Pivot method for global optimization

Pablo Serra,* Aaron F. Stanton, and Sabre Kais
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
(Received 26 September 1996)

A pivot algorithm for the location of a global minimum of a multiple-minimum problem is presented. The pivot method uses a series of randomly placed probes in phase space, moving the worst probes to be near better probes iteratively until the system converges. The approach chooses nearest-neighbor pivot probes to search the entire phase space by using a nonlocal distribution for the placement of the relocated probes. To test the algorithm, a standard suite of functions is given, as well as the energies and geometric structures of Lennard-Jones clusters, demonstrating the extreme efficiency of the method. Significant improvement over previous methods for high-dimensional systems is shown. [S1063-651X(97)08801-6]

PACS number(s): 02.70.−c, 02.60.Pn

Rational drug design, molecular modeling, quantum-mechanical calculations, and mathematical biological calculations are but a few examples of fields that rely heavily upon the location of a global minimum in a multiple-minimum problem [1–8]. An algorithm that is versatile enough to be utilized in any given problem, easy enough that gradients of the phase space need not be calculated, and robust enough to avoid entrapment in local minima would find many applications. In this paper we present such a method along with a set of test functions as well as solutions to structure of Lennard-Jones systems of up to 20 particles as a demonstration of its versatility.

The task is to locate a global minimum of a given function $f(\vec{x})$ within a predetermined phase space Ω defined by the maximum and minimum values of all parameters $x(i)$ of the function. A globally minimum value for this function is assumed to exist within the defined phase space at the corresponding phase space point

$$f_0 = f(\vec{x}_0) = \min_{\vec{x} \in \Omega} f(\vec{x}), \tag{1}$$

where x_0 is a phase-space point with the global minimum value f_0 for the function $f(\vec{x})$.

The pivot method begins with a given number of probes placed initially within the phase space Ω. If nothing is known of the problem, the probes are placed completely at random. However, if something is known of the problem, it is most convenient to use this information in the initial determination of probe location, which will be demonstrated in the Lennard-Jones example to follow. Each probe is simply a set of values for the parameters of the problem $x(i)$ within the boundaries of the phase space and therefore with a given functional value equal to the value of the function $f(\vec{x})$ at the probe point. In the current method, we start with $N=2m$ initial probes of which m probes will act as the pivot probes, and the remaining m probes will be relocated.

A local selection of the m pivot probes begins with a search at each probe for its nearest neighbor, based on the distance of the probes. Once we have paired the probes, the probe with the lower value for the function $f(\vec{x})$ is defined as the pivot probe, the other probe being the probe that will be relocated. There are several methods possible to do this pairing, but in our particular case we began with finding the nearest neighbor for probe 1 and removed those two probes from further consideration. This was repeated until all points had been paired.

For each pivot probe with parameter values $\vec{x}_{B,i}$, we explore phase space by placing the probe to be relocated near the pivot probe by changing its parameters $\vec{x}_{R,i}$ as

$$\vec{x}_{R,i} = \vec{x}_{B,i} + \Delta\vec{x}_i, \tag{2}$$

where $\Delta\vec{x}_i$ is a randomly generated vector according to an exploring distribution $g_e(\Delta\vec{x})$. We have tested several distributions, such as the standard Gaussian distribution used in the pivot method [9] and in simulated annealing [10], as well as the Cauchy-Lorentz distribution proposed by Zsu and Hartley for fast simulated annealing method [11]. Finally, we have tested a generalized q distribution based on the Tsallis entropy [12] and recently used with good results in the generalized simulated annealing method [13–16]. We chose to use the general q distribution for the placement of the probes near the pivot probes. This distribution is defined to be [14]

$$g_q(x) = \sqrt{\frac{q-1}{\pi}}\,\frac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\frac{1}{q-1}-\frac{1}{2}\right)} \times \frac{[\beta(t)]^{1/(3-q)}}{(1+(q-1)\{[\beta(t)]^{1/(3-q)}x\}^2)^{1/(q-1)}}, \tag{3}$$

where β is defined to be $1/T$ and T is an artificial temperature given by [14]

$$T(t) = \frac{2^{q-1}-1}{(1+t)^{q-1}-1}\,T(1), \quad t = 1,2,3,\ldots, \tag{4}$$

*Permanent address: Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina.

TABLE I. Average number of function evaluations in the global optimization of five test functions. The methods are the pure random search (PRS), simulated annealing types 1 and 2 (SA1 and SA2, respectively), and tabu search (TS). The references for these methods and results can be found in Ref. [17].

Method	GP	Branin	H3	H6	Shubert
PRS	5125	4850	5280	18 090	6700
SA1	5439	2700	3416	3975	241 215
SA2	563	505	1459	4648	780
TS	486	492	508	2845	727
Present	153	68	52	237	159

where $T(1)$ is the initial temperature and t is the discrete time corresponding to the computer iterations [17].

The most relevant fact about this distribution is the introduction of a new parameter q. Special cases that should be noted are the limit $q \to 1$, where the general q distribution approaches the Gaussian distribution, and $q=2$, where the q distribution is equal to the Cauchy-Lorentz distribution. The second moment of this distribution diverges for $q \geqslant 5/3$ and the distribution becomes unnormalizable for $q \geqslant 3$. As in generalized simulated annealing [14], we have found $q=2.5$ to be a good value for our global optimization method. A detailed comparison with other values of q and other distributions will be given elsewhere [18].

In our standard test suite of functions we include several well-known functions for comparison with established methods of optimization. These are the Goldstein-Price (GP), Branin, Hartman three- and six-dimensional variants (H3 and H6), and Shubert functions. The full details of these functions can be found elsewhere [19,20]. Note that all of the two-dimensional functions have similar results. The stopping criterion chosen was for the best probe to have a value no farther from the known global minimum than 3% or to stop if the number of iterations exceeded a certain value. This stopping criterion is the same as that in Ref. [17] and therefore allows an objective standard for comparison. The latter criterion was set sufficiently high to establish that the system was trapped in a local minimum if this criterion was triggered. A different number of probes was found to be optimal for each of the varying test functions. We have included in Table I our results to illustrate the improvements over previous methods. This represents an average of function calls over the successful runs. One thousand runs were done for each function. The improvement for two-dimensional methods is a factor 3.2–7.2, whereas the H3 and H6 functions show improvements of 9.8 and 12, respectively. Figure 1 illustrates the efficiency of using a value of $q=2.5$ graphically by plotting the number of times the Branin function was called upon in the minimization against values of q ranging from 1 to 3. Note that this represents a successful convergence rate in excess of 95%. A similar behavior was observed for the remainder of the functions in the test suite, but we have only included the one graph for reasons of space.

We apply the method to Lennard-Jones clusters of size 6–20 (phase-space dimension 12–54). Lennard-Jones clusters are excellent for testing the efficiency of global optimization algorithms. Regular Lennard-Jones clusters have well-established minima and minimum-energy structures for very large clusters [21]. However, the number of local minima apparently grows as $\exp(N^2)$ [22]. Wille and Vennik have shown that to find the global minimum in Lennard-Jones clusters is a NP hard problem [23]. Several global optimization methods have been applied to the energy function of Lennard-Jones clusters. These include simulated annealing [24], genetic algorithm [25], diffusion equation methods [26], quantum annealing [27], and others [28]. The total energy for a Lennard-Jones cluster of N particles is

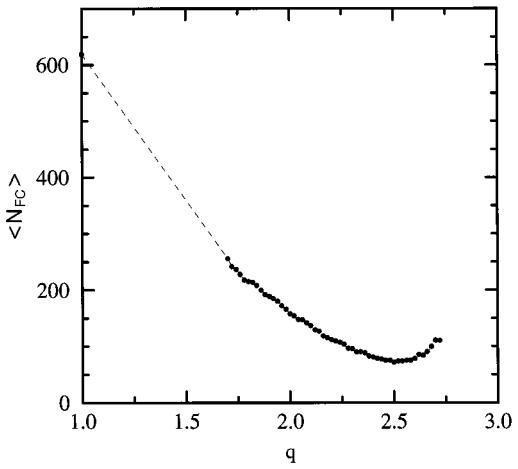

FIG. 1. Average number of function calls $\langle N_{\mathrm{FC}} \rangle$ vs q for the Branin function. This represents the average over the successful runs with a minimum convergence of 95%. One hundred runs were done for each value of q plotted.

$$E_N = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V_{\mathrm{LJ}}(r_{ij}), \tag{5}$$

where r_{ij} is the distance between the ith and the jth particles and $V_{\mathrm{LJ}}(r)$ is the Lennard-Jones two-body potential

$$V_{\mathrm{LJ}}(r) = \frac{1}{r^{12}} - \frac{2}{r^6}. \tag{6}$$

For small numbers of Lennard-Jones clusters ($N \leqslant 6$), the global energy minimum was located very quickly (less than 1 CPU second on an IBM RS/6000). For larger clusters, we incorporated the partial knowledge that we had by starting with the structure of the smaller $N-k$ clusters and adding k additional particle at random. In any ''growing'' problem, such as minimum-energy configuration of clusters, self-avoiding walks, and protein folding [21], this systematic approach to solving the structure of large clusters can be incorporated. One of the powerful features of this algorithm is that information such as this can be built into the initialization of the probes.

For an N cluster, we begin with $m \times k$ initial pivot probes chosen as follows: m, $N-1$ clusters + one random atom, m, $N-2$ clusters + 2 random atoms, ..., m, $N-k+1$ clusters + $k-1$ random atoms, and finally m completely random pivot probes (in our calculations we set $k=N/2$). With this set of the initial pivot probes, if the N cluster has a similar structure with a smaller cluster, the algorithm converges faster than purely random initial points. If the N cluster has a

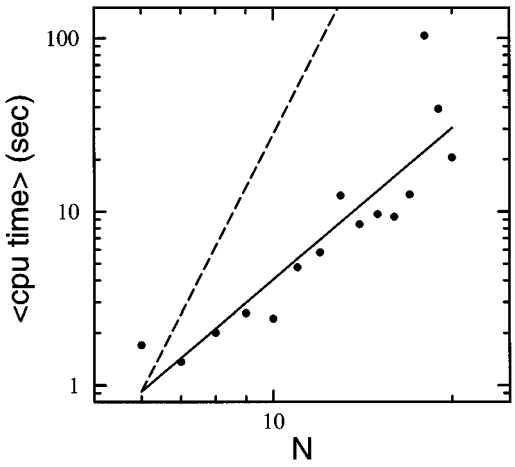

FIG. 2. log-log graph of CPU time in seconds vs number of Lennard-Jones particles. The dashed line represents the scaling of the genetic algorithm of Ref. [23], $N^{4.7}$. The points represent experimentally determined times for each of the clusters and the solid line represents the scaling of this method, $N^{2.9}$.

much different structure than the $N-k$ structure or has one or more local minima near the global minimum, as in the $N=18$ Lennard-Jones cluster, then the method works no less efficiently than it would with initial pivot probe locations chosen completely at random.

Figure 2 illustrates how our method scales with the number of Lennard-Jones particles to be minimized. Using a log-log scale we show that our method scales approximately as $N^{2.9}$, compared to the recently reported modified genetic algorithm, which scales as $N^{4.7}$ [25]. One should note that the probability of finding that the global energy minimum was strongly correlated with the size of the cluster. It was relatively difficult to obtain the global minimum for ''magic number'' clusters such as $N=6$ and $N=18$. This implies that the difficulty (CPU time) does not scale in a simple way with N but depends on the characteristics of the potential-energy hypersurface [29]. In order to reach the exact minimum for the Lennard-Jones clusters, a gradient descent minimization was used once our method met its convergence criteria. Our CPU time given includes this gradient minimization. In this case, we found that a value of $q=2.7$ worked better than $q=2.5$ for these criteria. The CPU time on this chart has been experimentally determined on an IBM RS/6000-580 for our method, the genetic algorithm [25] was included for comparison of scaling, and the exact CPU time was not determined. A comparison such as this shows that for extremely large systems it would allow considerable savings in time to use the present method.

We have presented a general method of optimization of arbitrary functions that has shown itself to be easy to implement, does not get easily trapped in local minima, and is extremely fast. Any initial knowledge of the behavior of the function in question is easy to incorporate in the initial conditions of the search, thus lending additional versatility. A 12-fold improvement over previous methods has been demonstrated for the six-dimensional Hartman function. CPU time has been shown to scale approximately as $N^{2.9}$ as compared to the modified genetic algorithm [25], which scales as $N^{4.7}$ for Lennard-Jones clusters. This method was empirically found to be extremely useful, and as of yet there does not exist a rigorous mathematical proof for convergence to the global minimum or for the rate of change of the temperature. Such a proof needs to be established, and various routes to do so are currently under investigation. We are using the cooling rate from generalized simulated annealing with no analytical proof that this is the best cooling rate for our method; however, through empirical testing it has been shown to yield very good results. Applications of this method for atomic and molecular clusters with complex potentials such as water, oxides, and halides of alkali metals will be given elsewhere [18].

The authors would like to thank Richard E. Bleil for many useful discussions and critically reading the manuscript. P.S. would like to acknowledge C. Tsallis for stimulating discussions about applications of generalized statistical mechanics to global optimization and D. Stariolo, who was kind enough to send us Ref. [14] prior to publication. This work was supported in part by a grant from the American Chemical Society's Petroleum Research Fund.

[1] *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1992), Vol. III.
[2] L. C. W. Dixon and G. P. Szego, *Towards Global Optimization* (North-Holland, New York, 1975), Vols. 1 and 2.
[3] D. R. Herschbach, J. Avery, and O. Goscinski, *Dimensional Scaling in Chemical Physics* (Kluwer Academic, Dordrecht, 1993).
[4] T. L. Blundell, B. L. Sibanda, M. J. E., Sternberg, and J. M. Thornton, Nature, **326**, 347 (1987).
[5] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, Protein Sci. **4**, 561 (1995).
[6] M. Oresic and D. Shalloway, J. Chem. Phys. **101**, 9844 (1994).
[7] L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem. **93**, 3339 (1989).
[8] R. S. Berry, Int. J. Quantum Chem. **58**, 657 (1996).
[9] A. F. Stanton, R. E. Bleil, and S. Kais, J. Comput. Chem. (to be published).
[10] S. Kirkpatrick, C. D. Gellat, and M. P. Vecchi, Science **220**, 671 (1983).
[11] H. Zsu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[12] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[13] D. A. Stariolo and C. Tsallis, in *Annual Reviews of Computational Physics II*, edited by D. Stauffer (World Scientific, Singapore, 1995), p. 343.
[14] C. Tsallis and D. A. Stariolo, Physica A (to be published).
[15] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).
[16] I. Andrichioaei and J. Straub, Phys. Rev. E **53**, R3055 (1996).
[17] The FORTRAN code for the random number generator of the q distribution is given in Ref. [14].
[18] P. Serra, R. E. Bleil, A. F. Stanton, and S. Kais (unpublished).
[19] D. Cvijovic and J. Klinowski, Science **267**, 664 (1995).
[20] A. Torn and A. Zilniskas, *Global Optimization* (Princeton University Press, Princeton, 1991).

[21] J. Gu and B. Du, in DIMACS 23, *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, edited by P. M. Pardalos, D. Shalloway, and G. Xue (American Mathematical Society, Providence, 1996), p. 65.
[22] M. R. Hoare, Adv. Chem. Phys. **40**, 49 (1979).
[23] L. T. Wille and J. Vennik, J. Phys. A **18**, L419 (1985).
[24] L. T. Wille, Chem. Phys. Lett. **133**, 405 (1987).
[25] S. K. Gregurick, M. H. Alexander, and B. Hartke, J. Chem. Phys. **104**, 2684 (1996).
[26] J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga, J. Phys. Chem. **95**, 4113 (1991).
[27] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Chem. Phys. Lett. **219**, 343 (1994).
[28] P. M. Pardalos and G. L. Xue, J. Global Opt. **4**, 117 (1994).
[29] J. Ma and J.E. Straub, J. Chem. Phys. **101**, 533 (1994).