3 Stage Hybrid Optimisation Algorithm for the Bonded Lennard-Jones Molecular Conformation Problem

Bennett Gale

August 2018

Contents

1 Introduction

The model molecular energy minimisation problem involves the geometric configuration of a chain of atoms in two dimensions, each connected by rigid bonds of unit length, such that the total potential energy yielded by the sum of individual Lennard-Jones potentials is globally minimised. The total potential energy of the cluster is given by

$$V(r) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(r_{ij}^{-12} - 2r_{ij}^{-6} \right)$$

Where r is the euclidean distance between a pair of atoms and N is the number of atoms in the chain.

The energy minimisation problem is historically difficult because of the exponential relationship between the locally stable structures and the number of atoms, N, in the molecule. That is, as the size of the cluster increases, the number of local minima increases exponentially such that when N approaches 100, there exist more than 10^{140} minima [1]. It is therefore necessary to employ sophisticated search algorithms in order to make larger input sizes attainable.

Section 2 will discuss some of the research that has already taken place in this area. Section 3 will detail the algorithm proposed by this paper.

2 Literature Review

Genetic algorithms have been applied [2], [3] successfully in order to extend the range of input size N that is reachable through structural optimisation. This population based approach emulates the Darwinian process of natural selection, and works by performing crossover operations on pairs of candidate configurations with selection probabilities based on their fitness, producing new child nodes which contain genetic material from both parents. With every crossover operation there is a small chance of subsequent random mutation of the child configurations, the purpose of which is to promote genetic diversity within the population and prevent premature convergence. Over many generations the population converges towards a global minimum as each configuration communicates information about the global search space through the crossover operation. This method has been successful in producing results that were previously unseen in literature for certain input sizes [3].

Judson *et al.* compare the performance of GAs with simulated annealing, a probabilistic global optimisation technique in which Monte Carlo steps are taken and new states are accepted if they are fitter, or with a probability based on a function of a temperature value [2]. When applied to the molecular conformation problem, they found that the GA method and the SA method, both applied as global meta-heuristics in conjunction with gradient-based local optimisers (conjugate gradient), similarly outperformed purely stochastic methods [2]. Judson *et al.* classify points on the model energy surface as being in one of

three states: knotted and of high energy, unknotted and of low but not minimal energy, and unknotted with globally minimised energy [2]. They applied both GA and SA in order to make the transition from the first state to the second state before applying local optimisation to descend to the bottom of the basin.

Genetic operators for the two dimensional bonded Lennard-Jones problem have since been further developed with successful results [3], proving the GA approach to be a viable global optimisation technique. Pullan [3] proposes three genetic crossover methods and four mutation operators and applies these in a parallel genetic algorithm to find global minima for most values of $N, 2 \leq N \leq 55$, as well as a minimum for $N = 61$ which was previously unseen in literature.

This two stage hybrid method has been successful, but Judson *et al.* also suggest the possibility of adding a third optimisation stage which is intermediate between the global and local gradient methods [2]. It is thought that a non-gradient optimisation method such as the simplex method may provide a means of stepping over barriers into possibly deeper minima surrounding those found by the global method in the first phase.

More recently, the Conformational Space Annealing Method [4] has been developed as a global optimisation algorithm that has produced very good results when applied to the problem of an unbound cluster of atoms with Lennard-Jones relationships. This algorithm combines concepts of Monte Carlo with minimisation, genetic algorithms, and simulated annealing. While this technique has not been applied to the bonded variant of the Lennard-Jones conformation problem as discussed here, it demonstrates that a hybrid approach combining SA and evolutionary population based algorithms may be a viable technique in a similar problem domain.

3 Proposed Algorithm

As discussed in section 2, global meta-heuristics such as the genetic algorithm, have been used successfully as a means of navigating the global search space and avoiding local minima that are caused by knots in the configurations [2], [3]. These methods use the global optimiser to find unknotted potential candidate configurations and then refine them using gradient-based methods. Judson *et al.* have suggested a three stage system in which a non-gradient optimiser such as the simplex method may be used as an intermediate stage between the global and local-gradient optimisation in order to traverse barriers into potentially deeper basins [2].

As simulated annealing has been found to be successful when combined with evolutionary algorithms in similar problem domains [4], this paper will explore its use as an intermediate non-gradient method such that the proposed algorithm has three levels: global meta-heuristic (GA), intermediate non-gradient optimisation (SA) and local gradient-based optimisation (BFGS). The performance of this three stage optimisation method will be compared with a purely evolutionary algorithm without the intermediate step.

3.1 Genetic Encoding

Encoding each molecule as a chain of two-dimensional cartesian coordinates would require constraints to be enforced so that the relative orientations and unit length bonds are maintained. In order to avoid this, therefore, molecules will be encoded as sequences of $N-2$ angles, where N is the number of atoms in the molecule and each angle, α_i represents the rotation of the segment following atom a_i. This encoding scheme means that cartesian coordinates must be calculated at each Lennard-Jones potential computation as it requires euclidean distances between pairs of atoms to be computed, but results in much simpler crossover and mutation operations.

3.2 Selection & Crossover Method

An elitist methodology will be utilised throughout the GA optimisation, meaning that the single fittest configuration will be passed on unmodified to subsequent generations in order to preserve beneficial traits and provide a mechanism for remembering information about the best configuration that has been found so far.

Pairs of parents to be crossed over will then be selected using the Roulette method, where the probability for a potential parent to be selected is based on a function of its fitness value. The 'elite' configurations that were inherited from the previous generation will also be candidates for selection using this method. Selected parents will undergo one of the following crossover operators with equal probabilities of occurring:

3.2.1 Single-Point Crossover

A single atom a_i will be randomly selected. Atoms to the left of this point in one parent will form the left segment of the child, and the right segment of the child will be similarly formed from the atoms to the right of the second parent. A second child may be generated by repeating with the opposite segments (the right segment from parent one and the left segment from parent two). Finally, the angle α_i at the crossover point will be iterated through discrete steps until an angle is found which results in the lowest energy, as described by Pullan [3].

3.2.2 Two-Point Crossover

Two crossover points will be selected at random and the molecules will be combined similarly to the single-point crossover, with each child containing two segments from one parent and one segment from the other. As with the single point crossover, the angles of both crossover points will be optimised.

3.3 Mutation and Intermediate Optimisation

The mutation mechanism is where the intermediate non-gradient optimisation algorithm will be incorporated. In traditional genetic algorithms, mutation has

a small chance of occurring and is random, resulting in some states which are less fit in the hopes that this will later lead to a deeper basin in the search space. By incorporating simulated annealing at the mutation stage it is hoped that the configuration will be dragged in a downward direction while also providing enough stochastic properties to allow boundaries between minima to be traversed, which is the goal of any mutation operator.

Mutation will have a low chance of occurring so as not to corrupt fit configurations too frequently and lose information provided by the global search. Each generation, child nodes that are produced during the crossover stage will have a small chance of becoming initial states in the simulated annealing search.

3.3.1 Annealing Schedule

During annealing random angles are selected and modified by random amounts. If the change produces an improved state, it is accepted. If the change produces a worsened state, it is accepted with a probability of $1 - e^{\delta E/kT}$ such that the current 'temperature' as well as how bad the state is determines whether it is accepted.

The goal at this intermediate stage is not to find the global minimum, but merely to provide a more localised search for deeper basins in the vicinity of the current configuration; it is hoped that if the GA search has found a local minimum in close proximity to the global minimum, the SA stage will take large enough steps in order to climb over the boundary into the neighbouring basin where it can be finally optimised by the gradient search in the next stage.

Therefore, an annealing schedule that prioritises speed over optimality will be chosen.

3.4 Local Optimisation

The final stage of the algorithm, after a candidate configuration has been found by the GA and its immediate vicinity has been further explored through mutation, is to refine the configuration using gradient-based local search. This stage will use the gradient of the objective function in order to guarantee that the bottom of the current basin will be reached. The quasi-Newton method Broyden–Fletcher–Goldfarb–Shanno (BFGS) will be applied at this stage resulting in a fast descent towards the current local minimum at each stage.

The coordinates of this locally optimised configuration will also be passed back to the GA. This is known as a 'Lamarkian' process and has been found to be beneficial when used with this problem [2].

References

[1] D. Daven, N. Tit, J. Morris, and K. Ho, "Structural optimization of lennard-jones clusters by a genetic algorithm," *Chemical Physics Letters*, vol. 256, no. 1, pp. 195–200, 1996, ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(96)00406-X. [Online]. Available: http://www.sciencedirect.com/science/article/pii/000926149600406X.

[2] J. R. S., C. M. E., M. J. C., H. A., and G. D., "Do intelligent configuration search techniques outperform random search for large molecules?" *International Journal of Quantum Chemistry*, vol. 44, no. 2, pp. 277–290, DOI: 10.1002/qua.560440214. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560440214. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560440214.

[3] W. Pullan, "Genetic operators for a two-dimensional bonded molecular model," *Computers & Chemistry*, vol. 22, no. 4, pp. 331–338, 1998, ISSN: 0097-8485. DOI: https://doi.org/10.1016/S0097-8485(97)00068-5. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097848597000685.

[4] J. Lee, I.-H. Lee, and J. Lee, "Unbiased global optimization of lennard-jones clusters for n¡=201 using the conformational space annealing method," *Phys. Rev. Lett.*, vol. 91, p. 080 201, 8 Aug. 2003. DOI: 10.1103/PhysRevLett.91.080201. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.91.080201.