

A New Method for Locating the Global Optimum: Application of the Cutting Angle Method to Molecular Structure Prediction

Kieran F. Lim (林百君)¹, Gleb Beliakov², and Lynn Batten²

¹ Centre for Chiral and Molecular Technologies, School of Biological and Chemical Sciences, Deakin University, Geelong, Victoria 3217, Australia
lim@deakin.edu.au

² School of Information Technology, Deakin University, Burwood 3125, Australia
{gleb, lmbatten}@deakin.edu.au

Abstract. Many problems in chemistry depend on the ability to identify the global minimum or maximum of a function. Examples include applications in chemometrics, optimization of reaction or operating conditions, and non-linear least-squares analysis. This paper presents the results of the application of a new method of deterministic global optimization, called the cutting angle method (CAM), as applied to the prediction of molecular geometries. CAM is shown to be competitive with other global optimization techniques for several benchmark molecular conformation problem. CAM is a general method that can also be applied to other computational problems involving global minima, global maxima or finding the roots of nonlinear equations.

1 Introduction

Many problems in chemistry depend on the ability to identify the global minimum or maximum of a function. Examples include important applications in chemometrics [1], flow rates or pH concentrations in the optimization of flow analysis methods [2–4], statistical process control in manufacturing, the maximization of yields in synthesis and manufacturing, non-linear least-squares analysis and the ability to predict molecular geometries. In particular, the last-mentioned application includes the areas of protein space structure elucidation, the investigation of host-guest interactions, the understanding of properties of superconductors and zeolites [5], and the identification of transition states [6]. In all of these situations, one needs to find the molecular geometry which corresponds to the global minimum of the potential energy surface (PES).

Global and deep local minima correspond to stable molecular conformations, and they dictate both physical and chemical properties of chemical substances. Hence, the search for the global minimum on the PES is one of the most important, albeit one of the most challenging, optimization problems in chemistry and mathematics. Typically,

a mix of local and global search techniques is necessary. The presence of large numbers of shallow local minima on the PES makes the problem unsuitable for traditional local descent techniques in the absence of other information, and so new local search strategies have emerged in the last few years. On the other hand, global search techniques perform poorly when only localized information is available [7, 8]. Thus, the success of a local/global search approach relies heavily on the availability of both local and global knowledge of the PES.

The cutting angle method described here is a new global search procedure. The authors implement it alongside known local search procedures. Existing global algorithms used in the determination of molecular structures can be classified into four (overlapping) categories: (1) deterministic methods, (2) stochastic methods, (3) heuristic methods and (4) surface smoothing methods [9]. Reviews of these methods have been given in [10]; applications to molecular conformation problems can be found in [7, 8, 11].

The computational cost of all known global search methods increases exponentially with the number of variables and so those methods are "NP-hard" [12]. For the purposes of practical results in the molecular conformation problem, it is therefore desirable to make the algorithms as efficient as possible so that conformations of a reasonably large size (or dimension) can still be managed.

This paper applies the cutting angle method, which is a new, efficient, deterministic method for global optimization, and demonstrates that it compares favorably with other deterministic techniques (systematic search, branch-and-bound, generalized descent) on a chosen set of chemical structures. Like other global optimization methods, it can also be applied to a range of other generic problems in theoretical chemistry (e.g., based on cluster analysis, similarity searching, etc.[7]).

2 Theory

The need for global optimization techniques in molecular modeling has been emphasized by many authors [7, 8, 11]. The number of local minima on the potential energy surface typically grows exponentially with the number of variables, and enumeration of all local minima using multistart local descent methods is not feasible. Moreover any method that relies only on the local information (such as derivatives), has very slim chances to locate the global minima.

The cutting angle method (CAM) was first proposed in 1999 by Andramonov, Bagirov, Glover and Rubinov, and subsequently improved by them [13-16]. It can be used for both differentiable and non-differentiable functions. The one-dimensional Piyavskii-Shubert algorithm [17], illustrated in Figure 1, emerges as a special case of CAM. The concept of the method is to replace the original optimization problem with a sequence of simpler (relaxed) optimization problems, so that the sequence of solutions of relaxed problems converge to the global minimum of a real-valued function $f(x^k), x^k \in R^n$. CAM is based on the construction of a sequence of *supporting functions* (e.g., hypercones), and these must be chosen so that the maximum of the supporting functions is the lower approximation of f .

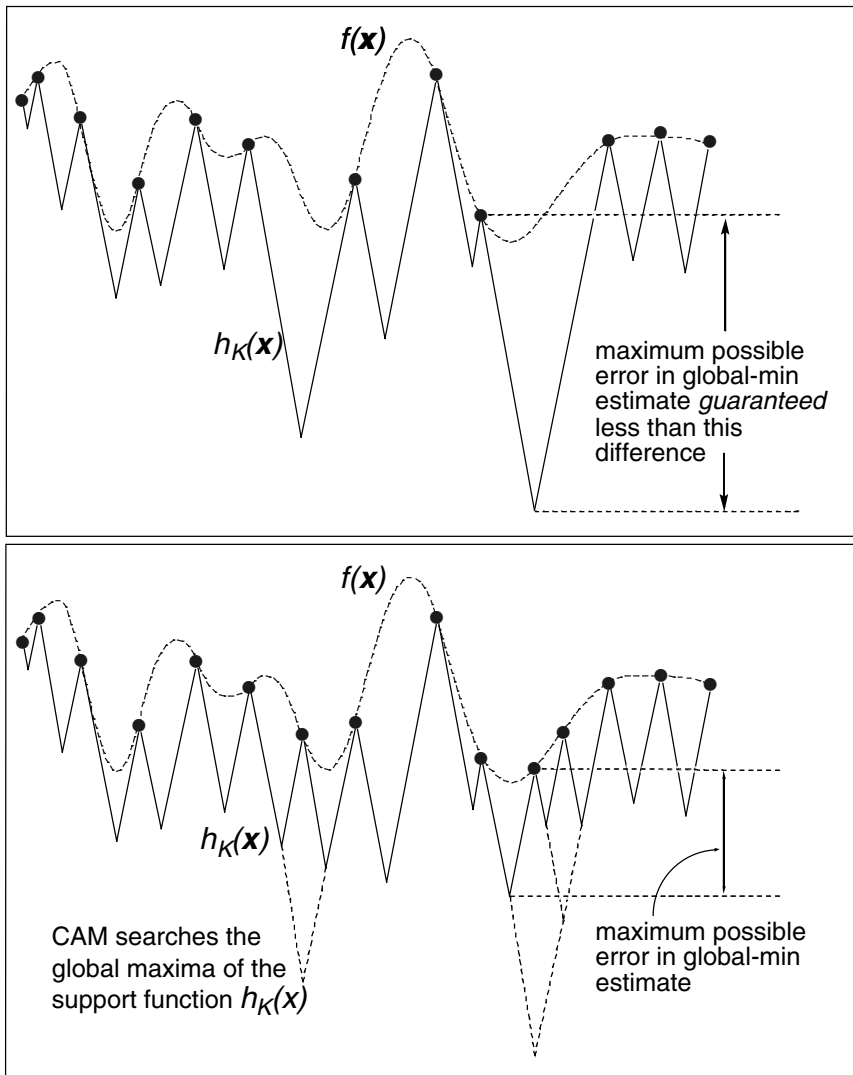


Fig. 1. Generalized cutting plane method of Lipschitz optimization (Pijavski-Shubert method). All local minima of the saw-tooth cover can be explicitly enumerated. As the number of function evaluations increases, the “teeth” become smaller and the supporting functions $h_k(x)$ converges pointwise to the required function $f(x)$. The global minimum of $h_k(x)$ converges to the global minimum of $f(x)$

For smooth functions, the *Lipschitz property* is usually interpreted as meaning that the function has a bounded first derivative. More rigorously, we define a real-valued function to be *Lipschitz continuous* if there is a real number L satisfying

$$|f(\mathbf{x}) - f(\mathbf{y})| < L \|\mathbf{x} - \mathbf{y}\| \quad (1)$$

for all points x and y for which f is defined. While we use the absolute value norm on the left hand side of (2), on the right, any norm can be used. We define the *Lipschitz constant* for f to be the least possible value for L in Eq. (1), over all x and y . The Lipschitz constant is important in our situation because it provides global information about the PES: global information is required in order to find and confirm the global minimum of the PES on any surface. In the mathematics literature, Lipschitz programming has attracted much attention as a method of (deterministic) global optimization [11, 17]. We make fundamental use of the choice of Lipschitz constant in applying CAM to the molecular structures considered in this article.

Figure 1 shows a function f that has been evaluated at several points x^1, x^2, x^3 , etc. From each evaluated point on the function f , it is possible to construct a hypercone, with slope L , and with its apex touching the function f . By using the l_∞ -norm in (2), we obtain min-type support functions. After K function evaluations at x^1, x^2, \dots, x^K , the support function

$$h_K(x) = \max_{k=1,2,\dots,K} \min_{i=1,\dots,n} \frac{x_i}{x_i^k} f(x^k), \quad x \in S_1^n \quad (2)$$

is a lower approximation to the function f [15]. Note that the form of $h_k(x)$ is very irregular (Figure 1), which is why it is called the *saw-tooth cover of f* . As more points $f(x)$ are evaluated, $h_k(x)$ converges point-wise to the function f from below. In particular, the sequence of global minima of $h_k(x)$ converges to the global minimum of the function f [15]. An important result is that, for any particular K value, the global minimum of $h_k(x)$ is *guaranteed* to be a lower bound for the true global minimum of the PES f , thus *measuring the uncertainty in the estimated global minimum* of the PES f , when the algorithm is terminated at some finite K .

Uncertainty in estimated global minimum

$$< \min\{f(x^k)\} - \min\{h_K(x)\}, \quad x \in S_1^n \quad (3)$$

$$\text{where } S_1^n = \left\{x \in R^n, x_i \geq 0, \sum x_i = 1\right\} \quad (4)$$

We have previously described a fast method of building support functions and presented the computational algorithm [18], including transformation of variables from R^n to S_1 . This improvement to earlier results has allowed practical application of the CAM to medium size problems involving 10-30 variables. In this paper, we apply a version of the CAM to molecular modelling problems where a global minimum of PES is needed.

3 Application to Molecular Structure Prediction

In this section, the application of the cutting angle method to the problem of prediction of molecular conformation is described. The function f to be minimized is the potential energy surface V .

Many models have been developed describing interactions within a molecule in terms of atomic bonds and effective interactions. Parameterizations of molecular potential functions (also known as Force Fields) include ECEPP, AMBER, CHARMM, GROMOS, MM3, DISCOVER and others (see the review [19]). Interfacing with these and other public domain and commercial software packages is important for practical applicability of novel optimization methods. In this study we chose ECEPP/3 force field (and software package) to model polypeptides.

In common with other authors, all the bond lengths and bond angles have been held fixed (eg, see References [11, 20-24]). Hence the conformations of the various molecules considered here, depend only on the torsional or dihedral angles. In turn, the potential energy depends only on the torsional potential energy terms and those corresponding to interactions between non-bonded atoms:

$$V = \sum_{i>j} V_{\text{non-bonded}}(r_{ij}) + \sum_k V_{\text{torsion}}(t_k), \quad (5)$$

where t_k are the dihedral (torsional) angles and $r_{ij} = r_{ij}(t_1, t_2, t_3, \dots)$ are the distances between atoms i and j , which in turn are functions of the dihedral angles t_k . In Eq.(5), the contribution of three-body and higher-order terms to $V_{\text{non-bonded}}$ has been neglected.

CAM was used to calculate the conformations of derivitised naturally-occurring amino acids of the form $\text{CH}_3\text{CO-NH-CRH-CO-NH-CH}_3$, where R is the amino acid residue. The parent amino acids were proline, alanine, histidine, tyrosine and glutamine, covering the range from 5 to 10 dihedral angles.

The CAM algorithm was interfaced with the ECEPP/3 potential energy surface of Scherga et al.[24] CAM was used as the "driver" algorithm to optimize the dihedral angles (global variables), but at each set of dihedral angles, the potential energies were evaluated by calls to ECEPP/3, and the three-atom angles ("θ angles") in the end groups were optimized within the ECEPP/3 package as local variables (described in the Appendix).

Conformations of several single aminoacids were computed initially to test the performance of the combined CAM-ECEPP/3 running on a cluster of DEC Alpha workstations.

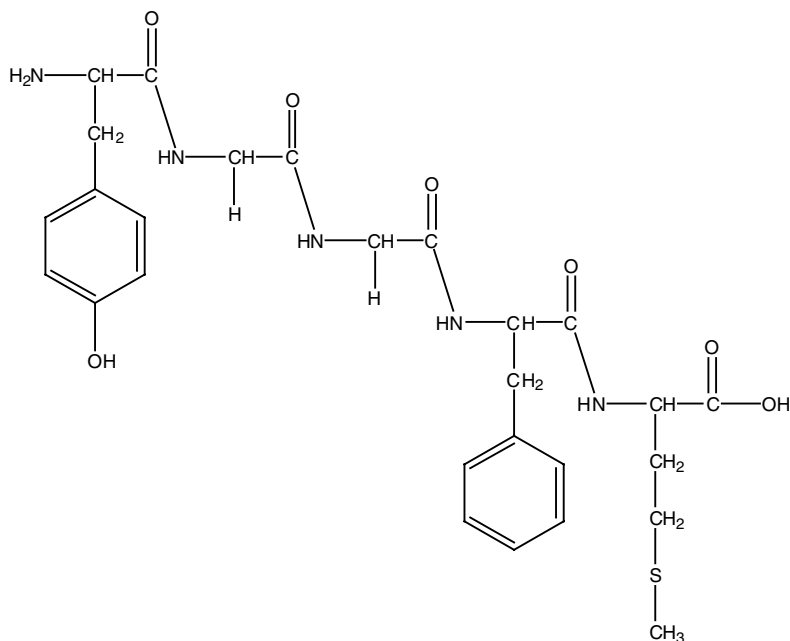
The CAM software was written in C. The ECEPP/3 package [24] was compiled on the same workstation and ran as an external process each time the potential energy was needed. The combined CAM-ECEPP/3 code worked as desired, and the choice of internal coordinates ensured that none of the chiral centres was inverted. (Racemisation can be a problem when Cartesian coordinates are used with Monte Carlo (Metropolis) methods.)

The CAM was then applied to a neutral form of met-enkephalin, shown in Figure 2, which is a benchmark molecular conformation problem [11]. It involves 24 independent

Table 1. Performance of the Cutting Angle Method on the prediction of conformation of single aminoacids and met-enkephalin

Parent aminoacid	Number of variables	Energy of most stable geometry	Function evaluations	CPU seconds
Proline (Pro)	5	-19.81	200	21
Alanine (Ala)	7	-5.18	5000	640
Histidine (His)	8	-8.92	8000	1067
Tyrosine (Tyr)	9	-8.48	4000	607
Glutamic acid (Glu)	10	-15.87	10000	4252
Met-enkephalin	24	-11.706	120,000	4740

Energies in kcal mol⁻¹. CPU time is for a single DEC Alpha workstation, except the last row, where CPU time is for a cluster of 36 DEC Alpha workstations

**Fig. 2.** The neutral form of met-enkephalin (Tyr-Gly-Gly-Phe-Met)

dihedral angles, giving rise to a very complex PES, involving, in the order of 10^{11} local minima [11]. This problem is very challenging because of the existence of several strong local minima, which terminate (or “trap”) local-descent algorithms. Local-search algorithms using random initial points fail to identify the global minimum; moreover, failures have been reported even if the initial point is chosen close to the

global minimum [11]. As with the substituted amino acids, CAM was interfaced with the ECEPP/3 package [24] (described in the Appendix).

The global minimum at $-11.706 \text{ kcal mol}^{-1}$ was found by the combination of the CAM and ECEPP/3 internal local optimizer using the procedure described in the Appendix.

4 Discussion

Floudas has previously reported that the α BB technique [11] has outperformed other methods (Monte Carlo, electrostatically driven Monte Carlo, diffusion equation, simulated annealing and others) for the met-enkephalin molecule. Calculations on this molecule have also illustrated the inadequacy of multistart local search techniques [11]. The α BB technique required about 400,000 function evaluations to locate the global minimum. Using the same ECEPP/3 potential function [24], the CAM of this work was able to locate the global minimum using 120,000 function evaluations, *thus out-performing the other optimization methods*. The met-enkephalin calculations have been performed in 79 minutes on a cluster of 36 DEC Alpha workstations.

ECEPP/3 was used as a standalone software package to which calls were made from our CAM driver module. CAM requires only the ability of third party software to compute the value of the objective function (PES), and thus poses little restriction on the choice of such software. This paper demonstrates that CAM can be implemented successfully on a multiprocessor computer.

Since CAM constructs a sequence of support functions that are strict lower bounds for the PES, Eq. (3) measures the uncertainty in the estimated global minimum of the PES when the algorithm is terminated after K iterations. Conversely, provided that there are sufficient computer resources, the algorithm can be continued until the estimated global minimum is within some *guaranteed tolerance* of the true global minimum. For computational convenience, the CAM algorithm uses normalized variables Eq. (4), which makes it compatible with third-party computer code describing any Lipschitz continuous function.

Although this paper has applied CAM to geometry optimization, CAM be applied to any global optimization method, including clustering and other problems in chemometrics [1], transition-state modelling [6], variational transition state theory [25], and optimization of flow analysis methods [2-4].

CAM can also be used to find the root(s) of nonlinear equations, e.g.

$$F(\mathbf{x}) = 0$$

through the transformation

$$G(\mathbf{x}) = \|F(\mathbf{x})\|^2 = 0$$

where G is the square of the norm of the original vector function F . The roots of F correspond to the global minima of G . Applications of this kind include the variational-theory method in quantum mechanics and Monte Carlo sampling in statistical

mechanics. These applications of the cutting angle method will be utilized in future work.

5 Conclusion

The problem of finding the global minimum on a hypersurface with multiple local minima has many important applications in chemistry. In this paper, we have applied a new global optimization method to the prediction of molecular conformation modeling. In these applications, the typical number of local minima grows exponentially with the number of variables, and is enormous even for relatively small molecules. Traditional local search techniques fail to detect global minima, and hence miss the likely molecular conformations, whereas stochastic global methods fail to confirm or give an estimate of the global minimum.

This paper has demonstrated that the cutting angle method (CAM) is competitive with (if not better than) other global optimization techniques in finding the minimum-potential-energy structures, when applied to benchmark molecular calculations for the met-enkephalin molecule.

CAM is a general global optimization method that can also be applied to other computational problems involving global minima or global maxima. It is a robust technique that can be applied to the optimization of non-differential (but Lipschitz continuous) functions.

Acknowledgments. The authors wish to acknowledge support from the Australian Partnership in Advanced Computing (APAC) (Merit scheme grant d88) and School of Information Technology, Deakin University (internal grant). KFL thanks Ms Jeanne Lee (李 静 宁) (Australian Catholic University) for encouraging and helpful discussions.

References

1. M. J. Adams, *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, London (1995)
2. A. Cladera, E. Gomez, J. M. Estela and V. Cerda, *Effect of Variables Influencing S-1/2 in Sequential Injection Analysis – Extrapolability of S-1/2 Based Results Between SIA Designs*, *Talanta*, 43 (1996), 1667–1674
3. S. M. Sultan, A. M. Almuaibed and A. Townshend, *Flow injection chemiluminescence determination of medazepam*, *Fresenius Journal of Analytical Chemistry*, 362 (1998), 167–169
4. A. Townshend and R. A. Wheatly, *Oxidative chemiluminescence assay of 2,4-dinitrophenylhydrazine*, *Analyst*, 123 (1998), 1041–1046

5. J.-R. Hill, C. M. Freeman and L. Subramanian, *Use of force fields in materials modeling*, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Wiley-VCH, New York, 2000, 141–216
6. J. E. Eksterowicz and K. N. Houk, *Transition state modeling with empirical force fields*, *Chemical Reviews*, 93 (1993), 2439–2461
7. A. Neumaier, *Molecular modeling of proteins and mathematical prediction of protein structure*, *SIAM Review*, 39 (1997), 407–460
8. P. M. Pardalos and C. A. Floudas, *Optimization in computational chemistry and molecular biology : local and global approaches*, Kluwer Academic Publishers, Boston (2000)
9. R. V. Pappu, R. K. Hart and J. W. Ponder, *Analysis and application of potential energy smoothing and search methods for global optimization*, *Journal of Physical Chemistry B*, 102 (1998), 9725–9742
10. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Oxford Science Publications, Oxford (1990)
11. C. A. Floudas, *Deterministic global optimization: Theory, methods, and applications*, Kluwer Academic Publishers, Dordrecht ; London (2000)
12. R. Horst and P. M. Pardalos, *Handbook of global optimization*, Kluwer Academic Publishers, Dordrecht ; Boston (1995)
13. M. Andramonov, A. Rubinov and B. Glover, *Cutting angle methods in global optimization*, *Applied Mathematics Letters*, 12 (1999), 95–100
14. A. Bagirov and A. Rubinov, *Global minimization of increasing positively homogeneous function over the unit simplex*, *Annals of Operations Research*, 98 (2000), 171–187
15. A. M. Rubinov, *Abstract convexity and global optimization*, Kluwer Academic Publishers, Dordrecht ; Boston (2000)
16. A. Bagirov and A. Rubinov, *Modified versions of the cutting angle method*, in N. Hadjisavvas and P. M. Pardalos, eds., *Convex analysis and global optimization*, Kluwer, Dordrecht, 2001, 245–268
17. P. Hansen and B. Jaumard, *Lipschitz optimization*, in R. Horst and P. Pardalos, eds., *Handbook of global optimization*, Kluwer, Dordrecht, 1995, 407–493
18. L. M. Batten and G. Beliakov, *Fast algorithm for the cutting angle method of global optimization*, *Journal of Global Optimization*, 24 (2002), 149–161
19. M. Jalaie and K. B. Lipkowitz, *Published force field parameters for molecular mechanics, molecular dynamics, and Monte Carlo simulations*, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Wiley-VCH, New York, 2000, 441–486
20. J.-P. Ryckaert, *Special geometrical constraints in the molecular dynamics of chain molecules*, *Mol. Phys.*, 55 (1985), 549–556
21. C. Ciccotti and J.-P. Ryckaert, *Molecular dynamics simulations of rigid molecules*, *Comput. Phys. Reports*, 4 (1986), 345–392
22. R. Kutteh and T. P. Straatsma, *Molecular dynamics with general holonomic constraints and application to internal coordinate constraints*, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Wiley-VCH, New York, 1998, 75–136
23. G. Némethy, M. S. Pottle and H. A. Scheraga, (*ECEPP/2: empirical conformational energy program for peptides and proteins*), *Journal of Physical Chemistry*, 87 (1983), 1883
24. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey and H. A. Scheraga, (*ECEPP/3: empirical conformational energy program for peptides and proteins*), *Journal of Physical Chemistry*, 96 (1992), 6472
25. R. G. Gilbert and S. C. Smith, *Theory of Unimolecular and Recombination Reactions*, Blackwells Scientific, Oxford (1990)

Appendix: Schematic Overview of the Software Setup

ECEPP/3 software was installed on the same computer as our CAM global optimization algorithm. The variables in CAM were transformed into the appropriate domain for the dihedral angles $[-\pi, \pi]$, and their values were supplied to the ECEPP/3 software in a suitable configuration file. ECEPP/3 was called as an external process, which minimized the potential energy with respect to local variables keeping the global variables constant. The CAM then read the value of the energy from the file generated by ECEPP/3. This process was repeated at each iteration of the CAM, and at the end (after a predefined number of iterations), CAM generated 100 good starting points for the ECEPP/3 local minimizer. Finally, the local optimization process with respect to all variables started from the specified 100 points to quickly improve the CAM solutions. The best solution was selected as the global minimum.

