithm 4.2 to the Lennard-Jones molecular configuration problem (2.1), with the number of atoms ranging from 5 to 55 (i.e. 15 to 165 variables). The following algorithmic parameters were used in the algorithm. 4000 sample points were generated in Step 1, and the best 40 of those were perturbed in Step 2. The threshold levels used in Step 2 varied for different problems sizes, and within each problem size, to ensure variability of start points, ranging from -1.0 to -20.0. Each atom selected in Step 2 was randomly sampled 100 times. At most 40 full dimensional minimizations were performed in Step 4, and the best 6 local minimizers were passed on to Step 5. Step 5 was performed for 20 iterations, where each iteration involves one complete one-atom global optimization. At the end of the one-atom global optimization, the four lowest configurations that resulted from the global optimization were polished using full dimensional local minimization. The algorithm was run on a network of five Sun 4 workstations, and was parallelized as summarized at the end of Section 4.1.

Tables 5.1a and 5.1b below give the results of these experiments. The third column gives the lowest energy value found by our algorithm, and the last column gives the lowest energy value for the problem that has been found by any method. All these lowest values have been found by the special purpose algorithms of [23] and [20], and sometimes, for smaller problems, by some other methods as well [15]. In order to assess the importance of step 5 (the improvement phase for local minimizers), the second column shows the best energy value found by Algorithm 4.2 through step 4.

These results show that Algorithm 4.2 finds the known lowest energy value for all problems with 5 through 30 atoms, but for the larger problems (31 - 55), it only finds the presumptive minimum energy value in three cases (N = 33, 35, and 38). The results also show that for every problem with more than 16 atoms, step 5 of the algorithm produces at least some improvement, and without this step, the known solutions for problems with greater than 16 atoms would not have been found.

To our knowledge, these are the best results obtained for these Lennard-Jones problems by any global optimization method that does not make use of the solution structure, and they may be among the first reported successes by any general purpose global optimization method for problems of up to 100 variables. The results compare favorably with those obtained for the Lennard-Jones problems by [15], where the solution structure

Table 5.1a: Experimental Results for LJ Problems

# of Atoms	Min. after Step 4	Min. from Alg. 4.2	Best Known Min.
5	-9.103852	-9.103852	-9.104
6	-12.712062	-12.712062	-12.712
7	-16.505384	-16.505384	-16.505
8	-19.821489	-19.821489	-19.822
9	-24.113360	-24.113360	-24.113
10	-27.479739	-28.422532	-28.420
11	-32.765970	-32.765970	-32.765
12	-37.967600	-37.967600	-37.967
13	-44.326801	-44.326801	-44.327
14	-47.845157	-47.845157	-47.845
15	-52.322627	-52.322627	-52.323
16	-56.815742	-56.815742	-56.816
17	-61.094499	-61.317995	-61.318
18	-66.284568	-66.530949	-66.531
19	-70.818708	-72.659782	-72.660
20	-75.094179	-77.177043	-77.177
21	-79.890928	-81.684571	-81.685
22	-85.214574	-86.809782	-86.810
23	-89.847243	-92.844472	-92.844
24	-95.184349	-97.348815	-97.349
25	-101.010521	-102.372663	-102.373
26	-105.432245	-108.315616	-108.316
27	-111.757532	-112.873584	-112.874
28	-116.216452	-117.822402	-117.822
29	-122.534706	-123.587371	-123.587
30	-126.201751	-128.286571	-128.287
31	-132.618957	-133.293822	-133.586
32	-135.782887	-138.823611	-139.636
33	-143.389903	-144.842719	-144.843
34	-148.366770	-149.010289	-150.045
35	-153.484883	-155.756643	-155.757
36	-157.444623	-159.881811	-161.825
37	-162.516048	-164.744147	-167.034
38	-170.659207	-173.134317	-173.134
39	-175.110320	-176.421266	-180.033
40	-179.464074	-180.962189	-185.250
41	-186.103672	-187.664377	-190.536
42	-192.282198	-194.327597	-196.278
43	-197.725819	-198.737089	-202.365
44	-202.145505	-204.561112	-207.689

Table 5.1b: Experimental Results for LJ Problems

# of Atoms	Min. after Step 4	Min. from Alg. 4.2	Best Known Min.
45	-209.146837	-212.132580	-213.785
46	-212.999826	-216.691708	-220.680
47	-219.078345	-221.308952	-226.012
48	-224.513300	-228.674021	-232.200
49	-231.696610	-234.482267	-239.092
50	-237.010457	-238.244280	-244.550
51	-241.988986	-245.327151	-251.254
52	-246.584642	-253.403586	-258.230
53	-252.619036	-254.291490	-265.203
54	-257.621963	-259.917199	-272.209
55	-263.438099	-268.117183	-279.248

is heavily utilized, and by [27], where the solution structure is not utilized. The results for over 30 atoms are not as good, however, as those in [23] and [20], except for three problems. As mentioned in Section 2, the methods in [23] and [20] make extensive use of the known lattice structure of the solution to Lennard-Jones problems, searching over lattices with the appropriate structure and then using the best of these as start points for local minimizations. On the one hand, it can be argued that these methods are not too relevant in assessing the potential of our algorithm for more general classes of problems, since the special purpose knowledge of the solution structure that they utilize heavily may not have an analog in many other more practical cases, including many practical molecular configuration problems. On the other hand, the results of these methods on the Lennard-Jones problem demonstrate that we do not yet have a general purpose algorithm that finds the global minimizer of large, difficult problems with any confidence, although we may have some confidence that the method finds a reasonably low local minimizer and represents a significant step towards the development of even more reliable methods.

Finally, in Table 5.2 we give an indication of the cost of our method on these problems. We have not attempted to optimize our implementation with respect to efficiency, and it is likely that both the numbers of function evaluations and the execution times could be improved considerably. But this table should give at least a rough indication of the computational effort. Regarding the function evaluations, recall that the one-atom function evaluations cost $\frac{2}{N}$ times as much as an N-atom function evaluation. Thus the cost of the one-atom function evaluations ranges from 10% to 28% of the cost of the N-atom function evaluations for the problems with 20 or more atoms. The execution times can only be interpreted very roughly, because while they were obtained using five Sun 4 workstations, these machines were of four different models with differing speeds, two were file servers that were also servicing other applications, and one key portion of the algorithm, the local minimization in step 5, was not parallelized although it could be. In addition, step 5 is partially sequential while steps 1-4 are entirely parallel, so the times

Table 5.2: Costs for Lennard-Jones Problems using Algorithm 4.2

Number of Atoms	N-atom function evals	1-atom function evals	Time in seconds for Steps 1-4	Time in seconds for Step 5	Time in seconds for Total Alg
5	7307	75306	21.0	129.8	150.8
10	12743	55852	90.4	136.8	227.2
15	14483	61859	251.8	173.4	425.2
20	18211	50780	479.0	227.1	706.1
25	21394	52250	658.1	399.8	1057.9
30	27429	71164	1410.4	720.6	2131.0
35	27381	80197	1783.3	690.9	2474.2
40	31932	80880	2151.2	1479.8	3631.1
45	33303	79980	2885.1	1130.3	4015.4
50	48104	137147	4679.3	3218.5	7897.8
55	52213	143441	5461.2	6838.8	12300

for step 5 are artificially large relative to the times for steps 1-4. Still, the execution time statistics show that the new local minimizer improvement phase of the algorithm (Step 5), which was seen previously to play a major role in its effectiveness, also constitutes a significant portion of its computational effort.

6 Summary and Future Research

The global optimization method discussed in this paper combines several new techniques with some techniques from standard stochastic global optimization methods to enable it to solve large scale problems effectively. The new techniques involve concentrating on small dimensional subproblems in two new phases of the method, improvement of sample points and improvement of local minimizers. For the Lennard-Jones problems, the small dimensional subproblems involve one atom, generally the one that contributes the most to the current energy function value. The sample point improvement phase perturbs selected sample points one atom at a time to generate points with much lower function values that will be better start points for local minimizations. The local minimizer improvement phase perturbs already low minimizers by selecting one atom and performing a global optimization on it with the remainder of the configuration fixed, followed by full-dimensional local minimization polishing steps of the best new configurations. This phase, which departs significantly from the standard approach of stochastic methods by taking a much more deterministic approach, constitutes a significant portion of the computational effort of the new algorithm and often substantially improves the results of the algorithm. The new method is designed to be applicable to any large, partially separable problem, although the selection of the appropriate unit of granularity for the small dimensional subproblems is application dependent.

One goal of this research is eventually to solve practical, complex molecular configuration problems involving hundreds or thousands of atoms, for example the configuration of proteins or polymers. As a step in this direction, we have recently begun experimenting with applying the techniques of this paper to clusters of water molecules. For this problem, the appropriate small dimensional subproblem is a water molecule (9 variables) as opposed to an atom (3 variables) for the Lennard-Jones problems. We are obtaining promising results on this problem, but are learning that further techniques for improving local minimizers are useful for more complex problems such as this one. We will report on this work in a subsequent paper. How effective the approach of this paper is for other classes of partially separable global optimization problems remains to be seen.

Another area of future research is making the approach of this paper more efficient. Important opportunities include utilizing local minimization procedures that are especially oriented to the particular class of large scale problems in the application, and adapting the algorithm to effectively utilize massively parallel computers. We have not pursued these directions for the Lennard-Jones problems as these problems are only stepping stones to more realistic large scale problems. They are certain to become important issues to address in the solution of larger and more practical problems, however.

A final comment is that it is clear from the results of Section 5 that the methods of this paper currently find low local minimizers, but not necessarily the global minimizer, in a reasonable amount of time on the computers we have used. In fact we have relinquished the probabilistic guarantee of finding the global minimizer since we do not iterate on the overall five step process, but doing so would not be expected to improve the practical performance. Indeed, our computational experience suggests that it may be practically intractable to find the global minimizer for many difficult large scale problems. However, there are many approaches that might enhance the effectiveness of the type of method discussed here. Efficient techniques that could find accurate lower bounds on the global minimum function value such as those used on small problems in [21] would be extremely useful. Also smoothing techniques like those described in [27] could improve performance. Certainly it is probable that if we were to implement this kind of procedure on a powerful supercomputer, we could find better solutions in a reasonable amount of time; thus the issue of parallel implementations is likely to be important.

References

[1] Aluffi-Pentini, F., Parisi, V., and Zirilli, F. (1988). A global optimization algorithm using stochastic differential equations. *ACM Transactions on Mathematical Software*, **14**, 345-365.

[2] Barrón, C. and Gómez, S. (1991) The Exponential Tunneling Method.

[3] Booker, L.B., Goldberg, D.E., and Holland, J.H. (1989). Classifier systems and genetic algorithms. *Artificial Intelligence*, **40**, 235-282.

[4] Branin, F.H. (1972). Widely convergent methods for finding multiple solutions of simultaneous nonlinear equations. *IBM Journal of Research Developments*, 504-522.

[5] Brooks, C.L., III, Karplus, M., and Petitt, B.M. (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics.* J. Wiley and Sons, New York.

[6] Byrd, R.H., Dert, C.L., Rinnooy Kan, A.H.G., and Schnabel, R.B. (1990). Concurrent stochastic methods for global optimization. *Mathematical Programming*, **46**, 1-29.

[7] Dixon, L.C.W. and Szego, G.P., eds. (1978). *Towards Global Optimization*, **2**. North-Holland, Amsterdam.

[8] Donnelly, R.A. (1990). Discrete generalized descent for global optimization. Preprint, Department of Chemistry, Auburn University.

[9] Farges, J., DeFeraudy, M.F., Raoult, B., and Torchet, G. (1985). Cluster models made of double icosahedron units. *Surface Sci.*, **156**, 370-378.

[10] Freeman, D.L. and Doll, J.D. (1985). Quantum Monte Carlo study of the thermodynamic properties of argon clusters: the homogeneous nucleation of argon in argon vapor and "magic number" distributions in argon vapor. *J. Chem. Phys.*, **82**, 462-471.

[11] Goldstein, A.A. and Price, J.F. (1971). On descent from local minima. *Mathematics of Computation*, **25**, 569-574.

[12] Griewank, A.O. and Toint, Ph.L. (1982). Partitioned variable metric updates for large sparse optimization problems. *Numerische Mathematik*, **39**, 119-137.

[13] Hansen, E.R. (1980). Global optimization using interval analysis - the multidimensional case. *Numerical Math*, **34**, 247-270.

[14] Hoare, M.R. (1979). Structure and dynamics of simple microclusters. *Adv. Chem. Phys.*, **40**, 49-135.

[15] Hoare, M.R. and Pal, P. (1971). Physical cluster mechanics: statics and energy surfaces for monatomic systems. *Adv. Phys.*, **20**, 161-196.

[16] Holland, J.H. (1975). *Adaptation in natural and artificial systems.* The University of Michigan Press, Ann Arbor.

[17] Judson, R.S.,Colvin, M.E.i,Meza, J.C., Huffer, A. and Gutierrez, D. (1991) Do Intelligent Configuration Search Techniques Outperform Random Search for Large Molecules? SAND91-8740.

[18] Kirkpatrick, S., Gelatt, C.D. Jr., Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, **220**, 671-680.

[19] Levy, A.V. and Montalvo, A. (1985). The tunneling algorithms for the global minimizer of functions. *SIAM Journal on Scientific and Statistical Computing*, **6**, 15-29.

[20] Maier, R.S., Rosen, J.B., Xue, G.L. (1992). A Discrete-Continuous Algorithm for Molecular Energy Minimization. AHPCRC Preprint 92-031.

[21] Maranas, C. D. and Floudas C. A. (1992). "A Global Optimization Approach for Lennard-Jones Microclusters", Preprint, Department of Chemical Engineering, Princeton University, June 1992.

[22] McCammon, J.A. and Harvey, S.A. (1987). *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, England.

[23] Northby, J.A. (1987). Structure and binding of Lennard-Jones clusters:$13 \leq N \leq 147$ *J. Chem. Phys.*, **87**, 6166-6177.

[24] Rinnooy Kan, A.H.G. and Timmer, G.T. (1984). A stochastic approach to global optimization. In *Numerical Optimization*, P. Boggs, R. Byrd, and R.B. Schnabel, eds., SIAM, Philadelphia, 245-262.

[25] Rinnooy Kan, A.H.G. and Timmer, G.T. (1989). Global optimization. In *Handbooks in operations research and management science, volume I : optimization*, G.L. Nemhauser, A.H.J. Rinnooy Kan, and M.J. Todd, eds., North-Holland, 631-662.

[26] Schnabel, R.B., Koontz, J.E., and Weiss, B.E. (1985). A modular system of algorithms of unconstrained minimization. *ACM Transactions on Mathematical Software*, **11**, 419-440.

[27] Shalloway, D. (1991). Packet annealing : a deterministic method for global minimization, with application to molecular conformation. Preprint, Section of Biochemistry, Molecular and Cell Biology, Cornell University. To appear in *Global Optimization*, C. Floudas and P. Pardalos, eds., Princeton University Press.

[28] Shubert, B.O. (1972). A sequential method seeking the global maximum of function. *SIAM Journal on Numerical Analysis*, **9**, 379-388.

[29] Smith, Sharon L. (1991). Adaptive asynchronous parallel algorithms in distributed computation. Ph.D. Thesis, Department of Computer Science, University of Colorado, Boulder, Colorado.

[30] Smith, S.L., Eskow, E., and Schnabel, R.B. (1989). Adaptive, asynchronous stochastic global optimization algorithms for sequential and parallel computation. In *Proceedings of the Workshop on Large-Scale Numerical Optimization*, T.F. Coleman and Y. Li, eds. SIAM, Philadelphia, 207-227.

[31] Smith, Sharon L. and Schnabel, Robert B. (1991). Centralized and distributed dynamic scheduling for adaptive, parallel algorithms. Technical Report CU-CS-516-91, Department of Computer Science, University of Colorado, Boulder, Colorado.

[32] Wille, L.T. (1975). Minimum-energy configurations of atomic clusters: new results obtained by simulated annealing. *Chem. Phys. Lett.*, **133**, 405-410.